Characterizing and Targeting the Chromatin Determinants of Cancer Cell Identity:

Cancer’s Addiction to its Originating Cellular Lineage

by

Abhijit Parolia

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Molecular and Cellular Pathology)
in The University of Michigan
2021

Doctoral Committee:

Professor Arul M. Chinnaiyan, Chair
Assistant Professor Marcin P. Cieslik
Professor Gregory R. Dressler
Professor Diane Robins
Adjunct Associate Professor Scott A. Tomlins

I am a part of all that I have met;
Yet all experience is an arch wherethro'
Gleams that untravell'd world whose margin fades
For ever and forever when I move.

- ALFRED LORD TENNYSON, “ULYSSES”

You are this poet’s magnum-opus,
blessed without creation yet hers to savor.
She is the thick mud beneath your lotus,
and will let you bloom with all her fervor.
Remember, forever the sun shall rise for you,
through the dark grey clouds
For you, my love, hold the strength to be
Fast, Furious and Loud.

- RINI KAUSHAL PAROLIA, “MY POETIC MUSE”

Abhijit Parolia

aparolia@umich.edu

ORCID iD: 0000-0003-0238-221X

DEDICATION

To my late grandmothers and foremost teachers,

Smt. Chandra Kala Gupta

&

Smt. Shashi Gupta,

who always stood by my side and inspired me to reach for the stars.

ACKNOWLEDGEMENTS

I truly am a part of all that I have met. In my academic journey in North America, now over a decade long, I have had the utmost privilege of learning in some of the best schools and being taught and trained by some of the most brilliant teachers and selfless mentors. There are many who have made this journey possible and I am immensely grateful to all of them:

I arrived in North America wanting to become a physicist. In fact, I had dropped biology as a subject in grades 11th and 12th. At the time, the mathematical logic and real-world application of basic physical laws gravitated me to the subject. This was primarily because of how I was taught physics by my high school teacher, Mr. Sami Ullah Kazmi, who is single-handedly responsible for igniting in me the love for science. In hindsight, however, it was a seemingly simple question asked by my Intro to Biology course instructor, Dr. Pamela Kalas, at the University of British Columbia (UBC) that resonated the strongest with my natural frequencies: "*What is a gene?*" she asked in her opening lecture. This highly interactive course went on to ask many similar questions that opened my eyes to a new biological discipline—one that did not involve memorization and regurgitation of anatomical labels and terminologies. I forever will be grateful to these two teachers who have majorly influenced the course of my academic journey.

I was fortunate to get selected into an honors program at UBC that was tailor-built to provide wet-lab experience to aspiring molecular biologists. The curriculum mandatorily demanded work experience in research laboratories, which landed me in the cancer-focused, academic laboratories of Dr. Cheryl D Helgason and Dr. Yuzhuo Wang at the BC Cancer Research Center, Canada. Here, I was put under the tutelage of Dr. Francesco Crea, who had recently joined the lab as a postdoctoral fellow from Italy. This would be an inflection point in my academic life, setting me on an unimaginable upward trajectory. Francesco has taught me everything I know about prostate cancer biology and the nuances surrounding its clinical management. I spent as much time as possible working with him and learning as much as I could for the next 2 years—taking on even part-time voluntary work during school semesters. I am truly thankful for all the long hours Francesco spent training me and answering all my questions. I had the joy of writing my very first first-author paper under his mentorship and learned from him the power of working in a team. Francesco has had a very strong influence on my career and, in many ways, is my scientific parent. He still continues to be the person I turn to when in need of advice. I am grateful to Francesco and his family for all the love and warmth they have given me.

In the last year of my undergraduate degree, Dr. Yuzhuo Wang (referred to as YZ in the group) called me in his office and counseled me on the laboratories I could explore for my Ph.D. training. He was very clear, however, that I was permitted to leave only if I could secure admission into a lab that he considered to be more resourceful than his. Looking back, I sincerely thank YZ for this candid conversation as one of the labs he highlighted was that of Dr. Arul M. Chinnaiyan at the University of Michigan. I was granted admission to the Michigan graduate program and was thrilled when Arul agreed to take me into his lab. I had been in awe of Arul's scientific work and accomplishments, and it has been a privilege to be mentored by him through my Ph.D. training. I would like to particularly thank him for the intellectual autonomy, cross-disciplinary expertise, and technological resources he provided, while always being available when I needed guidance.

Arul is a visionary leader and his lab functions as an independent institution. Being a small part of it has been truly inspirational for me. It has allowed me to grow as a scientist and reach an important conclusion: success in modern-day science requires working in large multi-disciplinary teams. I will always be grateful to Arul for his inspiring mentorship, for the many scientific lessons, and for admitting me to his lab. It placed me in the company of extremely brilliant individuals from whom I have learned so much in the past five years.

Dr. Marcin Cieslik happens to be one such individual, who at the time was a bioinformatics postdoc in the lab. I have worked very closely with Marcin during my Ph.D.—discussing with him ideas for potential projects and mulling over the mechanistic meaning of new data for hours. He is a brilliant scientist and I find his company to be very enriching. He challenged me to think deeper about every piece of data and taught me the finer concepts of cancer genomics. I largely owe the successful completion of my primary thesis project to him. He has been a remarkable mentor and provided pivotal support and guidance throughout.

During the second half of my Ph.D., I got the opportunity to work closely with Dr. Lanbo Xiao on a new drug we had received from a pharma collaborator. Lanbo is a CRISPR guru in the lab who has taught me almost everything I know about this technology. I was introduced to new CRISPR applications via weekly emails from Lanbo with ideas on how they can be used in our on-going projects. Lanbo is an extraordinary team player and I aspire to reach his level of collaborative spirit someday. I have learnt from Dr. Yuanyuan Qiao the skill to meticulously maintain experimental records and organize data. I have also had enriching interactions with Drs. Yuping Zhang, Pankaj Vats, and Pushpinder Bawa in the lab. They comprise the computational backbone of the lab and I am indebted to them for promptly responding to my uncountable emails and requests for complex analyses. Pankaj has been a calming influence in my personal life and

Pushpinder is a dear friend. I have also received immense good wishes and many surprise lunches from Dr. Rahul Mannan, Ingrid Apel, Jyoti Athanikar, Dr. Seema Chugh, Jonathan Gurkan, Dr. Palak Shah, Lakshmi Dommeti, and Dr. Sunita Shankar. I have shared numerous fond memories with these colleagues and their families. I also found a close friend in Jae Eun (Jenny) Choi—a fellow graduate student—who particularly made my courthouse wedding in Ann Arbor a memorable occasion.

The fondest memories from my Ph.D. are associated with my interactions with students who directly worked with me. We fondly referred to our group as “FOXES,” in reference to the name of the protein we study (i.e. FOXA1). Takahiro Ouchi, the very first student to join, has made meaningful contributions to this work. In the three-plus years, until his graduation, Taka spent many hours in the lab and helped run many important experiments. He is extremely organized and I repose great trust in his experimental ability. He is now at Wayne State University pursuing an M.D. degree. Divya Macha (a high school student at the time) and Mustapha Jaber followed next and injected the much-needed youthful energy into the group. Sandra Carson joined after them as a full-time technician. Sandra had a remarkable impact on the group as she helped organize many things and truly multiplied our productivity. I not only had a gifted apprentice in her but also found a very dear friend. She is now at Cornell, starting her Ph.D. journey. The most recent entrant is Sanjana Eyunni, who is a second-year Ph.D. student in Arul’s lab. She is extremely diligent, creative, and smart, and I look forward to working with her on various offshoot projects from our seminal study. I thank all the FOXES from the deepest bottom of my heart for being terrific team players. They have played a major role in enabling my success and have taught me a lot in the process.

I am also grateful to many members of the administrative and technical teams in the Chinnaiyan lab. A special mention goes to Xuhong Cao, our lab manager, who serves as a nodal point that connects the different groups within our lab—ensuring smooth and coordinated operation at all times. Xuhong has been instrumental in providing the logistical support for my projects and I am grateful to her for always being there in times of need. Also, I would like to thank Fengyun Su and Rui Wang for the countless number of sequencing libraries they have helped generate and quality check. Also, I thank Diana Banka, Jyoti Athanikar, Sisi Gao, Stephanie Ellison, Temuulen Johnsen, and Jennifer McPeake who have helped at various points to assemble and proofread grants and scientific manuscripts. I am particularly grateful to Jyoti who kindly offered her backyard and other critical resources for holding my wedding reception. Karen Giles, Nikki Cortis and Christine Betts for coordinating our lab schedules and running things smoothly. Melissa Dunn and Chia-Mei (Grace) Huang for keeping the lab well stocked and supplied with essential reagents and equipment. I also thank every member of the MCTP CLIA team for generating and curating the critical clinical database that laid the strong foundation for this research. I thank all other members of the Chinnaiyan family who have made many things possible for me and provided for a rich training environment—some of whom I am sure I am forgetting to mention here.

I sit here, with the wisdom of hindsight, attempting to pinpoint the precise moment we came up with the idea to classify the FOXA1 alterations, and I draw a blank. This makes me believe that ideas do not germinate in a vacuum; they rather originate from the ecosystem in which you live and function. Thus, I thank each and every member of the Chinnaiyan lab with whom I have ever interacted—be it either for a minute in a corridor or an hour in an office—for these interactions must have subliminally influenced me and my scientific thought process. In addition to the people mentioned above, my numerous unscheduled and passing conversations with Dr. Scott Tomlins, Dr. Rohit Malik, Jeremy D'Silva (he is now at Stanford pursing his Ph.D.), Dr. Chandan Kumar Sinha, Dr. Mohan Dhanasekaran, and Dr. Sriram Venneti come to mind. It is the company of every member of MCTP, therefore, that has shaped the projects in the form they are presented in this thesis and I am grateful to all of you for your collaboration.

I also thank my thesis committee members, Dr. Diane Robins, Dr. Greg Dressler, and Dr. Scott Tomlins. They have always been available and have provided constructive and meaningful feedback that has improved the quality of the science presented in this thesis. I am also grateful for all the support from my Ph.D. program. Dr. Zaneta Nikolovska-Coleska (director) and Laura Labut (student coordinator) have played instrumental roles in ensuring my well-being and success within the program. I have also established some strong friendships with fellow graduate students: Dr. Carrie-Anne Malinczak, Dr. Andi Cani, Rita Avelar, and Jessica McAnulty. I particularly enjoyed collaborating with Carrie-Anne who is my close friend and personal immunology instructor, and spending quality time with Rita who ensured that I was never alone on Christmas or Thanksgiving holidays.

I would like to thank my family next. In pursuit of quality higher education, I have now been living away from them for over a decade. I have been blessed to be raised in a household with grandparents. In fact, I had a bounty with five of them on the maternal and paternal sides: Surya Prakash Gupta, Shashi Gupta, Chandra Kala Gupta, Ramji Lal, and Kamlesh Gupta. Their lives and upright principles have been an inspiration to me. They have served in too many roles at different stages of my life for me to specifically recount. They have been my nannies, teachers, enablers, financiers, guarantors, protectors, motivators, cheerleaders, promoters, and idols. It would be sufficient for me to say that none of this would be possible without them. This is particularly true for my maternal grandfather Surya Prakash Gupta, a highly accomplished and decorated senior lawyer with honorary doctorates, whose magnanimous presence has been constantly reassuring and functioned as a safety net for my excursion into foreign lands. My only regret is that two of my grandmothers, Chandra Kala Gupta and Shashi Gupta, who taught me the subjects of Hindi and Sanskrit, are not with me to witness this day. I dedicate my Ph.D. thesis to them and their unrelenting faith in my abilities.

I reflect my deepest gratitude to my parents, Krishna Kumar Parolia and Arunima Parolia, and to my older brother, Aditya Parolia. My parents always ranked education at the very top and have made many sacrifices to afford the best education for me and my brother. They have spent their lives' savings on paying our international student fees and I cannot ever be grateful enough for this selfless act. While growing up, I always wanted to imbibe my mother's cerebral prowess and my father's amiable and humble spirit. My mother is a Ph.D. in Biochemistry. Sadly, she had to abruptly halt her career for the lack of quality academic institutions back home. Thus, she strongly advocated for sending me abroad for higher education, and my father, a chemical engineer, fully supported this view. I owe to my parents and their foresight my academic journey in this part of the world, and I hope they feel equal ownership of the work presented in this thesis. I am also grateful to my older brother and all the life lessons I have learned from him. His fourteen-page letter, which he wrote to me when I was leaving India, has become my holy scripture. I often re-read it to seek inspiration and clarity in disquiet times.

Last, but not least, I thank my best friend and life partner, Rini Kaushal Parolia. She is a therapist and a poet. She has been my rock and my sunshine—with me through all seasons—for the past ten plus years. We started dating back in high school and got married in the Ann Arbor courthouse on Sept 6th, 2019. It is my most significant accomplishment to this date. She truly deserves a lot of credit for allowing me to spend long hours in the lab and often postponing our vacations due to experiments. She is an incredible human being and an even superior poet, and I owe her all of my life's happiness and past, present, and future successes. I also thank Rakesh Kaushal, Nisha Kaushal and the rest of my in-laws, who have showered me with immense love and good wishes.

So today, I may appear taller for I stand on the broad shoulders of all of the people I mentioned above. To all of you, I forever will remain indebted. However, I promise to pay this debt forward by affording the same opportunities for my children and all the students who work with me in my scientific career.

Today I may appear taller, like the conifer,
slowly branching out into the limitless sky.
However, I must pen this gratitude metaphor
for all those who nurtured it, and never let it die.

\- RINI KAUSHAL PAROLIA

PREFACE

Cancer is a complex disease that is initiated by genetic perturbations in normal cells leading to uncontrollable cellular division. The ensuing tumor mass, at both primary and distant metastatic sites, hinders the normal functioning of essential organs and causes death. With the advent of modern multi-omics technologies, we have generated comprehensive molecular portraits of human cancers; however, the biological mechanisms that underlie the common genetic alterations in cancer remain largely uncharacterized and understudied. This has markedly impeded the clinical translation of extensive genomic information acquired from human malignancies.

As a young scientist, a feature of cancer genomic profiles that intrigues me the most is that somatic alterations across cancer types show a tissue-specific pattern of recurrence and distribution. This implies that the originating/precursor cell lineage gene programs persist in cancer cells (albeit aberrantly) and dictate the set of oncogenes that are competent to drive transformation and/or progression. As articulated in detail in this following thesis, systematic interrogation of driver lineage-oncogene associations can uncover novel molecular dependencies that can be co-targeted to inhibit cancer cell growth and survival—a concept herein termed as “targeting the cancer cell identity.” These lineage-targeted therapies are most likely to be synergistic with existing therapies that target oncogene-addiction, potentially leading to more durable therapeutic benefits in patients with advanced incurable cancers.

Given the remarkable technological advances that allow us to read, write, and edit the chemical letters that make up our DNA, and our ability to acquire single-cell or even single-molecule level information, I believe we are stepping into another golden age for molecular biology. Particularly, the view of cancer genomes as natural CRISPR screens can help elucidate not only the pathogenic roles of driver proteins, but also specific functions of key constituent residues and regulatory protein domains in normal physiology. I eagerly anticipate and look forward to the discoveries in this research discipline in the years to come.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

ABSTRACT

Lineage-restricted somatic alterations are a prominent feature of human cancers. This strongly implies that the originating cell lineage programs persist in the transformed state and determine which oncogenes can drive hyper-proliferation and/or disease progression. This concept is exemplified by sex hormone receptors, namely androgen receptor (AR) and estrogen receptor (ER), which themselves are not mutated/amplified in the primary prostate or breast tumors, respectively; however, inhibition of their lineage-specific programs results in significant therapeutic benefit. This raises some pertinent questions: 1) how do somatically altered oncogenes interact with and repurpose the originating cell lineage pathway to enable carcinogenesis? and 2) do the aberrant lineage–oncogene networks engender specific molecular dependencies in cancer cells that can be therapeutically targeted? These questions lie at the center of my doctoral thesis, wherein we molecularly characterize the AR and FOXA1 lineage–oncogene interaction as well as evaluate the targetability of a shared chromatin dependency (i.e., the SWI/SNF complex) to derive therapeutic benefit in lineage-addicted subtypes of prostate cancer.

Mounting evidence suggests an instructive role for pioneer transcription factor activity at enhancer sites in defining cellular identity. Pioneer factors independently bind to and de-compact chromatin at enhancer elements, thereby priming them for transcriptional activation. In prostate epithelia,

FOXA1 pioneers the binding sites for AR and enables its lineage-specific proliferative gene program. Intriguingly, FOXA1 is frequently altered in prostate tumors and in the primary study published on my thesis we showed that FOXA1 alterations—present in over 35% of metastatic prostate cancer—recur into three structural classes that diverge in clinical incidence. While missense and in-frame indel mutations in the FOXA1 coding region clustered within the C-terminal edge of the DNA-binding domain (class 1), frameshift mutations accumulated in the C-terminal half of the protein (class 2). As an entirely novel discovery, we also found structural rearrangements within the highly-syntenic FOXA1 locus in 20-30% of metastatic prostate cancer (class 3). Notably, FOXA1 alterations show class-specific gain-of-function: the initiating class 1 mutants gain nuclear mobility to strongly activate the luminal AR program of prostate tumorigenesis; the acquired class 2 mutants gain cistromic-dominance due to increased DNA affinity and de-repress the WNT program of cancer metastasis; and the stage-wise enriched class 3 rearrangements duplicate or reposition a conserved enhancer element—which we annotated as FOXA1 mastermind (*FOXMIND*)—to drive overexpression of *FOXA1* or other oncogenes.

Relative to the normal prostate tissue, in prostate cancer both AR and FOXA1 cistromes are markedly expanded to commission novel *cis*-regulatory elements. A proteomics-based search uncovered some of the core SWI/SNF—a nucleosome remodeling complex—components as shared chromatin interactors of AR and FOXA1 at enhancer elements in cancer cells. Thus, we developed a novel proteolysis targeting chimera (PROTAC) degrader against ATPase subunits of the SWI/SNF complex—SMARCA2 and SMARCA4—and assessed its anti-tumor efficacy. In a panel of over 65 cell lines, the SWI/SNF ATPase degrader showed preferential cytotoxicity in AR/FOXA1-positive prostate cancer relative to benign prostate, as well as other cancer cell lines. Mechanistically, SMARCA2/4 degradation led to rapid nucleosomal compaction and chemical

decommissioning of the core AR/FOXA1-enhancer circuitry, dislodging both AR and FOXA1 from the chromatin, and attenuating their collaborative gene program that drives proliferation. Furthermore, SWI/SNF inactivation also disrupted DNA-looping interactions between super-enhancers and gene promoters, thereby tempering hyper-expression of the AR, FOXA1, TMPRSS2-ERG, and MYC genes themselves in cancer cells. The SMARCA2/4 degrader showed remarkable synergism with AR antagonists, inducing disease remission in all animals receiving the combinatorial regimen. This work has been recently submitted for peer-review and received major funding for clinical evaluation of our SWI/SNF compound in early-phase human trials.

In summary, targeting lineage addiction in cancer cells is a novel concept and is likely to synergize with existing oncogene-targeted therapies. This doctoral thesis establishes FOXA1 as a principal oncogene and uncovers SWI/SNF complex as a novel lineage-associated dependency in AR-addicted prostate cancer. This work offers the first classification scheme of the highly recurrent FOXA1 alterations in human cancers and is the first preclinical demonstration that impeding physical chromatin accessibility can be a potent therapeutic strategy in enhancer-addicted malignancies.

Chapter 1

Lineage Addiction in Human Cancer and Prostate Carcinogenesis

The originating lineage conditions somatic alterations in cancer

In the post-genomics era, we have gone onto sequence thousands of cancer genomes identifying recurrent somatic alterations in all common human malignancies. These comprehensive genomic catalogs of cancer have revealed that somatic driver alterations recur in a lineage-restricted fashion [1–3]. For instance, activating KRAS mutations (e.g. G12C/D) that are very common in pancreatic ductal (>95% of the cases [4]) and colorectal adenocarcinomas (>50% of the cases [5]), are rarely found in cancers originating from the prostate or breast epithelia [6–8]. Similarly, the genetic fusion between the *BCR* and *ABL1* genes resulting in a hybrid oncoprotein is an abnormality found in leukemia cells, particularly chronic myeloid leukemia [9,10], and is rarely detected in solid malignancies. This strongly suggests that the originating cell lineage influences carcinogenesis by dictating which genes—when somatically mutated or rearranged—can competently drive transformation and/or metastatic progression. In other words, only select genes have the ability to activate tumorigenic pathways within the confines of a given cellular lineage. This raises some pertinent biological questions: 1) What molecular features of the originating lineage persist in cancer cells? 2) Do driver oncogenes rely on and repurpose the lineage-instructing transcriptional machinery to activate neoplastic gene programs? 3) Can we target the chromatin and transcriptional determinants of lineage identity to derive therapeutic benefit in human cancers? These questions lie at the center of the research presented in this doctoral thesis, wherein I posit the emergence of tumor cells from their normal precursors to involve a complex interplay of genetics, epigenetics, and, most importantly, the cell lineage (**Figure 1-1**). The study of lineage-specifying pathways in the tumor cell-of-origin and associated cancer cells will provide new insights into tumor biology and uncover a distinct class of lineage-associated cancer dependencies that can be therapeutically co-targeted to extort more potent and durable disease remissions—a concept herein termed as "targeting the cancer cell identity".

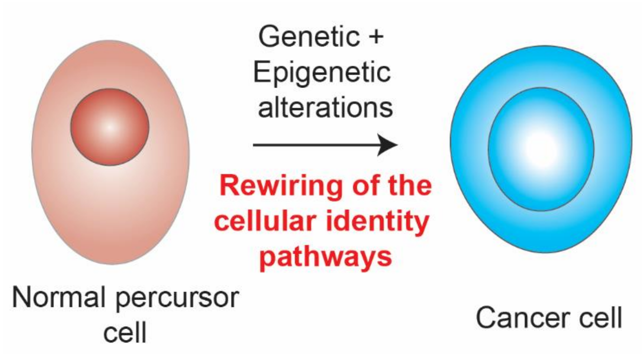


Figure 1-1: Lineage-addiction in cancer. Lineage-addiction of cancer: The originating cell lineage condition somatic alterations in human cancers. This model proposes hijacking of the lineage machinery by classical oncogenes to drive the malignant phenotype.

Prostate cancer statistics and clinical management

Prostate cancer is the most commonly diagnosed cancer and the second-leading cause of cancer-related death in North American men, accounting for more than 33,300 lives in 2020 alone [11]. It is estimated that 248,530 new cases of prostate cancer will be diagnosed in 2021 (accounting for 28% of all cancers diagnosed in men in the same year [11,12]), resulting in more than 34,000 deaths in the United States. However, prostate cancer mortality has declined by more than 50% from 1993 to 2017 [13,14]. This is mainly due to the advent of better diagnostic tools that detect the disease in early stages [15] and the development of effective therapies.

In the early 1940s, a clinical study showed the ablation of testicular androgen (through either bilateral orchiectomy or androgenic neutralization by estrogen administration) in men with prostate cancer triggered marked atrophy of the prostate epithelia [16,17]. Ever since, androgen deprivation therapy (ADT) has been the mainstay therapy for prostate cancer patients [18], and was also recognized with the Nobel Prize in Medicine in 1966 [19]. Much later, in the early 1980s, androgen receptor (AR)—a nuclear transcription factor—was identified as the primary effector of androgenic male hormones, i.e. testosterone [20]. This led to the development of direct AR antagonistic drugs as second-generation therapeutics [21]. Thus, clinical management of prostate cancer involves generic local therapies (e.g., surgery or radiation) as well as specific systemic therapies involving blockade of the AR/androgen signaling axis that fuels cancer growth, as depicted in **Figure 1-2** and described below.

In current times, about 85% of prostate cancer cases are diagnosed when the disease is organ-confined [22–24]. The recent incorporation of multi-parametric magnetic resonance imaging in prostate cancer screening will further enhance our ability to detect clinically relevant primary disease earlier, and the recent FDA approval of PSMA-based imaging tests will facilitate early detection of local invasion or metastasis [25,26]. The primary disease at the time of diagnosis is histologically graded using multiple biopsies. If the primary tumor is deemed histologically aggressive as per the Gleason score/Grade Group method [27] (where a score of 7—Grade Group 2—generally demarcates an aggressive tumor [27,28]), it is treated by prostatectomy (i.e. surgical resection of the prostate gland) or radiation (through either external beam or internal brachytherapy [29]) [30]. The patients with Gleason score 6/Grade Group 1 disease are generally put on active surveillance or watchful waiting [31]. The last decade has particularly witnessed many concentrated efforts to stratify the indolent from the aggressive (likely to metastasize or be resistant to therapy) subtype of the primary disease using minimally invasive approaches [32]. These include detection of tumor-specific prognostication markers (including both genomic alterations or aberrantly expressed transcripts) in cell-free DNA from patient plasma or urine or in DNA from circulating tumor cells [33–36]. To this end, usage of lineage-specific long non-coding RNAs [37–39] and RNase-resistant circular RNAs [40–42] have emerged as meaningful biological analytes. The primary intent behind these diagnostic technologies is to avoid unnecessary overtreatment of indolent disease [31,43], which may never progress to affect patients' overall survival or quality of life.

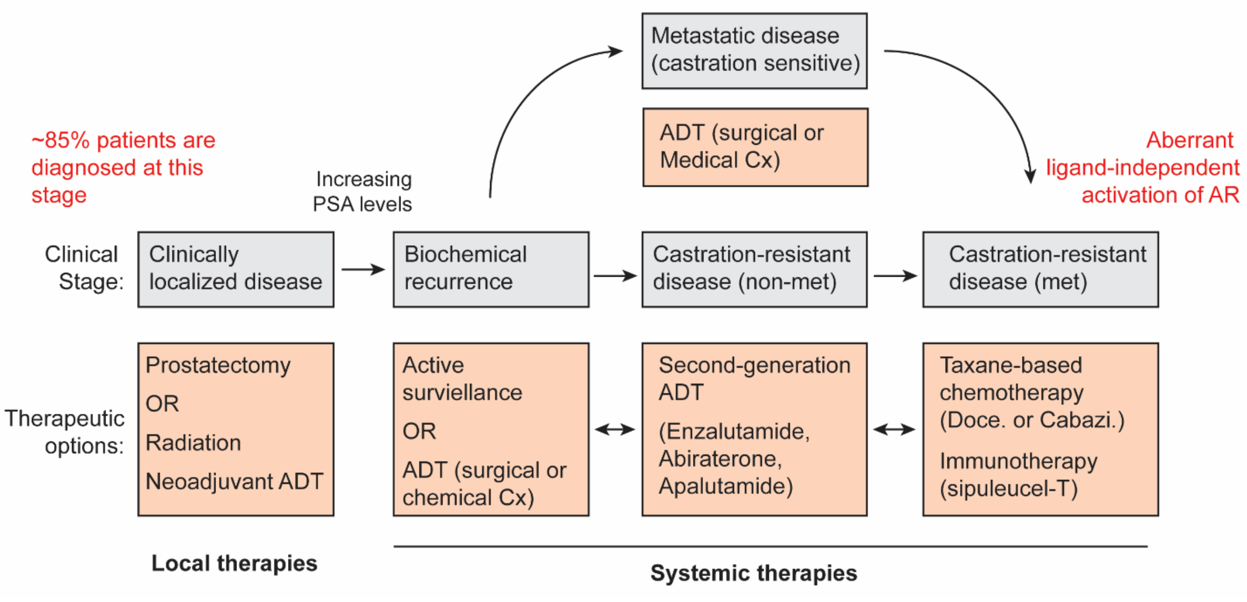


Figure 1-2: Clinical workflow for AR-driven prostate cancer treatment. Patients with histologically aggressive disease are initially treated with local therapies like surgery or radiation. Relapsed patients receive systemic therapies like pharmacologic that target the androgen/AR signaling axis or various taxane-based chemotherapeutics. Double-headed arrows indicate possible early combination of therapies. PSA, prostate specific antigen; Cx, Castration; ADT, androgen deprivation therapy; Doce, docetaxel; Cabazi, Cabazitaxel; AR, androgen receptor. This figure has been adapted from Drake CG. Prostate cancer as a model for tumour immunotherapy. Nat Rev Immunol. 2010;10: 580–593.

Prostate cancer patients on active surveillance are periodically monitored by imaging tests, prostate biopsies, and measurement of serum concentration of prostate-specific antigen (PSA)—a bona fide AR up-regulated gene [44]—which is used as a surrogate marker for disease burden [45]. Among men treated with prostatectomy or radiation therapy (i.e. local therapies), about 20-30% present with androgen-dependent biochemical recurrence of the tumor with a sudden elevation of serum PSA levels. In patients treated with radical prostatectomy, biochemical recurrence often correlates with metastatic spread of the disease to secondary sites [18,46]. At this point, the patients undergo either surgical or medical castration to block androgen production from the testes as the first-line of ADT [18,46,47]. Medical castration is a mechanism-informed approach to androgen ablation, wherein pharmacologic agonists or antagonists of the luteinizing hormone-releasing hormone (LHRH)-receptor are administered to inhibit the pituitary production of luteinizing hormone (LH) that signals the testes to produce androgen [48,49].

Castration, however, triggers only a transient remission of the tumor and almost all patients eventually progress to develop an androgen-independent disease—clinically referred to as castration-resistant prostate cancer (CRPC). The relapsed disease is treated with direct AR antagonists, namely enzalutamide [50–52] or apalutamide [53,54], that out-compete DHT by blocking its binding pocket, or drugs that inhibit the production of residual testosterone from the adrenal gland, namely abiraterone [55,56]. Many of the patients are inherently resistant to the second line of ADTs, or it induces only a transient disease remission due to the acquisition of drug resistance. These advanced patients are treated with chemotherapy (using drugs such as docetaxel [57,58] and cabazitaxel [59]) as the last resort for therapy. In modern practice, first-line ADT is often upfront combined with AR antagonists or chemotherapeutics as several clinical studies report higher over survival rates for the combinatorial regimens. The taxane chemotherapeutics have been shown to inhibit AR signaling by drug-induced microtubule stabilization that suppresses nuclear translocation of AR [60]. The median overall survival of CRPC patients being treated with taxanes is 16-18 months [18,46,61], highlighting the clinical subtype of prostate cancer that is in dire need of new and effective treatment options.

In recent years, the wider clinical use of potent direct AR inhibitors more frequently forces an adaptive evolution of CRPC into an AR-independent state, where tumor cells trans-differentiates into the small cell or neuroendocrine lineage. Therapy-emergent neuroendocrine or the recently-defined "double-negative" (negative for both AR and neuroendocrine markers) prostate cancer cases are increasing in frequency (comprising more than 30% of advanced CRPC patients by some estimates [62–64]) and pose a major clinical challenge moving forward. Accordingly, many recent studies have genomically and transcriptomically profiled these tumors in search of new therapeutic targets. The work presented in this thesis primarily focuses on the molecular and mechanistic characterization of AR-dependent prostate carcinogenesis and/or metastatic progression, and highlights novel molecular dependencies in this disease subtype that can be therapeutically exploited.

Somatic alterations in the prostate cancer genome

Cancer is a genetic disease—that is, it is caused by alterations in genes encoding proteins that govern important cellular functions such as survival, DNA replication, and cell division [65]. Loss of the *NKX3.1* gene on chromosome 8p21 was amongst the first and the most common genetic events associated with primary prostate cancer [66–68]. Subsequent studies revealed a high recurrence of *PTEN* [69,70] and rarer *TP53* [71,72] deletions in primary prostate cancer, both of which were also detected in other solid malignancies. Several studies also reported the recurrent copy-amplification of the *MYC* oncogene coded at chromosome 8q24 [68,73–75]. However, the most impactful discovery pre-dating the genomics era was chromosomal rearrangements that fuse an androgen-responsive promoter (the 5’ partner) to an ETS-domain family gene (the 3’ partner) in over 50% of primary prostate cancer [76,77]. This made ETS gene rearrangements the most common, prostate cancer-specific aberrations and the first report of gene fusions driving transformation in solid human malignancies driving the majority of cancer morbidity and mortality [78,79].

In the genomic era, these gene loci-focused studies were followed by genome-wide profiling of prostate cancer with the use of high-throughput sequencing technologies. These studies yielded unprecedented insights into the genetic and molecular makeup of human cancers. They uncovered recurrent structural and mutational alterations—both germline and somatic—that likely causally contribute to prostate tumorigenesis (**Figure 1-3**). The first whole-genome and exome sequencing studies revealed a higher recurrence of structural variations and chromosomal rearrangements as opposed to somatic base mutations in prostate cancer genomes [68,80–83]. Notably, the breakpoints of recurrent structural rearrangements (such as translocations or duplications) clustered in close proximity to AR binding sites, with some reports even suggesting a driver role for AR in the genesis of the *TMPRSS2-ERG* gene fusion [81,84]. One of these seminal studies also reported a novel category of rearrangements involving multiple loci to be commonly found in prostate cancer genomes—which they termed “chromoplexy” [85,86]. Altogether, these studies confirmed and provided accurate frequencies of copy-number loss and/or deleterious mutations in tumor suppressor genes such as *NKX3.1, PTEN* (predominantly monoallelic), *CHD1, TP53,* and *RB1,* and copy-number amplification of the *MYC, CCND1,* and *PIK3CA* oncogenes in primary prostate cancer. They also confirmed the recurrence of ETS-fusion in over half of the primary cases, while uncovering novel somatic mutations in genes such as *SPOP*, *FOXA1*, *NCOA2*, *NCOR2*, and *MED12*.

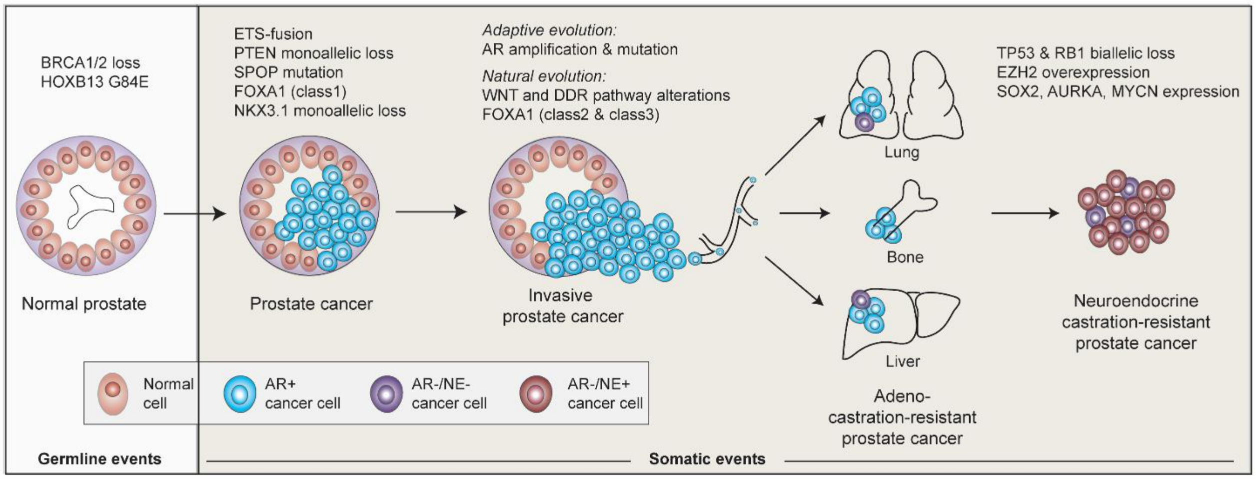


Figure 1-3: Evolutionary trajectory of the prostate cancer genome. Prostate cancer genomes are dominated by structural rearrangements over single nucleotide variations in the initial stages. Major germline and somatic genomic drivers of the disease are identified as per their clonal detection in distinct clinical and molecular subtypes of prostate cancer. AR+, androgen receptor positive; NE+, neuroendocrine positive.

The Cancer Genome Atlas (TCGA) program—a joint effort between the National Cancer Institute and the National Human Genome Research Institute—released the first comprehensive blueprint of the primary prostate cancer genome constructed from 333 patient specimens in 2015 [87]. This study identified the mutually exclusive genomic subtypes of primary prostate cancer that are characterized either by activating fusion in the ETS genes (namely, in order of frequency, *ERG*, *ETV1*, *ETV4*, and *FLI1*) or mutations in the *SPOP*, *FOXA1*, or *IDH1* genes. But, these categories accounted for only 74% of the primary tumors. The remaining 26% of the cases harbored no known subtype-defining single gene driver alterations, suggesting these to be driven by yet-to-be-uncovered molecular mechanisms [87]. A recent study explained about 2-3% of these primary cases to be independently initiated by biallelic inactivation of the *CDK12* gene [88], which is a key regulator of transcriptional elongation [89–91]. Notably, these tumors harbored a very distinct genomic signature characterized by the genome-wide accumulation of focal tandem duplications [88], which were earlier also reported in the CDK12-deficient ovarian cancers [92]. The remaining unclassified tumors showed low single-digit recurrence of alterations in epigenetic modifier genes (such as *KMT2C/MLL3*, *KMT2D/MLL2*, etc.), components of the DNA damage response machinery (such as *ATM, BRCA2,* etc.), or outlier *SPINK1* expression—all of which still remain mechanistically under investigation [93,94]. These efforts were recently supplemented by genomic profiling of non-indolent primary prostate cancer, where the aggressive tumors distinctively showed a higher recurrence of structural variations within non-coding genomic regions [95–97]. These included large-scale genomic rearrangements and inversions that results in transcriptional attenuation of the *cis*-neighborhood [96].

It is possible that some of the unclassified tumors may stem from germline events that predispose to prostate carcinogenesis—some of which may still be unknown. Like in other cancers, *BRCA2* inactivating alterations are associated with hereditary prostate cancer [98–100]. Also, the TCGA genomic profiles do not report the presence of prostate cancer-predisposing polymorphisms, such as the *HOXB13 G84E* population variant [101]. The G84E polymorphism in the *HOXB13* gene has been shown to result in an earlier onset of the disease [102,103]. Notably, 5-10% of prostate cancer diagnoses are associated with a familial history of the disease and familial cases comprise up to 30-40% of the early-onset disease [104]. Close relatives of prostate cancer patients are also twice as likely to develop the disease in their lifetime [104,105]. Thus, hereditary prostate cancer could be associated with a distinct set of genomic drivers that likely remain unclassified.

The most surprising finding from localized, primary prostate cancer genomes in my opinion was the stark absence of recurrent alterations in the *AR* gene itself or its genomic loci. As detailed above, owing to a serendipitous clinical observation in the early 1940s, the androgen/AR signaling axis has been at the center of all targeted therapies in prostate cancer following surgical resection or radiational shrinkage of the primary disease. Although the tumor initially responds to AR-targeted therapeutics, there is an inevitable onset of resistance to these therapies [18,61]. Genomic profiles of these resistant tumors further reinforced the idea of AR-addiction in prostate cancer survival and maintenance [2,3,6,106]. Upwards of 70-80% of the CRPC cases harbor copy-amplification of the *AR* gene, amplification of *AR* super-enhancers, activating *AR* coding mutations (e.g., T878A, H875Y, F877L, etc.) that render AR antagonists to instead serve as agonists, *AR* coding mutations that enable promiscuous activation by non-canonical steroids, expression of ligand-independent AR splice-variants (e.g. *AR-V7*), and/or AR compensation by aberrant expression of other nuclear receptors (e.g. glucocorticoid receptor) [18,61,106–113]. Altogether, coding and non-coding alterations in the *AR* gene itself or AR cofactor genes (namely *FOXA1, NCOR1/2, NCOA, ZBTB16,* etc.) are by far the most common acquired alterations in hormone-refractory prostate cancers [6,106].

The metastatic castration-resistant disease also shows a considerable increase in the recurrence of biallelic *TP53, RB1, CDK12,* and *PTEN* deletions [114], enrichment for *FOXA1 and KMT2D* alterations, and acquisition of novel genomic mutations in *GNAS, PIK3CA/B, PIK3R1,* and ancillary cancer pathways such as WNT (e.g. *CTNNB1* activating mutations, *APC* deletion, etc.) and DNA damage response (*ATM* inactivating mutations, and *BRCA1/2* deletions) [6]. Notably, both WNT and DNA repair response pathway genes are altered in more than 20% of the CRPC cases [6,114], which is remarkably higher than their recurrence in the primary disease. These alterations are most likely associated with disease aggressiveness and metastatic propensity. *CDK12* loss-of-function alterations (including copy deletions, frameshifting mutations, missense mutations within the kinase domain) are considerably increased in CRPC cases with a parallel increase in the accumulation of focal tandem duplications across the genome [88,115]—the mechanistic basis of which still remains unknown and is profoundly intriguing to study. The CDK12-deficient tumors are predicted to have a higher neo-antigen burden and thus are more responsive to immune checkpoint blockade [88,116]. In this respect, CDK12-aberrant tumors are similar to the tumors that harbor mismatch repair deficiencies: While the average mutation rate for CRPC is 4.4 mutations/Mb, cases that harbor alterations in mismatch repair genes, such as *MLH1*, *MSH2,* or *MSH6*, showed a mutation rate of nearly 50 per Mb [2,6]. This results in microsatellite instability (MSI) in these tumors that in turn imparts higher immunogenicity. Interestingly, of all initiating alterations, *SPOP* coding mutations are the only events that are depleted in the CRPC tumors [114], suggesting a better prognosis for this genomic subtype of prostate cancer upon treatment with the first line of ADTs.

The subsequent use of second-generation, potent AR antagonists in advanced prostate cancer coerces an adaptive evolution of the tumor genome, wherein tumor cells can completely lose detectable expression of AR and its cofactors (such as FOXA1, HOXB13, etc.), and transdifferentiate into the small cell or neuroendocrine lineage. The umbrella term "lineage plasticity" is often used to capture this concept in the cancer biology field [117,118]. Neuroendocrine prostate cancer is clinically identified via distinct histopathologic features and expression of neuroendocrine lineage markers, such as synaptophysin (SYP) and chromogranin A (CHGA) [62,63,119,120]. These tumors are genetically very similar to CRPC and are primarily thought to evolve through epigenetic mechanisms (**Figure 1-4**). In comparison to CRPC, neuroendocrine prostate cancers show only an increased frequency of *RB1* and *TP53* co-deletions [62,121,122], and low recurrence of the basal-phenotype promoting *FOXA1 R219H* mutations [123]. However, the expression of the FOXA1 transcript or protein (wild-type or mutant) in the neuroendocrine disease is yet to be compellingly demonstrated. The neuroendocrine disease is primarily thought to involve a dramatic reprogramming of the epigenetic landscape and transcriptional gene networks, with outlier expression of *SOX2, MYCN, AURKA, EZH2,* and *SPDEF* oncogenes [120,122,124–126]. The molecular mechanism of the trans-differentiation of prostate adenocarcinoma into a neuroendocrine/small cell disease is very much a subject of active research, which will hopefully uncover novel targets to impede this lethal progression of the disease.

Despite the sudden burst in genomic information from prostate cancer genomes from all its clinical subtypes, outside of BRCA2 and MSI status, these findings are yet to be meaningfully translated into the clinic. This is largely due to the lack of mechanistic insights and/or identification of druggable dependencies that distinct genomic alterations engender in tumor cells. Thus, functional and mechanistic annotations of genomic drivers in human cancers should be the primary focus of experimental/molecular biologists in the years to follow.

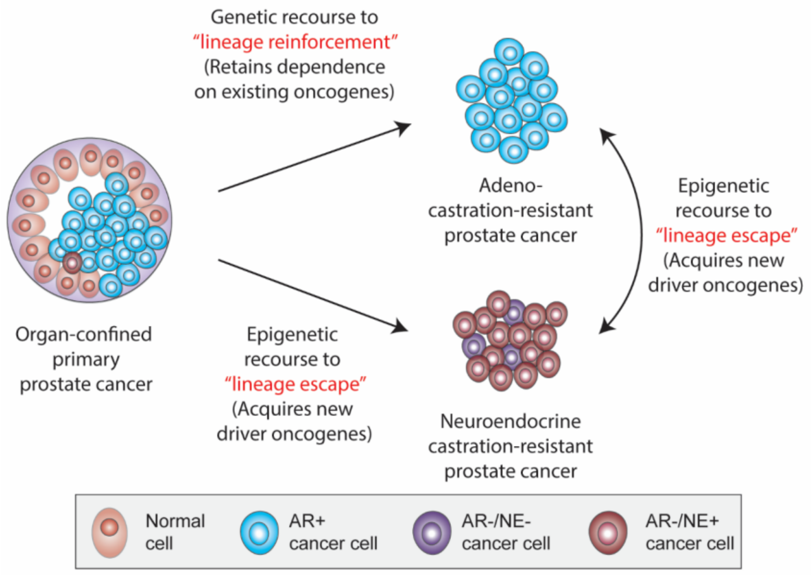


Figure 1-4: Major routes of prostate cancer progression. Upon sustained inhibition of lineage-specific AR activity, advanced prostate cancer progresses to acquire resistance either via genetic alterations to reinforce the lineage transcriptional network or epigenetic reprogramming to escape the prostatic lineage along with the acquisition of novel oncogenic drivers.

The centrality of AR signaling in prostate cancer pathobiology

As described above, AR is the principal mediator of prostate tumorigenesis and progression. Upon pharmacologic inhibition of AR activity, prostate cancer cells evolve resistant mechanisms that aberrantly reinstate the transcriptional competence of AR through convergent genetic alterations. AR is a type I nuclear receptor that in absence of its steroid hormone ligand (i.e., testosterone or its active cytosolic variant, called dihydrotestosterone) remains in the cytosol in a transcriptionally inactive conformation [127]. Androgens bind to the C-terminal ligand-binding domain which is connected via a nuclear translocation sequence-containing hinge region to the DNA-binding and transactional domains [128,129]. Upon ligand binding, AR dissociates from cytosolic heat shock proteins, self-dimerizes, and instantly translocates into the nucleus, where it binds to genomic sites that harbor the 15-bp palindromic androgen-responsive element (consensus sequence: 5' AGAACAnnnTGTTCT 3') [130]. The genome-wide profile of AR chromatin binding (aka the AR cistrome) shows the majority of AR sites to lie within intergenic or non-coding genomic regions that are enriched for enhancer elements [131]. Modern proteomics assays have revealed the AR transcriptional complex to comprise hundreds of epigenetic and chromatin remodeling co-factors that work in concert to activate the expression of genes that instruct the prostatic lineage identity and normal cellular function [132–134].

Experimental studies over the years have revealed the primary drivers of prostate cancer to rely on and/or repurpose AR signaling to drive oncogenesis. This was firmly corroborated by the first profiles of the AR cistrome from primary prostate tumors and matched normal prostate tissues, which showed an augmentation and extensive redistribution of AR chromatin occupancy in tumor cells vs the normal prostate tissue [131,135]. Notably, consensus motifs of FOXA1 and HOXB13 were enriched at cancer-specific binding sites of AR, and ectopic overexpression of these lineage-specific AR cofactors in the normal prostate cells confirmed their causal role in neoplastic reprogramming of the AR cistrome [131]. *FOXA1* is over-expressed and recurrently mutated in primary and metastatic prostate cancer [2,82,87], while *HOXB13* harbors predisposing, high-penetrance polymorphisms that are associated with the early onset of prostate cancer [101,102]. The highly-recurrent *TMPRSS2-ERG* gene fusion in prostate cancer directly relies on AR activity for its expression [76]. Some studies implicate AR-signaling in the genesis of these gene fusions, where liganded AR facilitates a bridging event between the *TMPRSS2* and ETS-gene loci to enable DNA recombination [84,136]. Furthermore, the ERG oncoprotein functions as a transcription factor and recruits AR to a distinct subset of binding sites to activate oncogenic gene programs [137,138]. ERG also co-opts FOXA1 and HOXB13 for activation of new *cis*-regulatory elements and aberrant gene transcription [139]. Thus, the ERG oncoprotein works in close collaboration with AR and its transcriptional cofactors to drive the malignant phenotype.

Recurrent point mutations in *SPOP* (at residues Y87, F102, Y123, W131, and F133) define a distinct molecular subclass of prostate cancer. SPOP is an E3 ubiquitin ligase adaptor that mediates the ubiquitination of target proteins, leading to their proteasomal degradation [140]. All missense mutations recur within the MATH domain of the SPOP protein, which disrupts its recognition of AR [141,142] and other essential AR co-activators (like BET proteins [143,144], TRIM24/28 [145,146], etc.) as substrates. This prevents the ubiquitin-mediated proteasomal degradation of these proteins and thus increases their activity in cancer cells. The recurrent inactivation of CHD1, a chromatin remodeler, alters chromatin accessibility and redistributes AR on the chromatin to occupy cancer-specific enhancer elements [147]. This activates a distinct AR gene program that drives the transformation of prostate epithelial cells [147]. *CHD1* is a prostate cancer-specific tumor suppressor gene and is often co-altered along with the *SPOP* gene [147,148], thus the two oncogenes collaboratively work to reprogram AR activity during transformation. Loss of the prostate lineage-specific cofactor NKX3.1 disables its AR-repressive functions, in addition to having other oncogenic effects, to drive prostate carcinogenesis [67,149]. That said, the functional effect of *NKX3.1* deletion and/or mutations in prostate cancer still remains to be fully understood. In a good fraction of the primary tumors, only one allele of the NKX3.1 gene is deleted, there are recurrent base substitutions within the DNA-binding homeodomain (e.g. Y177C), and copy-neutral loss of heterozygosity—all of which remain mechanistically uncharacterized (personal knowledge).

In summary, prostate carcinogenesis is molecularly constructed around the lineage-specific AR transcriptional complex, often involving genomic alterations in key AR cofactors, thus explaining the acute addiction of malignant cells to robust and sustained AR activity. Adding to this knowledge base, this thesis offers a novel scheme for structural classification of the highly-recurrent FOXA1 alterations, mechanistically delineates their class-specific pathobiology in prostate carcinogenesis and metastatic progression, and identifies the chromatin-remodeling SWI/SNF complex as a novel, targetable dependency in AR/FOXA1-driven prostate cancers.

Pioneer factors instruct lineage identity by demarcating physical chromatin accessibility at *cis*-regulatory elements

Years of research focused on animal embryogenesis and organogenesis has revealed a pivotal role for chromatin structure in cellular differentiation (i.e. stepwise specification of a pluripotent stem cell into more specialized cell types) and terminal cell identity [150,151]. In mammalian cells, the nucleosome is a basic structural unit of the chromatin and comprises DNA (147 bp long) wrapped around a histone octamer [152–155]. It is the physical positioning as well as the chemical modification on histone tails of nucleosomes that determine the functional competence of specific DNA sequences [156–160]. Nascent chromatin is inaccessible to transcription factors and this is transcriptionally silent [161,162]. Thus chromatin accessibility has emerged as a primary hallmark of functionally-competent DNA, particularly within the non-coding genomic regions [163–165]. In this context, profiling of the accessible chromatin landscape across tissue types and the identification of its primary regulators has become a central focus to fully understand the lineage-specificity of gene transcription.

An emerging class of chromatin regulators, known as the pioneer transcription factors, can independently bind to their DNA targets within compacted chromatin and subsequently displace nucleosomes and linker histones (e.g. H1) to make the DNA more accessible [166–168]. As per the current transcriptional model, pioneer factors first bind to specific regions in the chromatin and facilitate the binding of prototypical, non-pioneer transcription factors [168]. A bona fide pioneer transcription factor must: 1) engage its target sites in closed, inactive chromatin prior to gene activity, and 2) increase accessibility of a target site upon binding, thereby recruiting other regulatory proteins (e.g., transcription factors, chromatin remodelers, chromatin modifiers, histone variants, and repressors) to that site [169]. Thus, pioneer factors play a major role in cell programming and cell fate determination [169–171], and, not surprisingly, known pioneer factor genes show a very lineage-specific expression pattern (e.g., *FOXA1*, *GATA3*, etc.). Pioneer factors are able to access and recognize their target sites on nucleosomal DNA as they harbor specialized DNA-binding domains. This includes the "winged" helix-turn-helix (found in the FOX [172,173] and NF-Y [174,175] pioneer factor families) and zinc-finger (found in the Groucho/TLE [176] and GATA [177] pioneer factors) DNA-binding domains. In the normal prostate epithelia, the pioneer factor FOXA1 de-compacts chromatin at *cis*-regulatory elements and allows AR to bind to and activate the expression of prostatic genes [178–182]. Intriguingly, FOXA1 is recurrently mutated in 12-13% of metastatic CRPC patients, showing 2-3 fold enrichment over the primary disease [6]. At the time I started my Ph.D., the functional implication of FOXA1 mutations remained heavily contested with contradictory evidence projecting FOXA1 as an oncogene as well as a tumor suppressor. This thesis summarizes the hypermorphic or neomorphic functions of distinct classes of FOXA1 alterations in advanced prostate cancer—unambiguously establishing FOXA1 as a principal oncogene in all stages of AR-dependent disease.

ATP-dependent chromatin remodelers mediate dynamic regulation of chromatin accessibility

While pioneer factors passively change chromatin accessibility, multimeric chromatin remodelers actively utilize energy from ATP-hydrolysis to make the DNA more accessible [183]. These enzymes share the highly-conserved ATPase catalytic domain but harbor distinct sets of epigenetic reader domains in the flanking regions. Based on the presence of reader domains, chromatin remodeling complexes are split into four families, namely SWI/SNF (also known as BRG1/BRM-associated factor complexes (BAF)), CHD, ISWI, and INO80 [184–187]. These chromatin remodelers are responsible for modulating nucleosomal positioning at specific regions in the genome and work in concert with the core transcriptional machinery to dynamically regulate gene expression [188,189].

Of these, the SWI/SNF family complexes have come to the forefront of mechanistic interrogations as its constituent subunits are altered in about 20-25% of human cancers (e.g. PBRM1, ARID1A, or SMARCA4) [190–192]. The SWI/SNF remodelers are defined by the presence of two functional domains, 1) an N-terminally located helicase-SANT (HSA) domain that is known to recruit actin and actin-related proteins, and 2) a C-terminally located bromodomain that is known to bind to the acetylated-lysines on histone tails [193,194]. In mammalian cells, SWI/SNF complexes are 1-1.5-MDa in size and sequentially assemble from 29 subunits that comprise the core, secondary, or ATPase modules [195–197]. Primarily dictated by the composition of the secondary module, the SWI/SNF complexes exist in three non-redundant final form assemblies:

the canonical SWI/SNF (exclusively comprises ARID1A/B and DPF1/2/3), the polybromo-associated SWI/SNF (exclusively comprises ARID2, BRD7, and PBRM1), and the non-canonical SWI/SNF (exclusively comprises GLTSCR1/1L and BRD9) [195,198,199]. Many of the SWI/SNF subunits have multiple paralogs in human cells that lead to extensive compositional diversity of end complexes—each speculated to mediate distinct cellular functions [200]. There are two paralogs of the SWI/SNF ATPase subunit, SMARCA2 (also known as BRM) and SMARCA4 (also known as BRG1)—only one of which assembles into each of the SWI/SNF complexes [201]. These ATPases can functionally compensate for each other and the ATPase activity is absolutely essential for the chromatin remodeling ability of SWI/SNF complexes [197].

Although yet to be mechanistically explained, different SWI/SNF complexes are recruited to specific genomic sites where they help displace or eject nucleosomes to make the underlying DNA functionally accessible [202]. In SWI/SNF-aberrant tumors, the residual complex is thought to reshape the chromatin structure and enable oncogenic gene programs [203,204]. In synovial sarcomas, the SWI/SNF subunit *SS18* is fused to the *SSX* gene resulting in a neomorphic hybrid protein that alters the chromatin recruitment of the SWI/SNF complex [205,206]. Other recurrent alterations in SWI/SNF subunits in human cancers, such as *SMARCA4* and *PBRM1* deletions, are largely thought to reconfigure the composition of the SWI/SNF complexes that alters their interaction with transcription and other chromatin regulators [191,195]. Thus, there is mounting interest in exploring therapeutic targetability of SWI/SNF activity in human cancers; however, these efforts have been thwarted by the lack of potent pharmacologic agents. In this thesis, we share a novel proteolysis targeting chimera (PROTAC) of the SWI/SNF ATPases, SMARCA2 and SMARCA2, that show exquisite and preferential cytotoxicity in enhancer-binding transcription factor-driven prostate and plasma cell tumors.

Chapter acknowledgements:

This chapter was written by Abhijit Parolia. It elaborates on various scientific ideas and concepts that in the past have been discussed with Dr. Arul Chinnaiyan in individual meetings. Figures 1-1, 1-3, and 1-4 were originally drawn for this thesis. Figure 1-2 was adapted form a similar summary figure in Drake CG. Prostate cancer as a model for tumor immunotherapy. Nat Rev Immunol. 2010;10: 580–593.

References:

1. Franco I, Helgadottir HT, Moggio A, Larsson M, Vrtačnik P, Johansson A, et al. Whole genome DNA sequencing provides an atlas of somatic mutagenesis in healthy human cells and identifies a tumor-prone cell type. Genome Biol. 2019;20: 285.
2. Robinson DR, Wu Y-M, Lonigro RJ, Vats P, Cobain E, Everett J, et al. Integrative clinical genomics of metastatic cancer. Nature. 2017;548: 297–303.
3. Zehir A, Benayed R, Shah RH, Syed A, Middha S, Kim HR, et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. Nat Med. 2017;23: 703–713.
4. Cancer Genome Atlas Research Network. Electronic address: andrew_aguirre@dfci.harvard.edu, Cancer Genome Atlas Research Network. Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma. Cancer Cell. 2017;32: 185–203.e13.
5. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. Nature. 2012;487: 330–337.
6. Robinson D, Van Allen EM, Wu Y-M, Schultz N, Lonigro RJ, Mosquera J-M, et al. Integrative clinical genomics of advanced prostate cancer. Cell. 2015;161: 1215–1228.
7. Ciriello G, Gatza ML, Beck AH, Wilkerson MD, Rhie SK, Pastore A, et al. Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. Cell. 2015;163: 506–519.
8. Rheinbay E, Parasuraman P, Grimsby J, Tiao G, Engreitz JM, Kim J, et al. Recurrent and functional regulatory mutations in breast cancer. Nature. 2017;547: 55–60.
9. Bartram CR, de Klein A, Hagemeijer A, van Agthoven T, Geurts van Kessel A, Bootsma D, et al. Translocation of c-ab1 oncogene correlates with the presence of a Philadelphia chromosome in chronic myelocytic leukaemia. Nature. 1983;306: 277–280.
10. Groffen J, Stephenson JR, Heisterkamp N, de Klein A, Bartram CR, Grosveld G. Philadelphia chromosomal breakpoints are clustered within a limited region, bcr, on chromosome 22. Cell. 1984;36: 93–99.
11. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin. 2020;70: 7–30.
12. Kelly SP, Anderson WF, Rosenberg PS, Cook MB. Past, Current, and Future Incidence Rates and Burden of Metastatic Prostate Cancer in the United States. Eur Urol Focus. 2018;4: 121–127.
13. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA Cancer J Clin. 2019;69: 7–34.

14. Simon S. Facts & figures 2020 reports largest one-year drop in cancer mortality. In: American Cancer Society [Internet]. 8 Jan 2020 [cited 6 Feb 2021]. Available: https://www.cancer.org/latest-news/facts-and-figures-2020.html
15. Tests to Diagnose and Stage Prostate Cancer. [cited 6 Feb 2021]. Available: https://www.cancer.org/cancer/prostate-cancer/detection-diagnosis-staging/how-diagnosed.html
16. Huggins C, Stevens RE, Hodges CV. STUDIES ON PROSTATIC CANCER: II. THE EFFECTS OF CASTRATION ON ADVANCED CARCINOMA OF THE PROSTATE GLAND. Arch Surg. 1941;43: 209–223.
17. Huggins C. EFFECT OF ORCHIECTOMY AND IRRADIATION ON CANCER OF THE PROSTATE. Ann Surg. 1942;115: 1192–1200.
18. Mills IG. Maintaining and reprogramming genomic androgen receptor activity in prostate cancer. Nat Rev Cancer. 2014;14: 187–198.
19. The Nobel Prize in Physiology or Medicine 1966. [cited 8 Feb 2021]. Available: https://www.nobelprize.org/prizes/medicine/1966/summary/
20. Mulder E, Vrij AA, Brinkmann AO. DNA and ribonucleotide binding characteristics of two forms of the androgen receptor from rat prostates. Biochem Biophys Res Commun. 1983;114: 1147–1153.
21. Rice MA, Malhotra SV, Stoyanova T. Second-Generation Antiandrogens: From Discovery to Standard of Care in Castration Resistant Prostate Cancer. Front Oncol. 2019;9: 801.
22. Desireddi NV, Roehl KA, Loeb S, Yu X, Griffin CR, Kundu SK, et al. Improved stage and grade-specific progression-free survival rates after radical prostatectomy in the PSA era. Urology. 2007;70: 950–955.
23. Dong F, Reuther AM, Magi-Galluzzi C, Zhou M, Kupelian PA, Klein EA. Pathologic stage migration has slowed in the late PSA era. Urology. 2007;70: 839–842.
24. Caire AA, Sun L, Lack BD, Lum K, Tang P, Stackhouse DA, et al. Predicting non-organ-confined prostate cancer in men diagnosed after 2000. Prostate Cancer Prostatic Dis. 2010;13: 248–251.
25. Fortuin A, Rooij M de, Zamecnik P, Haberkorn U, Barentsz J. Molecular and functional imaging for detection of lymph node metastases in prostate cancer. Int J Mol Sci. 2013;14: 13842–13875.
26. Office of the Commissioner. FDA approves first PSMA-targeted PET imaging drug for men with prostate cancer. 12 Jan 2020 [cited 6 Feb 2021]. Available: https://www.fda.gov/news-events/press-announcements/fda-approves-first-psma-targeted-pet-imaging-drug-men-prostate-cancer
27. Epstein JI, Egevad L, Amin MB, Delahunt B, Srigley JR, Humphrey PA, et al. The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma: Definition of Grading Patterns and Proposal for a New Grading System. Am J Surg Pathol. 2016;40: 244–252.
28. Humphrey PA. Gleason grading and prognostic factors in carcinoma of the prostate. Mod Pathol. 2004;17: 292–306.
29. Koukourakis G, Kelekis N, Armonis V, Kouloulias V. Brachytherapy for prostate cancer: a systematic review. Adv Urol. 2009; 327945.
30. Yap TA, Zivi A, Omlin A, de Bono JS. The changing therapeutic landscape of castration-resistant prostate cancer. Nat Rev Clin Oncol. 2011;8: 597–610.
31. Morash C, Tey R, Agbassi C, Klotz L, McGowan T, Srigley J, et al. Active surveillance for

the management of localized prostate cancer: Guideline recommendations. Can Urol Assoc J. 2015;9: 171–178.

32. Sowalsky AG. Following bloody footprints: Cancer diagnosis from cell-free DNA. Sci Transl Med. 2019;11. doi:10.1126/scitranslmed.aaz3724
33. Seyedolmohadessin SM, Akbari MT, Nourmohammadi Z, Basiri A, Pourmand G. Assessing the Diagnostic Value of Plasma-Free DNA in Prostate Cancer Screening. Iran Biomed J. 2018;22: 331–337.
34. Casadio V, Calistri D, Salvi S, Gunelli R, Carretta E, Amadori D, et al. Urine Cell-Free DNA Integrity as a Marker for Early Prostate Cancer Diagnosis: A Pilot Study. Biomed Res Int. 2013;2013. doi:10.1155/2013/270457
35. Cario CL, Chen E, Leong L, Emami NC, Lopez K, Tenggara I, et al. A machine learning approach to optimizing cell-free DNA sequencing panels: with an application to prostate cancer. BMC Cancer. 2020;20: 820.
36. Ulz P, Belic J, Graf R, Auer M, Lafer I, Fischereder K, et al. Whole-genome plasma sequencing reveals focal amplifications as a driving force in metastatic prostate cancer. Nat Commun. 2016;7: 12008.
37. Iyer MK, Niknafs YS, Malik R, Singhal U, Sahu A, Hosono Y, et al. The landscape of long noncoding RNAs in the human transcriptome. Nat Genet. 2015;47: 199–208.
38. Mehra R, Udager AM, Ahearn TU, Cao X, Feng FY, Loda M, et al. Overexpression of the Long Non-coding RNA SChLAP1 Independently Predicts Lethal Prostate Cancer. Eur Urol. 2016;70: 549–552.
39. Mehra R, Shi Y, Udager AM, Prensner JR, Sahu A, Iyer MK, et al. A novel RNA in situ hybridization assay for the long noncoding RNA SChLAP1 predicts poor clinical outcome after radical prostatectomy in clinically localized prostate cancer. Neoplasia. 2014;16: 1121–1127.
40. Xia Q, Ding T, Zhang G, Li Z, Zeng L, Zhu Y, et al. Circular RNA Expression Profiling Identifies Prostate Cancer- Specific circRNAs in Prostate Cancer. Cell Physiol Biochem. 2018;50: 1903–1915.
41. Vo JN, Cieslik M, Zhang Y, Shukla S, Xiao L, Zhang Y, et al. The Landscape of Circular RNA in Cancer. Cell. 2019;176: 869–881.e13.
42. Tucker D, Zheng W, Zhang D-H, Dong X. Circular RNA and its potential as prostate cancer biomarkers. World J Clin Oncol. 2020;11: 563–572.
43. Litwin MS, Tan H-J. The Diagnosis and Treatment of Prostate Cancer: A Review. JAMA. 2017;317: 2532–2542.
44. Kim J, Coetzee GA. Prostate specific antigen gene regulation by androgen receptor. J Cell Biochem. 2004;93: 233–241.
45. Prostate-Specific Antigen (PSA) Test. 6 Sep 2017 [cited 6 Feb 2021]. Available: https://www.cancer.gov/types/prostate/psa-fact-sheet
46. Imamura Y, Sadar MD. Androgen receptor targeted therapies in castration-resistant prostate cancer: Bench to clinic. Int J Urol. 2016;23: 654–665.
47. Kachur E, Hematology/Oncology P, BMT Pharmacist Clinical Coordinator Atrium Health, Levine Cancer Institute Charlotte, Carolina N. Prostate Cancer Review. [cited 6 Feb 2021]. Available: https://www.uspharmacist.com/article/prostate-cancer-review
48. Freckman HA. Chemical Castration. JAMA. 1964;189. doi:10.1001/jama.1964.03070090050025
49. Weiner AB, Cohen JE, DeLancey JO, Schaeffer EM, Auffenberg GB. Surgical versus

Medical Castration for Metastatic Prostate Cancer: Use and Overall Survival in a National Cohort. J Urol. 2020;203: 933–939.

50. Tran C, Ouk S, Clegg NJ, Chen Y, Watson PA, Arora V, et al. Development of a second-generation antiandrogen for treatment of advanced prostate cancer. Science. 2009;324: 787–790.
51. Scher HI, Fizazi K, Saad F, Taplin M-E, Sternberg CN, Miller K, et al. Increased survival with enzalutamide in prostate cancer after chemotherapy. N Engl J Med. 2012;367: 1187–1197.
52. Beer TM, Armstrong AJ, Rathkopf DE, Loriot Y, Sternberg CN, Higano CS, et al. Enzalutamide in metastatic prostate cancer before chemotherapy. N Engl J Med. 2014;371: 424–433.
53. Clegg NJ, Wongvipat J, Joseph JD, Tran C, Ouk S, Dilhas A, et al. ARN-509: a novel antiandrogen for prostate cancer treatment. Cancer Res. 2012;72: 1494–1503.
54. Koshkin VS, Small EJ. Apalutamide in the treatment of castrate-resistant prostate cancer: evidence from clinical trials. Ther Adv Urol. 2018;10: 445–454.
55. Fizazi K, Scher HI, Molina A, Logothetis CJ, Chi KN, Jones RJ, et al. Abiraterone acetate for treatment of metastatic castration-resistant prostate cancer: final overall survival analysis of the COU-AA-301 randomised, double-blind, placebo-controlled phase 3 study. Lancet Oncol. 2012;13: 983–992.
56. Ryan CJ, Smith MR, Fizazi K, Saad F, Mulders PFA, Sternberg CN, et al. Abiraterone acetate plus prednisone versus placebo plus prednisone in chemotherapy-naive men with metastatic castration-resistant prostate cancer (COU-AA-302): final overall survival analysis of a randomised, double-blind, placebo-controlled phase 3 study. Lancet Oncol. 2015;16: 152–160.
57. Tannock IF, de Wit R, Berry WR, Horti J, Pluzanska A, Chi KN, et al. Docetaxel plus Prednisone or Mitoxantrone plus Prednisone for Advanced Prostate Cancer. N Engl J Med. 2004;351: 1502–1512.
58. Petrylak DP, Tangen CM, Hussain MHA, Lara PN Jr, Jones JA, Taplin ME, et al. Docetaxel and estramustine compared with mitoxantrone and prednisone for advanced refractory prostate cancer. N Engl J Med. 2004;351: 1513–1520.
59. de Bono JS, Oudard S, Ozguroglu M, Hansen S, Machiels J-P, Kocak I, et al. Prednisone plus cabazitaxel or mitoxantrone for metastatic castration-resistant prostate cancer progressing after docetaxel treatment: a randomised open-label trial. Lancet. 2010;376: 1147–1154.
60. Zhu M-L, Horbinski CM, Garzotto M, Qian DZ, Beer TM, Kyprianou N. Tubulin-targeting chemotherapy impairs androgen receptor activity in prostate cancer. Cancer Res. 2010;70: 7992–8002.
61. Watson PA, Arora VK, Sawyers CL. Emerging mechanisms of resistance to androgen receptor inhibitors in prostate cancer. Nat Rev Cancer. 2015;15: 701–711.
62. Beltran H, Prandi D, Mosquera JM, Benelli M, Puca L, Cyrta J, et al. Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. Nat Med. 2016;22: 298–305.
63. Vlachostergios PJ, Puca L, Beltran H. Emerging Variants of Castration-Resistant Prostate Cancer. Curr Oncol Rep. 2017;19: 32.
64. Bluemn EG, Coleman IM, Lucas JM, Coleman RT, Hernandez-Lopez S, Tharakan R, et al. Androgen Receptor Pathway-Independent Prostate Cancer Is Sustained through FGF Signaling. Cancer Cell. 2017;32: 474–489.e6.
65. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144: 646–

674.
66. He WW, Sciavolino PJ, Wing J, Augustus M, Hudson P, Meissner PS, et al. A novel human prostate-specific, androgen-regulated homeobox gene (NKX3.1) that maps to 8p21, a region frequently deleted in prostate cancer. Genomics. 1997;43: 69–77.
67. Bhatia-Gaur R, Donjacour AA, Sciavolino PJ, Kim M, Desai N, Young P, et al. Roles for Nkx3.1 in prostate development and cancer. Genes Dev. 1999;13: 966–977.
68. Lapointe J, Li C, Giacomini CP, Salari K, Huang S, Wang P, et al. Genomic profiling reveals alternative genetic pathways of prostate tumorigenesis. Cancer Res. 2007;67: 8504–8510.
69. Li J, Yen C, Liaw D, Podsypanina K, Bose S, Wang SI, et al. PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. Science. 1997;275: 1943–1947.
70. Cairns P, Okami K, Halachmi S, Halachmi N, Esteller M, Herman JG, et al. Frequent inactivation of PTEN/MMAC1 in primary prostate cancer. Cancer Res. 1997;57: 4997–5000.
71. Hall MC, Navone NM, Troncoso P, Pollack A, Zagars GK, von Eschenbach AC, et al. Frequency and characterization of p53 mutations in clinically localized prostate cancer. Urology. 1995;45: 470–475.
72. Kubota Y, Shuin T, Uemura H, Fujinami K, Miyamoto H, Torigoe S, et al. Tumor suppressor gene p53 mutations in human prostate cancer. Prostate. 1995;27: 18–24.
73. Ahmadiyeh N, Pomerantz MM, Grisanzio C, Herman P, Jia L, Almendro V, et al. 8q24 prostate, breast, and colon cancer risk loci show tissue-specific long-range interaction with MYC. Proc Natl Acad Sci U S A. 2010;107: 9742–9746.
74. Jenkins RB, Qian J, Lieber MM, Bostwick DG. Detection of c-myc oncogene amplification and chromosomal anomalies in metastatic prostatic carcinoma by fluorescence in situ hybridization. Cancer Res. 1997;57: 524–531.
75. Nag A, Smith RG. Amplification, rearrangement, and elevated expression of c-myc in the human prostatic carcinoma cell line LNCaP. Prostate. 1989;15: 115–122.
76. Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun X-W, et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science. 2005;310: 644–648.
77. Tomlins SA, Laxman B, Dhanasekaran SM, Helgeson BE, Cao X, Morris DS, et al. Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. Nature. 2007;448: 595.
78. Kumar-Sinha C, Tomlins SA, Chinnaiyan AM. Recurrent gene fusions in prostate cancer. Nat Rev Cancer. 2008;8: 497–511.
79. Tomlins SA, Mehra R, Rhodes DR, Cao X, Wang L, Dhanasekaran SM, et al. Integrative molecular concept modeling of prostate cancer progression. Nat Genet. 2007;39: 41–51.
80. Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y, Carver BS, et al. Integrative genomic profiling of human prostate cancer. Cancer Cell. 2010;18: 11–22.
81. Berger MF, Lawrence MS, Demichelis F, Drier Y, Cibulskis K, Sivachenko AY, et al. The genomic complexity of primary human prostate cancer. Nature. 2011;470: 214–220.
82. Barbieri CE, Baca SC, Lawrence MS, Demichelis F, Blattner M, Theurillat J-P, et al. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. Nat Genet. 2012;44: 685–689.
83. Pflueger D, Terry S, Sboner A, Habegger L, Esgueva R, Lin P-C, et al. Discovery of non-ETS gene fusions in human prostate cancer using next-generation RNA sequencing. Genome Res. 2011;21: 56–67.

84. Mani R-S, Tomlins SA, Callahan K, Ghosh A, Nyati MK, Varambally S, et al. Induced chromosomal proximity and gene fusions in prostate cancer. Science. 2009;326: 1230.
85. Shen MM. Chromoplexy: a new category of complex rearrangements in the cancer genome. Cancer cell. 2013. pp. 567–569.
86. Baca SC, Prandi D, Lawrence MS, Mosquera JM, Romanel A, Drier Y, et al. Punctuated evolution of prostate cancer genomes. Cell. 2013;153: 666–677.
87. Cancer Genome Atlas Research Network. The Molecular Taxonomy of Primary Prostate Cancer. Cell. 2015;163: 1011–1025.
88. Wu Y-M, Cieślik M, Lonigro RJ, Vats P, Reimers MA, Cao X, et al. Inactivation of CDK12 Delineates a Distinct Immunogenic Class of Advanced Prostate Cancer. Cell. 2018;173: 1770–1782.e14.
89. Bartkowiak B, Liu P, Phatnani HP, Fuda NJ, Cooper JJ, Price DH, et al. CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. Genes Dev. 2010;24: 2303–2316.
90. Krajewska M, Dries R, Grassetti AV, Dust S, Gao Y, Huang H, et al. CDK12 loss in cancer cells affects DNA damage response genes through premature cleavage and polyadenylation. Nat Commun. 2019;10: 1–16.
91. Tellier M, Zaborowska J, Caizzi L, Mohammad E, Velychko T, Schwalb B, et al. CDK12 globally stimulates RNA polymerase II transcription elongation and carboxyl-terminal domain phosphorylation. Nucleic Acids Res. 2020;48: 7712–7727.
92. Popova T, Manié E, Boeva V, Battistella A, Goundiam O, Smith NK, et al. Ovarian Cancers Harboring Inactivating Mutations in CDK12 Display a Distinct Genomic Instability Pattern Characterized by Large Tandem Duplications. Cancer Res. 2016;76: 1882–1891.
93. Cannataro VL, Townsend JP. Wagging the long tail of drivers of prostate cancer. PLoS Genet. 2019;15: e1007820.
94. Armenia J, Wankowicz SAM, Liu D, Gao J, Kundra R, Reznik E, et al. The long tail of oncogenic drivers in prostate cancer. Nat Genet. 2018;50: 645–651.
95. Irshad S, Bansal M, Castillo-Martin M, Zheng T, Aytes A, Wenske S, et al. A molecular signature predictive of indolent prostate cancer. Sci Transl Med. 2013;5: 202ra122.
96. Fraser M, Sabelnykova VY, Yamaguchi TN, Heisler LE, Livingstone J, Huang V, et al. Genomic hallmarks of localized, non-indolent prostate cancer. Nature. 2017;541: 359–364.
97. Fraser M. The somatic clinico-genomics of localized, non-indolent prostate cancer. J transl genet genom. 2018. doi:10.20517/jtgg.2018.27
98. Grönberg H, Ahman AK, Emanuelsson M, Bergh A, Damber JE, Borg A. BRCA2 mutation in a family with hereditary prostate cancer. Genes Chromosomes Cancer. 2001;30: 299–301.
99. Junejo NN, AlKhateeb SS. BRCA2 gene mutation and prostate cancer risk. Comprehensive review and update. Saudi Med J. 2020;41: 9–17.
100. Cavanagh H, Rogers KMA. The role of BRCA1 and BRCA2 mutations in prostate, pancreatic and stomach cancers. Hered Cancer Clin Pract. 2015;13: 16.
101. Ewing CM, Ray AM, Lange EM, Zuhlke KA, Robbins CM, Tembe WD, et al. Germline mutations in HOXB13 and prostate-cancer risk. N Engl J Med. 2012;366: 141–149.
102. Pilie PG, Giri VN, Cooney KA. HOXB13 and other high penetrant genes for prostate cancer. Asian J Androl. 2016;18: 530–532.
103. Stott-Miller M, Karyadi DM, Smith T, Kwon EM, Kolb S, Stanford JL, et al. HOXB13 mutations in a population-based, case-control study of prostate cancer. Prostate. 2013;73: 634–641.

104. Bratt O. Hereditary prostate cancer: clinical aspects. J Urol. 2002;168: 906–913.
105. Alberti C. Hereditary/familial versus sporadic prostate cancer: few indisputable genetic differences and many similar clinicopathological features. Eur Rev Med Pharmacol Sci. 2010;14: 31–41.
106. Quigley DA, Dang HX, Zhao SG, Lloyd P, Aggarwal R, Alumkal JJ, et al. Genomic Hallmarks and Structural Variation in Metastatic Prostate Cancer. Cell. 2018;174: 758–769.e9.
107. Ho Y, Dehm SM. Androgen Receptor Rearrangement and Splicing Variants in Resistance to Endocrine Therapies in Prostate Cancer. Endocrinology. 2017;158: 1533–1542.
108. Jernberg E, Bergh A, Wikström P. Clinical relevance of androgen receptor alterations in prostate cancer. Endocr Connect. 2017;6: R146–R161.
109. Paschalis A, Sharp A, Welti JC, Neeb A, Raj GV, Luo J, et al. Alternative splicing in prostate cancer. Nat Rev Clin Oncol. 2018;15: 663–675.
110. Yamashita S, Arai Y. AR Splice Variant in Prostate Cancer. In: Arai Y, Ogawa O, editors. Hormone Therapy and Castration Resistance of Prostate Cancer. Singapore: Springer Singapore; 2018. pp. 293–298.
111. Puhr M, Hoefer J, Eigentler A, Ploner C, Handle F, Schaefer G, et al. The Glucocorticoid Receptor Is a Key Player for Prostate Cancer Cell Survival and a Target for Improved Antiandrogen Therapy. Clin Cancer Res. 2018;24: 927–938.
112. Takeda DY, Spisák S, Seo J-H, Bell C, O'Connor E, Korthauer K, et al. A Somatically Acquired Enhancer of the Androgen Receptor Is a Noncoding Driver in Advanced Prostate Cancer. Cell. 2018;174: 422–432.e13.
113. Viswanathan SR, Ha G, Hoff AM, Wala JA, Carrot-Zhang J, Whelan CW, et al. Structural Alterations Driving Castration-Resistant Prostate Cancer Revealed by Linked-Read Genome Sequencing. Cell. 2018;174: 433–447.e19.
114. Abida W, Cyrta J, Heller G, Prandi D, Armenia J, Coleman I, et al. Genomic correlates of clinical outcome in advanced prostate cancer. Proc Natl Acad Sci U S A. 2019. doi:10.1073/pnas.1902651116
115. Antonarakis ES, Isaacsson Velho P, Fu W, Wang H, Agarwal N, Sacristan Santos V, et al. CDK12-Altered Prostate Cancer: Clinical Features and Therapeutic Outcomes to Standard Systemic Therapies, Poly (ADP-Ribose) Polymerase Inhibitors, and PD-1 Inhibitors. JCO Precis Oncol. 2020;4: 370–381.
116. Lotan TL, Antonarakis ES. CDK12 Deficiency and the Immune Microenvironment in Prostate Cancer. Clinical cancer research: an official journal of the American Association for Cancer Research. 2021. pp. 380–382.
117. Blee AM, Huang H. Lineage plasticity-mediated therapy resistance in prostate cancer. Asian J Androl. 2019;21: 241–248.
118. Beltran H, Hruszkewycz A, Scher HI, Hildesheim J, Isaacs J, Yu EY, et al. The Role of Lineage Plasticity in Prostate Cancer Therapy Resistance. Clin Cancer Res. 2019;25: 6916–6924.
119. Conteduca V, Oromendia C, Eng KW, Bareja R, Sigouros M, Molina A, et al. Clinical features of neuroendocrine prostate cancer. Eur J Cancer. 2019;121: 7–18.
120. Beltran H, Rickman DS, Park K, Chae SS, Sboner A, MacDonald TY, et al. Molecular characterization of neuroendocrine prostate cancer and identification of new drug targets. Cancer Discov. 2011;1: 487–495.
121. Ku SY, Rosario S, Wang Y, Mu P, Seshadri M, Goodrich ZW, et al. Rb1 and Trp53 cooperate

to suppress prostate cancer lineage plasticity, metastasis, and antiandrogen resistance. Science. 2017;355: 78–83.

122. Mu P, Zhang Z, Benelli M, Karthaus WR, Hoover E, Chen C-C, et al. SOX2 promotes lineage plasticity and antiandrogen resistance in TP53- and RB1-deficient prostate cancer. Science. 2017;355: 84–88.
123. Adams EJ, Karthaus WR, Hoover E, Liu D, Gruet A, Zhang Z, et al. FOXA1 mutations alter pioneering activity, differentiation and prostate cancer phenotypes. Nature. 2019;571: 408–412.
124. Lee JK, Phillips JW, Smith BA, Park JW, Stoyanova T, McCaffrey EF, et al. N-Myc Drives Neuroendocrine Prostate Cancer Initiated from Human Prostate Epithelial Cells. Cancer Cell. 2016;29: 536–547.
125. Zhang Y, Zheng D, Zhou T, Song H, Hulsurkar M, Su N, et al. Androgen deprivation promotes neuroendocrine differentiation and angiogenesis through CREB-EZH2-TSP1 pathway in prostate cancers. Nat Commun. 2018;9: 4080.
126. Mosquera JM, Beltran H, Park K, MacDonald TY, Robinson BD, Tagawa ST, et al. Concurrent AURKA and MYCN gene amplifications are harbingers of lethal treatment-related neuroendocrine prostate cancer. Neoplasia. 2013;15: 1–10.
127. Simental JA, Sar M, Lane MV, French FS, Wilson EM. Transcriptional activation and nuclear targeting signals of the human androgen receptor. J Biol Chem. 1991;266: 510–518.
128. Jenster G, van der Korput HA, van Vroonhoven C, van der Kwast TH, Trapman J, Brinkmann AO. Domains of the human androgen receptor involved in steroid binding, transcriptional activation, and subcellular localization. Mol Endocrinol. 1991;5: 1396–1404.
129. Rundlett SE, Wu XP, Miesfeld RL. Functional characterizations of the androgen receptor confirm that the molecular basis of androgen action is transcriptional regulation. Mol Endocrinol. 1990;4: 708–714.
130. Beato M, Chalepakis G, Schauer M, Slater EP. DNA regulatory elements for steroid hormones. J Steroid Biochem. 1989;32: 737–747.
131. Pomerantz MM, Li F, Takeda DY, Lenci R, Chonkar A, Chabot M, et al. The androgen receptor cistrome is extensively reprogrammed in human prostate tumorigenesis. Nat Genet. 2015;47: 1346–1351.
132. Stelloo S, Nevedomskaya E, Kim Y, Hoekman L, Bleijerveld OB, Mirza T, et al. Endogenous androgen receptor proteomic profiling reveals genomic subcomplex involved in prostate tumorigenesis. Oncogene. 2018;37: 313–322.
133. Chmelar R, Buchanan G, Need EF, Tilley W, Greenberg NM. Androgen receptor coregulators and their involvement in the development and progression of prostate cancer. Int J Cancer. 2007;120: 719–733.
134. Heinlein CA, Chang C. Androgen receptor (AR) coregulators: an overview. Endocr Rev. 2002;23: 175–200.
135. Sharma NL, Massie CE, Ramos-Montoya A, Zecchini V, Scott HE, Lamb AD, et al. The androgen receptor induces a distinct transcriptional program in castration-resistant prostate cancer in man. Cancer Cell. 2013;23: 35–47.
136. Lin C, Yang L, Tanasa B, Hutt K, Ju B-G, Ohgi K, et al. Nuclear receptor-induced chromosomal proximity and DNA breaks underlie specific translocations in cancer. Cell. 2009;139: 1069–1083.
137. Yu J, Yu J, Mani R-S, Cao Q, Brenner CJ, Cao X, et al. An integrated network of androgen receptor, polycomb, and TMPRSS2-ERG gene fusions in prostate cancer progression.

Cancer Cell. 2010;17: 443–454.

138. Zhang Z, Chng KR, Lingadahalli S, Chen Z, Liu MH, Do HH, et al. An AR-ERG transcriptional signature defined by long-range chromatin interactomes in prostate cancer cells. Genome Res. 2019;29: 223–235.
139. Kron KJ, Murison A, Zhou S, Huang V, Yamaguchi TN, Shiah Y-J, et al. TMPRSS2-ERG fusion co-opts master transcription factors and activates NOTCH signaling in primary prostate cancer. Nat Genet. 2017;49: 1336–1345.
140. Cuneo MJ, Mittag T. The ubiquitin ligase adaptor SPOP in cancer. FEBS J. 2019;286: 3946–3958.
141. An J, Wang C, Deng Y, Yu L, Huang H. Destruction of full-length androgen receptor by wild-type SPOP, but not prostate-cancer-associated mutants. Cell Rep. 2014;6: 657–669.
142. Geng C, Rajapakshe K, Shah SS, Shou J, Eedunuri VK, Foley C, et al. Androgen receptor is the key transcriptional mediator of the tumor suppressor SPOP in prostate cancer. Cancer Res. 2014;74: 5631–5643.
143. Zhang P, Wang D, Zhao Y, Ren S, Gao K, Ye Z, et al. Intrinsic BET inhibitor resistance in SPOP-mutated prostate cancer is mediated by BET protein stabilization and AKT-mTORC1 activation. Nat Med. 2017;23: 1055–1062.
144. Janouskova H, El Tekle G, Bellini E, Udeshi ND, Rinaldi A, Ulbricht A, et al. Opposing effects of cancer-type-specific SPOP mutants on BET protein degradation and sensitivity to BET inhibitors. Nat Med. 2017;23: 1046–1054.
145. Fong K-W, Zhao JC, Song B, Zheng B, Yu J. TRIM28 protects TRIM24 from SPOP-mediated degradation and promotes prostate cancer progression. Nat Commun. 2018;9: 5007.
146. Groner AC, Cato L, de Tribolet-Hardy J, Bernasocchi T, Janouskova H, Melchers D, et al. TRIM24 Is an Oncogenic Transcriptional Activator in Prostate Cancer. Cancer Cell. 2016;29: 846–858.
147. Augello MA, Liu D, Deonarine LD, Robinson BD, Huang D, Stelloo S, et al. CHD1 Loss Alters AR Binding at Lineage-Specific Enhancers and Modulates Distinct Transcriptional Programs to Drive Prostate Tumorigenesis. Cancer Cell. 2019;35: 603–617.e8.
148. Burkhardt L, Fuchs S, Krohn A, Masser S, Mader M, Kluth M, et al. CHD1 is a 5q21 tumor suppressor required for ERG rearrangement in prostate cancer. Cancer Res. 2013;73: 2795–2805.
149. Lei Q, Jiao J, Xin L, Chang C-J, Wang S, Gao J, et al. NKX3.1 stabilizes p53, inhibits AKT activation, and blocks prostate cancer initiation caused by PTEN loss. Cancer Cell. 2006;9: 367–378.
150. Perino M, Veenstra GJC. Chromatin Control of Developmental Dynamics and Plasticity. Dev Cell. 2016;38: 610–620.
151. Yadav T, Quivy J-P, Almouzni G. Chromatin plasticity: A versatile landscape that underlies cell fate and identity. Science. 2018;361: 1332–1336.
152. DNA Packaging: Nucleosomes and Chromatin. [cited 7 Feb 2021]. Available: https://www.nature.com/scitable/topicpage/dna-packaging-nucleosomes-and-chromatin-310/
153. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Chromosomal DNA and Its Packaging in the Chromatin Fiber. Garland Science; 2002.
154. Kornberg RD. Chromatin structure: a repeating unit of histones and DNA. Science. 1974;184: 868–871.

155. Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature. 1997;389: 251–260.
156. Constanze B, Cockerill PN. Chromatin Mechanisms Regulating Gene Expression In Health And Disease. Landes Bioscience; 2013.
157. Seitan VC, Faure AJ, Zhan Y, McCord RP, Lajoie BR, Ing-Simmons E, et al. Cohesin-based chromatin interactions enable regulated gene expression within preexisting architectural compartments. Genome Res. 2013;23: 2066–2077.
158. Lindsay S. Chromatin control of gene expression: the simplest model. Biophysical journal. 2007. p. 1113.
159. Jenuwein T, David Allis C. Translating the Histone Code. Science. 2001;293: 1074–1080.
160. Mellor J. The dynamics of chromatin remodeling at promoters. Mol Cell. 2005;19: 147–157.
161. Morse RH. Nucleosomes inhibit both transcriptional initiation and elongation by RNA polymerase III in vitro. The EMBO Journal. 1989. pp. 2343–2351. doi:10.1002/j.1460-2075.1989.tb08362.x
162. Li G, Margueron R, Hu G, Stokes D, Wang Y-H, Reinberg D. Highly compacted chromatin formed in vitro reflects the dynamics of transcription activation in vivo. Mol Cell. 2010;38: 41–53.
163. Natarajan A, Yardımcı GG, Sheffield NC, Crawford GE, Ohler U. Predicting cell-type–specific gene expression from regions of open chromatin. Genome Res. 2012;22: 1711–1722.
164. Klemm SL, Shipony Z, Greenleaf WJ. Chromatin accessibility and the regulatory epigenome. Nat Rev Genet. 2019;20: 207–220.
165. Stewart-Morgan KR, Reverón-Gómez N, Groth A. Transcription Restart Establishes Chromatin Accessibility after DNA Replication. Mol Cell. 2019;75: 284–297.e6.
166. Zaret KS, Carroll JS. Pioneer transcription factors: establishing competence for gene expression. Genes Dev. 2011;25: 2227–2241.
167. Dobersch S, Rubio K, Barreto G. Pioneer Factors and Architectural Proteins Mediating Embryonic Expression Signatures in Cancer. Trends Mol Med. 2019;25: 287–302.
168. Zaret KS, Lerner J, Iwafuchi-Doi M. Chromatin Scanning by Dynamic Binding of Pioneer Factors. Mol Cell. 2016;62: 665–667.
169. Iwafuchi-Doi M, Zaret KS. Pioneer transcription factors in cell reprogramming. Genes Dev. 2014;28: 2679–2692.
170. Mayran A, Sochodolsky K, Khetchoumian K, Harris J, Gauthier Y, Bemmo A, et al. Pioneer and nonpioneer factor cooperation drives lineage specific chromatin opening. Nat Commun. 2019;10: 3807.
171. Morris SA. Direct lineage reprogramming via pioneer factors; a detour through developmental gene regulatory networks. Development. 2016;143: 2696–2705.
172. Weigel D, Jäckle H. The fork head domain: a novel DNA binding motif of eukaryotic transcription factors? Cell. 1990;63: 455–456.
173. Kaestner KH, Knochel W, Martinez DE. Unified nomenclature for the winged helix/forkhead transcription factors. Genes Dev. 2000;14: 142–146.
174. Oldfield AJ, Yang P, Conway AE, Cinghu S, Freudenberg JM, Yellaboina S, et al. Histone-fold domain protein NF-Y promotes chromatin accessibility for cell type-specific master transcription factors. Mol Cell. 2014;55: 708–722.
175. Oldfield AJ, Henriques T, Kumar D, Burkholder AB, Cinghu S, Paulet D, et al. NF-Y controls fidelity of transcription initiation at gene promoters through maintenance of the nucleosome-depleted region. Nat Commun. 2019;10: 3072.

176. Chen G, Courey AJ. Groucho/TLE family proteins and transcriptional repression. Gene. 2000;249: 1–16.
177. Rodriguez-Bravo V, Carceles-Cordon M, Hoshida Y, Cordon-Cardo C, Galsky MD, Domingo-Domenech J. The role of GATA2 in lethal prostate cancer aggressiveness. Nat Rev Urol. 2017;14: 38–48.
178. Cirillo LA, Lin FR, Cuesta I, Friedman D, Jarnik M, Zaret KS. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. Mol Cell. 2002;9: 279–289.
179. Gao N, Ishii K, Mirosevich J, Kuwajima S, Oppenheimer SR, Roberts RL, et al. Forkhead box A1 regulates prostate ductal morphogenesis and promotes epithelial cell maturation. Development. 2005;132: 3431–3443.
180. Lupien M, Eeckhoute J, Meyer CA, Wang Q, Zhang Y, Li W, et al. FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. Cell. 2008;132: 958–970.
181. Yang YA, Yu J. Current perspectives on FOXA1 regulation of androgen receptor signaling and prostate cancer. Genes Dis. 2015;2: 144–151.
182. Wang Q, Li W, Liu XS, Carroll JS, Jänne OA, Keeton EK, et al. A hierarchical network of transcription factors governs androgen receptor-dependent prostate cancer growth. Mol Cell. 2007;27: 380–392.
183. Clapier CR, Iwasa J, Cairns BR, Peterson CL. Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. Nat Rev Mol Cell Biol. 2017;18: 407–422.
184. Längst G, Manelyte L. Chromatin Remodelers: From Function to Dysfunction. Genes . 2015;6: 299–324.
185. Vignali M, Hassan AH, Neely KE, Workman JL. ATP-dependent chromatin-remodeling complexes. Mol Cell Biol. 2000;20: 1899–1910.
186. Sima X, He J, Peng J, Xu Y, Zhang F, Deng L. The genetic alteration spectrum of the SWI/SNF complex: The oncogenic roles of BRD9 and ACTL6A. PLoS One. 2019;14: e0222305.
187. Reisman D, Glaros S, Thompson EA. The SWI/SNF complex and cancer. Oncogene. 2009;28: 1653–1668.
188. Clapier CR, Cairns BR. The Biology of Chromatin Remodeling Complexes. Annu Rev Biochem. 2009;78: 273–304.
189. Tyagi M, Imam N, Verma K, Patel AK. Chromatin remodelers: We are the drivers!! Nucleus. 2016;7: 388–404.
190. Kadoch C, Hargreaves DC, Hodges C, Elias L, Ho L, Ranish J, et al. Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. Nat Genet. 2013;45: 592–601.
191. Masliah-Planchon J, Bièche I, Guinebretière J-M, Bourdeaut F, Delattre O. SWI/SNF chromatin remodeling and human malignancies. Annu Rev Pathol. 2015;10: 145–171.
192. Kadoch C. Diverse compositions and functions of chromatin remodeling machines in cancer. Sci Transl Med. 2019;11. doi:10.1126/scitranslmed.aay1018
193. Turegun B, Baker RW, Leschziner AE, Dominguez R. Actin-related proteins regulate the RSC chromatin remodeler by weakening intramolecular interactions of the Sth1 ATPase. Commun Biol. 2018;1: 1.
194. Fujisawa T, Filippakopoulos P. Functions of bromodomain-containing proteins and their roles in homeostasis and cancer. Nat Rev Mol Cell Biol. 2017;18: 246–262.

195. Mashtalir N, D'Avino AR, Michel BC, Luo J, Pan J, Otto JE, et al. Modular Organization and Assembly of SWI/SNF Family Chromatin Remodeling Complexes. Cell. 2018;175: 1272–1288.e20.
196. Wilson BG, Roberts CWM. SWI/SNF nucleosome remodellers and cancer. Nat Rev Cancer. 2011;11: 481–492.
197. Kadoch C, Crabtree GR. Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics. Sci Adv. 2015;1: e1500447.
198. Schubert HL, Wittmeyer J, Kasten MM, Hinata K, Rawling DC, Héroux A, et al. Structure of an actin-related subcomplex of the SWI/SNF chromatin remodeler. Proc Natl Acad Sci U S A. 2013;110: 3345–3350.
199. Wang X, Wang S, Troisi EC, Howard TP, Haswell JR, Wolf BK, et al. BRD9 defines a SWI/SNF sub-complex and constitutes a specific vulnerability in malignant rhabdoid tumors. Nat Commun. 2019;10: 1881.
200. Dechassa ML, Zhang B, Horowitz-Scherer R, Persinger J, Woodcock CL, Peterson CL, et al. Architecture of the SWI/SNF-nucleosome complex. Mol Cell Biol. 2008;28: 6010–6021.
201. Centore RC, Sandoval GJ, Soares LMM, Kadoch C, Chan HM. Mammalian SWI/SNF Chromatin Remodeling Complexes: Emerging Mechanisms and Therapeutic Strategies. Trends Genet. 2020;36: 936–950.
202. Mathur R, Roberts CWM. SWI/SNF (BAF) Complexes: Guardians of the Epigenome. Annu Rev Cancer Biol. 2018;2: 413–427.
203. Hohmann AF, Vakoc CR. A rationale to target the SWI/SNF complex for cancer therapy. Trends Genet. 2014;30: 356–363.
204. Mittal P, Roberts CWM. The SWI/SNF complex in cancer - biology, biomarkers and therapy. Nat Rev Clin Oncol. 2020;17: 435–448.
205. Przybyl J, Sciot R, Rutkowski P, Siedlecki JA, Vanspauwen V, Samson I, et al. Recurrent and novel SS18-SSX fusion transcripts in synovial sarcoma: description of three new cases. Tumour Biol. 2012;33: 2245–2253.
206. Baranov E, McBride MJ, Bellizzi AM, Ligon AH, Fletcher CDM, Kadoch C, et al. A Novel SS18-SSX Fusion-specific Antibody for the Diagnosis of Synovial Sarcoma. Am J Surg Pathol. 2020;44: 922–933.
207. Iwafuchi-Doi M, Donahue G, Kakumanu A, Watts JA, Mahony S, Pugh BF, et al. The Pioneer Transcription Factor FoxA Maintains an Accessible Nucleosome Configuration at Enhancers for Tissue-Specific Gene Activation. Mol Cell. 2016;62: 79–91.
208. Factor DC, Corradin O, Zentner GE, Saiakhova A, Song L, Chenoweth JG, et al. Epigenomic comparison reveals activation of "seed" enhancers during transition from naive to primed pluripotency. Cell Stem Cell. 2014;14: 854–863.
209. Parolia A, Cieslik M, Chu S-C, Xiao L, Ouchi T, Zhang Y, et al. Distinct structural classes of activating FOXA1 alterations in advanced prostate cancer. Nature. 2019;571: 413–418.

Chapter 2

Distinct Structural Classes of Activating FOXA1 Alterations in Advanced Prostate Cancer[1]

Abstract

Forkhead box A1 (FOXA1) is a pioneer transcription factor that is essential for the normal development of several endoderm-derived organs, including the prostate gland. FOXA1 is frequently mutated in hormone-receptor driven prostate, breast, bladder, and salivary gland tumors. However, it is unclear how FOXA1 alterations affect the development of cancer, and FOXA1 has previously been ascribed both tumor-suppressive and oncogenic roles. Here we assemble an aggregate cohort of 1,546 prostate cancers and show that FOXA1 alterations fall into three structural classes that diverge in clinical incidence and genetic co-alteration profiles, with a collective prevalence of 35%. Class-1 activating mutations originate in early prostate cancer without alterations in drivers such as ETS-factors or SPOP, selectively recur within the wing-2 region of the DNA-binding forkhead domain, enable enhanced chromatin mobility and binding frequency, and strongly transactivate a luminal androgen-receptor program of prostate oncogenesis. By contrast, class-2 activating mutations are acquired in metastatic prostate cancers,

---

[1] Contents of this chapter were previously published as part of the following manuscript: Parolia A, Cieslik M, et al. Distinct structural classes of activating FOXA1 alterations in advanced prostate cancer. Nature. 2019;571: 413–418. Specific author contributions are included in chapter acknowledgements.

truncate the C-terminal domain of FOXA1, enable dominant chromatin binding by increasing DNA affinity, and strongly activate the WNT program of prostate cancer metastasis. Finally, class-3 genomic rearrangements are enriched in metastatic prostate cancers, consist of duplications and translocations within the *FOXA1* locus, and structurally reposition a conserved regulatory element—herein denoted *FOXA1* mastermind (*FOXMIND*)—to drive overexpression of *FOXA1* or other oncogenes. Our study reaffirms the central role of FOXA1 in mediating AR-driven oncogenesis and provides mechanistic insights into how different classes of FOXA1 alterations uniquely promote prostate cancer initiation and/or metastatic progression. Furthermore, these results have direct implications in understanding the biology of other hormone-receptor driven cancers and rationalize therapeutic co-targeting of FOXA1 activity.

Introduction

FOXA proteins are endodermal development factors that show restricted expression in organs that originate from the endoderm, including the bladder, pancreas, and prostate glands [1,2]. Breast tissue is an exception to this rule and is the only non-endodermal organ that shows high expression of the *FOXA1* gene [3]. Genomic sequencing studies of human cancers has revealed a high-recurrence of FOXA1 mutations in nuclear hormone-driven prostate, breast and salivary gland cancers [4–7]. FOXA1 is an essential modulator of the hormone receptor-signaling complex, including AR [8,9]. FOXA1 functions as a pioneer factor by independently binding to and de-compacting condensed chromatin to reveal the underlying binding sites of partnering transcription factors (TFs) [10,11]. The “winged helix” DNA-binding domain of FOXA1 structurally resembles the linker histone H1 [12], the latter of which functions as a molecular glue to compactly package the nucleosomes [13,14]. FOXA1 binding to the chromatin has been shown to displace these linker histones thereby making the DNA functionally accessible [15,16]. Thus, FOXA1 and other pioneer

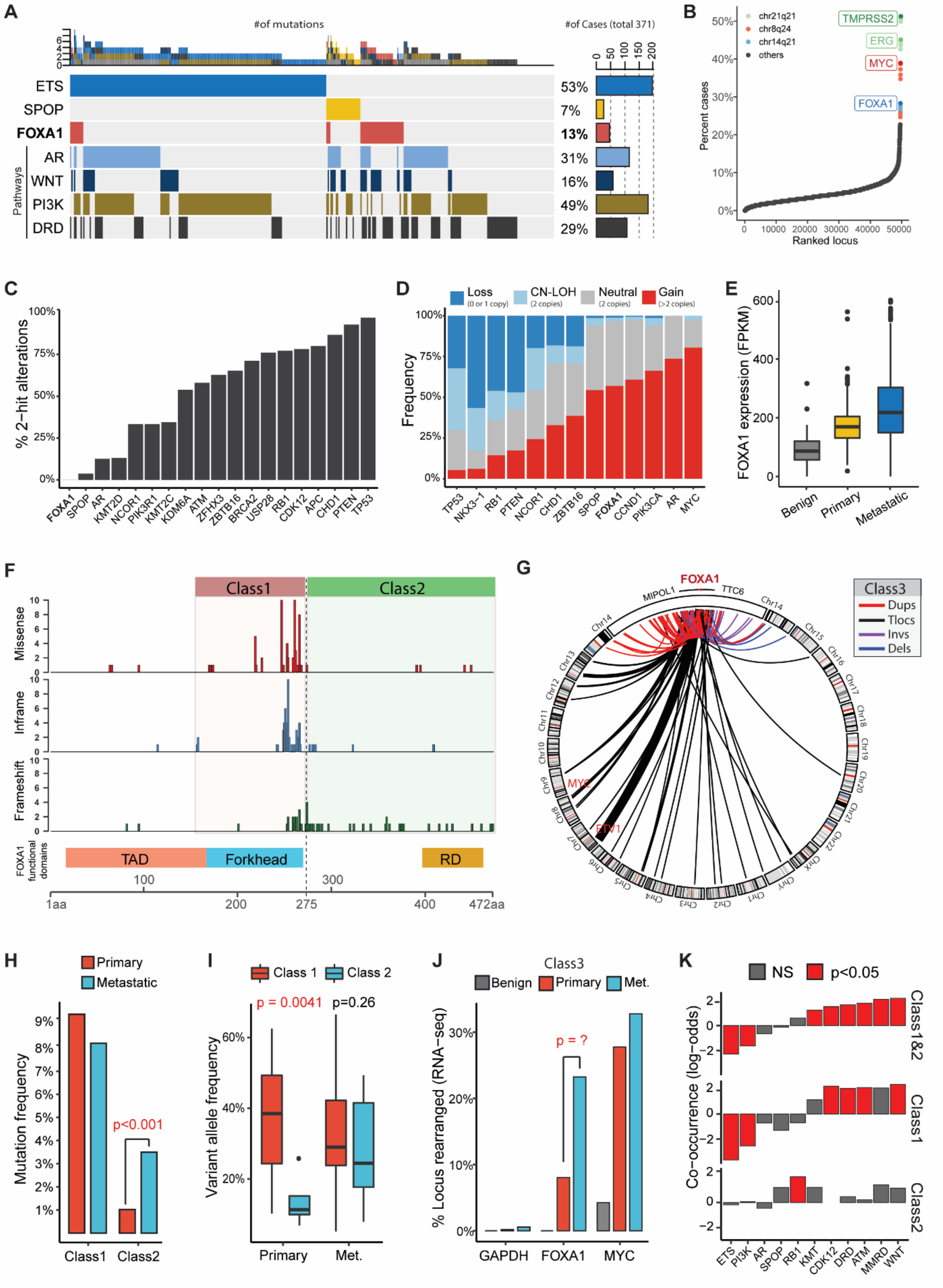

A
#of mutations
#of Cases (total 371)
0
2
4
6
0
50
100
150
200
ETS
SPOP
FOXA1
AR
WNT
PI3K
DRD
Pathways
53%
7%
13%
31%
16%
49%
29%
B
chr21q21
chr8q24
chr14q21
others
TMPRSS2
ERG
MYC
FOXA1
Percent cases
50%
40%
30%
20%
10%
0%
0
10000
20000
30000
40000
50000
Ranked locus
C
% 2-hit alterations
100%
75%
50%
25%
0%
FOXA1
SPOP
AR
KMT2D
NCOR1
PIK3R1
KMT2C
KDM6A
ATM
ZFHX3
ZBTB16
BRCA2
USP28
RB1
CDK12
APC
CHD1
PTEN
TP53
D
Loss
(0 or 1 copy)
CN-LOH
(2 copies)
Neutral
(2 copies)
Gain
(>2 copies)
Frequency
100%
75%
50%
25%
0%
TP53
NKX3−1
RB1
PTEN
NCOR1
CHD1
ZBTB16
SPOP
FOXA1
CCND1
PIK3CA
AR
MYC
E
FOXA1 expression (FPKM)
600
400
200
0
Benign
Primary
Metastatic
F
Class1
Class2
Missense
Inframe
Frameshift
10
8
6
4
2
0
FOXA1 functional domains
TAD
Forkhead
RD
100
300
1aa
200
275
400
472aa
G
FOXA1
MIPOL1
TTC6
Class3
Dups
Tlocs
Invs
Dels
Chr14
Chr15
Chr16
Chr17
Chr18
Chr19
Chr20
Chr21
Chr22
ChrX
ChrY
Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Chr8
Chr9
Chr10
Chr11
Chr12
Chr13
MYC
ETV1
H
Primary
Metastatic
Mutation frequency
9%
8%
7%
6%
5%
4%
3%
2%
1%
p<0.001
Class1
Class2
I
Class 1
Class 2
p = 0.0041
p=0.26
Variant allele frequency
60%
40%
20%
Primary
Met.
J
Class3
Benign
Primary
Met.
% Locus rearranged (RNA−seq)
30%
20%
10%
0%
p = ?
GAPDH
FOXA1
MYC
K
NS
p<0.05
Co−occurrence (log−odds)
2
0
−2
Class1&2
Class1
Class2
ETS
PI3K
AR
SPOP
RB1
KMT
CDK12
DRD
ATM
MMRD
WNT

Figure 2-1: Distinct structural classes of FOXA1 aberrations. (A) Landscape of FOXA1 mutations relative to other key alterations in mCRPC. Mutations in AR, WNT, PI3K were aggregated at the pathway level, ETS represents ETS gene fusions, DRD represents DNA repair defects and included alterations in BRCA1/2, ATM and CDK12, MMRD represents mismatch repair deficiency. Overall alteration prevalence is tallied in the right panel. (B) Locus-level (1.5Mb sliding windows) recurrence of structural variations (SVs) based on RNA-seq. (C) Frequency of bi-allelic inactivation, and (D) frequency of copy-number changes, across the mCRPC cohort. (E) Tumor-content adjusted expression of FOXA1 in metastatic relative to primary prostate tumors. (F) Structural classification of FOXA1 mutations. TAD, transactivation domain; Forkhead, Forkhead DNA-binding domain; RD, regulatory domain (G) Structural classification of FOXA1 locus rearrangements. Dups, Tandem duplications; Tlocs, translocations; Invs, inversions; Dels, deletions. (H) Mutation frequency of FOXA1 by stage and class (Fisher's exact test). (I) Variant allele frequency by stage and class (t-test). (J) Locus-level recurrence of SVs based on RNA-seq by stage (Fisher's exact test). (K) Mutual exclusivity or co-occurrence of FOXA1 mutations (categories same as in A, Fisher's exact test).

factors function as primary instructors of the non-coding *cis*-regulome, which in turn establishes the cellular identity.

In prostate luminal epithelial cells, FOXA1 delimits a tissue-specific enhancer landscape that is transcriptionally activated by AR to drive the expression of prostatic genes [8,9,17]. FOXA1 also plays a central role in the extensive redistribution of AR on the chromatin to bind cancer-specific sites in the transformed prostate epithelia [18,19]. Accordingly, FOXA1 and AR are co-expressed in prostate cancer cells (**Supplementary Figure A1-1A,B**), wherein, akin to AR, FOXA1 activity is indispensable for survival and proliferation [9,19] (**Supplementary Figure A1-1C-E**). Elevated protein expression of FOXA1, at the time of prostatectomy, also correlates with poorer clinicopathological features and shorter time to biochemical recurrence [20–22]. Furthermore, overexpression of WT FOXA1 has been experimentally shown to promote aggressive phenotypes in prostate cancer and AR-driven apocrine breast cancer cells [20,23,24].

In this context, it is intriguing that FOXA1 is the third most-highly mutated gene [4,5], and, as shown here for the first time, among the most-highly rearranged genomic loci in AR-dependent metastatic prostate cancer. These insights raise a pertinent biological question: What are the functional consequences of FOXA1 alterations and how do they affect disease biology? Counterintuitively, recent studies have suggested these alterations to be inactivating [25,26] and have described FOXA1 to be a tumor suppressor in AR-driven metastatic prostate cancer [27,28]. However, there has been no systematic effort to experimentally characterize FOXA1 alterations in cancer. In this study, we describe genomic characteristics and mechanistic roles of FOXA1 mutations in prostate cancer. We demonstrate that FOXA1 mutations are activating and that they fall into three structural classes that have distinct clinical incidence, genetic co-alterations profiles, and gain-of-function that uniquely promote prostate cancer initiation and/or metastatic progression.

Results

FOXA1 alterations recur into three structural classes

For a comprehensive genomic characterization of FOXA1 aberrations, we first curated an aggregate prostate cancer cohort comprising of 888 localized and 658 metastatic prostate cancer samples from major clinical sequencing studies [4,5,29,30], 498, and 357 cases of which had matched RNA-seq data, respectively. The metastatic cases in our cohort were castration-resistant, predominantly AR-positive, and comprised of very few neuroendocrine (NE) tumors. FOXA1 mutations recurred at an overall frequency of 8% in the primary localized disease that increased to 13% in metastatic castration-resistant prostate cancer (mCRPC; Figure 2-1A and **Supplementary Figure A1-2A**). Intriguingly, RNA-seq based calls of structural variants revealed a high prevalence (**Figure 2-1B**) and density (**Supplementary Figure A1-2B**) of variants within the FOXA1 locus, akin to previously reported MYC locus rearrangements [31,32].

To validate these calls, we carried out whole-genome sequencing (WGS) of four index cases and identified a matched genomic breakpoint in all of them (**Supplementary Figure A1-2C**). We also called copy-number breakpoints based on whole-exome sequencing (WES) across 370 metastatic samples, and copy-number variant (CNV) analysis. Bearing in mind that only a subset of genomic rearrangements results in CNVs or fusions, we detected a highly significant overlap between the two readouts (odds-ratio=6.81, p=4.3e-09; Supplementary Figure A1-2D). WGS of three additional cases with copy-number breakpoints again confirmed a precise genomic breakpoint (**Supplementary Figure A1-2C**). Overall, we estimate the recurrence of FOXA1 locus rearrangements at 20%-30% in mCRPC (**Supplementary Figure A1-2E**). For FOXA1 mutations, all the cases were heterozygous (**Figure 2-1C**), and the gene itself showed copy-number gains in over 50% of the cases without a single event of homozygous deletion (**Figure 2-1D**). Next, we assessed the expression of FOXA1 in different clinical stages of prostate cancer. We found a stage-wise increasing trend in the mRNA expression of FOXA1 (**Figure 2-1E**). This is consistent with a parallel protein-level increase described in multiple immunohistochemical studies [20–22]. Together, this genomic alteration profile strongly suggests that FOXA1 functions as an oncogene in AR-dependent prostate cancer.

Next, on mapping individual mutations onto the functional domains of FOXA1, we found two structural patterns to their distribution: 1) missense and in-frame indel mutations were clustered at the C-terminal end of the FKHD, while 2) truncating frameshift mutations were restricted to the C-terminal half of the protein (**Figure 2-1F**). FOXA1 structural variations predominantly consisted of tandem-duplications and translocations, which clustered in close proximity to the gene (**Figure 2-1G** and **Supplementary Figure A1-2B**). Notably, these alterations did not disrupt the FOXA1 coding sequence and rarely involved translocations with other oncogenes such as ETV1 or MYC. Thus, we categorized FOXA1 aberrations into three structural classes: class1 comprising all the mutations within the FKHD, class2 comprising mutations in the C-terminal end following the FKHD (**Figure 2-1F** and **Supplementary Figure 2-3A**), and class3 comprising structural variations impacting FOXA1 regulatory regions (**Figure 2-1G**). Notably, in breast cancer cases, where FOXA1 is also recurrently mutated [6], distribution of the FOXA1 alterations largely followed the structural patterns defined above (**Supplementary Figure A1-3B,C**).

Most remarkably, we found that the majority of FOXA1 mutations in primary prostate cancer belonged to class1, which showed no enrichment in metastatic disease (**Figure 2-1H**). Conversely, class2 mutations were significantly enriched (p=0.01) in metastatic prostate cancer (**Figure 2-1H**). In the rare primary cases with class2 mutations, the mutant allele was detected at significantly (p=0.0041) lower sub-clonal variant allele frequencies as opposed to clonal levels in metastatic prostate cancer (**Figure 2-1I** and **Supplementary Figure A1-4A**). Class3 aberrations were also significantly (odds ratio=3.46, p=7.9e-10) enriched in metastatic tumors (**Figure 2-1J**). Assessment of concurrent alterations further revealed that class1 mutations were significantly mutually exclusive with other primary events such as ETS fusions (odds ratio=0.078, p=9.6e-08), while class2 mutant cases were enriched for RB1 loss in metastatic prostate cancer (odds-ratio=4.17, p=0.034) (**Figure 2-1K** and **Supplementary Figure A1-4B**). Both mutational classes were further enriched for alterations in DNA repair (DRD), mismatch repair (MMRD) and WNT signaling pathways (**Figure 2-1K** and **Supplementary Figure A1-4F**). Additionally, both mutant classes had statistically higher mRNA expression of FOXA1 relative to WT cases in primary and metastatic prostate cancer (**Supplementary Figure A1-4G**). Taken together, these data reveal distinct structural patterns, clinical incidence, and genetic associations of FOXA1 aberrations; and demonstrate that class1 mutations are early driver alterations that emerge in localized prostate cancer, while class2 mutations and class3 structural variations are acquired or enriched, respectively, in the course of metastatic progression.

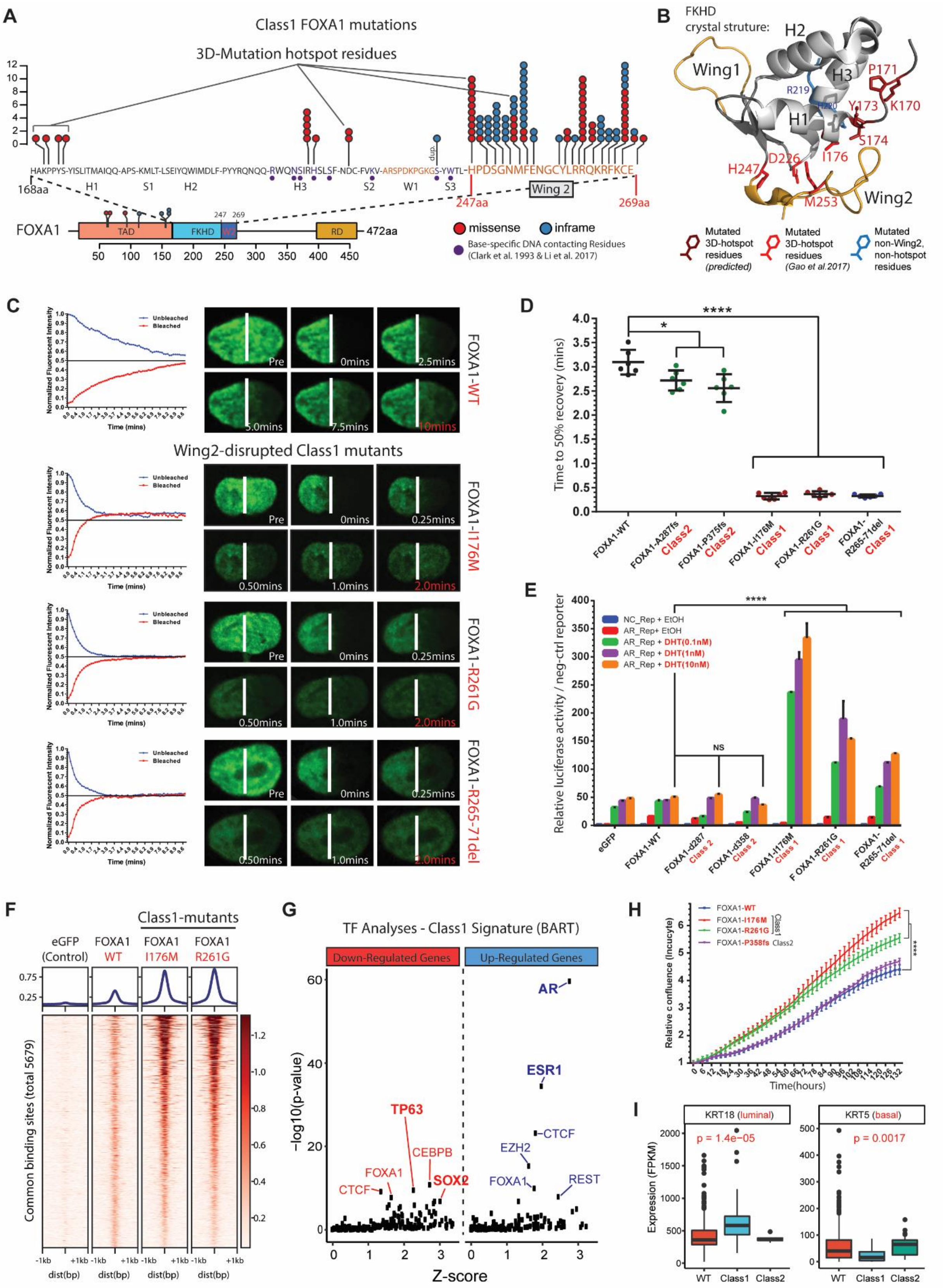
A
Class1 FOXA1 mutations
3D-Mutation hotspot residues
HAKPPYS-YISLITMAIQQ-APS-KMLT-LSEIYQWIMDLF-PYYRQNQQ-RWQNSIRHSLSF-NDC-FVKV-ARSPDKPGKGS-YWTL-HPDSGNMFENGCYLRRQKRFKCE
dup
168aa
H1
S1
H2
H3
S2
W1
S3
Wing 2
247aa
269aa
FOXA1
TAD
FKHD
RD
472aa
247
269
50 100 150 200 250 300 350 400 450
missense
inframe
Base-specific DNA contacting Residues (Clark et al. 1993 & Li et al. 2017)
B
FKHD crystal struture:
H2
Wing1
H3
P171
R219
H1
Y173
K170
S174
H247
D226
I176
M253
Wing2
Mutated 3D-hotspot residues (predicted)
Mutated 3D-hotspot residues (Gao et al.2017)
Mutated non-Wing2, non-hotspot residues
C
Normalized Fluorescent Intensity
Unbleached
Bleached
Time (mins)
Pre
0mins
2.5mins
5.0mins
7.5mins
10mins
FOXA1-WT
Wing2-disrupted Class1 mutants
0.25mins
0.50mins
1.0mins
2.0mins
FOXA1-I176M
FOXA1-R261G
FOXA1-R265-71del
D
****
*
Time to 50% recovery (mins)
FOXA1-WT
FOXA1-A287fs Class2
FOXA1-P375fs Class2
FOXA1-I176M Class1
FOXA1-R261G Class1
FOXA1-R265-71del Class1
E
****
NC_Rep + EtOH
AR_Rep+ EtOH
AR_Rep + DHT(0.1nM)
AR_Rep + DHT(1nM)
AR_Rep + DHT(10nM)
NS
Relative luciferase activity / neg-ctrl reporter
eGFP
FOXA1-WT
FOXA1-d287 Class 2
FOXA1-d358 Class 2
FOXA1-I176M Class 1
F OXA1-R261G Class 1
FOXA1-R265-71del Class 1
F
Class1-mutants
eGFP (Control)
FOXA1 WT
FOXA1 I176M
FOXA1 R261G
Common binding sites (total 5679)
-1kb +1kb
dist(bp)
G
TF Analyses - Class1 Signature (BART)
Down-Regulated Genes
Up-Regulated Genes
-log10(p-value)
AR
ESR1
TP63
CEBPB
FOXA1
CTCF
SOX2
EZH2
FOXA1
CTCF
REST
Z-score
H
FOXA1-WT
FOXA1-I176M
FOXA1-R261G
Class1
FOXA1-P358fs Class2
****
Relative confluence (Incucyte)
Time(hours)
I
KRT18 (luminal)
p = 1.4e-05
KRT5 (basal)
p = 0.0017
Expression (FPKM)
WT
Class1
Class2

Figure 2-2: Functional characterization of Class1 mutations of FOXA1. (A) Distribution of class1 mutations on the protein map of FOXA1 functional domains and FKHD secondary structures. (B) Crystal structure of the FKHD (from FOXA3) with visualization of all recurrent mutations outside of Wing2 (247-269aa). 3D-hotspot mutations are shown in shades of red. (C) FRAP recovery kinetic plots (left) and representative time-lapse images from pre-bleaching ('Pre') to equilibrated state (red timestamps). (D) Quantified FRAP data depicting time taken to recover to 50% fluorescence post-bleaching (6 nuclei/group; ANOVA and Tukey's test). (E) AR dual-luciferase reporter assays with transient overexpression of labeled FOXA1 variants in HEK293-AR cells with or without DHT stimulation (6 replicates/group; ANOVA and Tukey's test). (F) Averaged peak profiles (top) and read density heatmaps (bottom) from ChIP-Seq data shown at genomic loci bound by WT and class1 mutants in HEK293 cells. (G) BART prediction of specific TFs mediating observed transcriptional changes. The most likely (significance) and strong (Z-score) mediators of transcriptional responses mediated by TFs in class1 tumors are highlighted. (H) Growth curves (IncuCyte assay) of 22RV1 cells with stable, dox-inducible overexpression of FOXA1 mutants or WT FOXA1 in androgen-depleted medium (6 replicates/group; ANOVA and Tukey's test). (I) Expression of luminal and basal markers in class1 tumors vs WT and class2 tumors (ANOVA).

Class 1 mutants are stronger AR co-activators and promote prostate oncogenesis

The FOX gene family is defined by the highly-conserved, winged-helix DNA-binding domain (aka FKHD) [33]. In our aggregate cohort, we found class1 mutations to be particularly clustered at the C-terminal edge of the FKHD, prompting us to further map them on the constituent secondary structures. The majority of the class1 mutations were located within the Wing2 region, extending from amino acid 247 to 269 (**Figure 2-2A** and **Supplementary Figure 2-5A**). Rare mutations outside of the wing were part of a 3D-hotspot that spatially oriented in close proximity to Wing2 [34] **(Figure 2-2B** and **Supplementary Figure A1-5B**). Notably, class1 mutations did not alter key residues within Helix 3 (H3) and Wing1, which make direct DNA base-specific interactions identified from DNA-bound FOXA crystal structures [12,35] (purple dots; **Figure 2-2A** and **Supplementary Figure A1-5C**). Notably, Wing2 is not strongly conserved amongst FOX family genes [33,35]. In FOXA proteins, Wing2 residues stabilize chromatin-binding by making base-independent (i.e., non-specific) contacts with the sugar-phosphate backbone31, and these interactions have been shown to impede nuclear mobility [36]. These data strongly suggest that class1 mutations selectively disrupt Wing2 of the FKHD, are unlikely to abolish the DNA-binding ability or sequence-specificity of FOXA1 but may affect its nuclear mobility.

WT FOXA1 is a relatively sluggish TF that takes over 10 mins to fully recover in the fluorescence recovery after photobleaching (FRAP) assay [36] (top, **Figure 2-2C**). This is markedly slower than MYC, which takes 1 min to recover [36]. Thus, we hypothesized that Wing2-disrupted class1 mutants would display faster mobility through the chromatin. We cloned three representative class1 mutants I176M (3D-hotspot mutation), R261G (missense), and R265-71del (in-frame deletion), and fused them to GFP for FRAP assays (Supplementary Figure A1-6A). Like WT FOXA1, all mutants showed predominant nuclear localization and had equal expression (**Supplementary Figure A1-6B**). Remarkably, we found that class1 mutants had 5-6 times faster nuclear mobility than WT FOXA1, irrespective of the mutation-type that disrupted Wing2 (**Figure 2-2C**). Class1 mutants took roughly 30 seconds to reach 50% recovery post-bleaching, while the WT protein took 3 minutes for the same (**Figure 2-2D**). Notably, class2 mutants with intact Wing2 were still sluggish in their nuclear movement, taking over 2.5mins for 50% recovery (**Figure 2-2D and Supplementary Figure A1-6C,D**).

Given the marked increase in nuclear mobility, we next assessed the effect of class1 mutants on the transactivation of AR-signaling. Notably, like WT FOXA1, both class1 and class2 mutants continued to interact with the AR-signaling complex (**Supplementary Figure A1-7A-C**). Strikingly, in reporter assays, at comparable expression, class1 mutants enabled a 3-6 fold increase in AR-activity relative to WT FOXA1 (**Figure 2-2E** and **Supplementary Figure A1-7D**). This increase was evident even under castrate levels of androgen (0.1nM and 1nM; **Figure 2-2E**). Accordingly, class1 mutants markedly rescued the inhibitory effect of enzalutamide on AR-activity (**Supplementary Figure A1-7E**). In parallel assays, class2 mutants had no significant differences relative to WT FOXA1 (**Figure 2-2E** and **Supplementary Figure A1-7E**). Next, we performed chromatin-immunoprecipitation (ChIP) assays in an isogenic system to profile the cistromes of class1 mutants. At comparable levels of expression, the mutants showed larger cistromes and stronger binding at common or mutant-specific sites relative to WT FOXA1 (**Figure 2-2F** and **Supplementary Figure A1-8A,B**), and retained comparable binding at FOXA1 WT sites (**Supplementary Figure A1-8C**). This strongly suggests that faster nuclear mobility dramatically enhances the ability of class1 mutants to interact with target DNA and transactivate the AR-signaling complex.

To identify corroborating evidence in patient samples, we identified genes differentially expressed in tumors with class1 mutations relative to a WT group of patients. The class1-specific gene signature revealed marked activation of hyper-proliferative and pro-tumorigenesis pathways (**Supplementary Figure A1-9A,B**), and enrichment of primary (luminal) prostate cancer gene expression program (**Supplementary Figure A1-9C**). Next, we used a computational approach, BART [37], to predict likely transcriptional regulators of genes that were differentially expressed in tumors with class1 mutations. Here, AR was predicted to be the most significant TF for the up-regulated genes (**Figure 2-2G**). We experimentally confirmed WNT7B and CASP2, which were amongst the most significantly up-regulated and down-regulated genes respectively, to be direct AR/FOXA1 targets in prostate cancer cells (**Supplementary Figure A1-9D-F**). Interestingly, WNT7B and CASP2 have well-described oncogenic [38,39] or tumor suppressive [40,41] roles, respectively. To further validate this approach, we carried out BART analysis for genes that were differentially expressed upon FOXA1 knockdown in LNCaP cells. As expected, FOXA1 and AR were predicted as top TFs (**Supplementary Figure A1-10A**), verifying FOXA1 as an essential cofactor of oncogenic AR-activity. Concordantly, isogenic 22RV1 cells with overexpression of class1 mutants showed significantly higher growth in the androgen-depleted medium (**Figure 2-2H**), but not in androgen-supplemented complete medium (**Supplementary Figure A1-10B**).

Interestingly, in the BART analyses, TP63 and SOX2 were predicted to be among the top transcriptional regulators for genes down-regulated in class1 tumors (**Figure 2-2G**). Notably, expression of TP63 and SOX2 itself was significantly down-regulated in class1 mutant cases versus both class2 mutant and WT cases (**Supplementary Figure A1-10C**). Since these TFs are implicated in driving a basal prostatic gene program [42,43], we looked at the expression of canonical basal and luminal markers in class1 tumors. Basal markers such as KRT5 and KRT14 were significantly downregulated, while luminal markers such as KRT8 and KRT18 were significantly upregulated in class1 mutant tumors (**Figure 2-2I** and **Supplementary Figure A1-10D**). Class1 tumors also had a significantly higher AR-score with a parallel reduction in the NE-score relative to WT tumors (**Supplementary Figure A1-10E**). Together, these data demonstrate that class1 mutants have faster nuclear mobility and interact more efficiently with their genomic targets, which confer the mutants higher competence to enhance the luminal oncogenic program of AR.

Class 2 mutants are cistromically-dominant and promote metastasis

Class2 mutations comprise frameshifting alterations that introduce a premature stop codon and truncate the C-terminal regulatory domain of FOXA1 (**Figure 2-3A**). Thus, we used N-terminal and C-terminal antibodies to characterize the cistromes of class2 mutants. We validated both antibodies using recombinant proteins where the N-terminal antibody recognized both the full-length and truncated FOXA1 variants, while the C-terminal antibody recognized only the full-length WT protein (**Supplementary Figure A1-11**).

Amongst CRPC-derived AR-dependent prostate cancer cells, LAPC4 endogenously harbors a heterozygous class2 mutation in FOXA1 (i.e. P358fs; **Supplementary Figure A1-12A**). In these

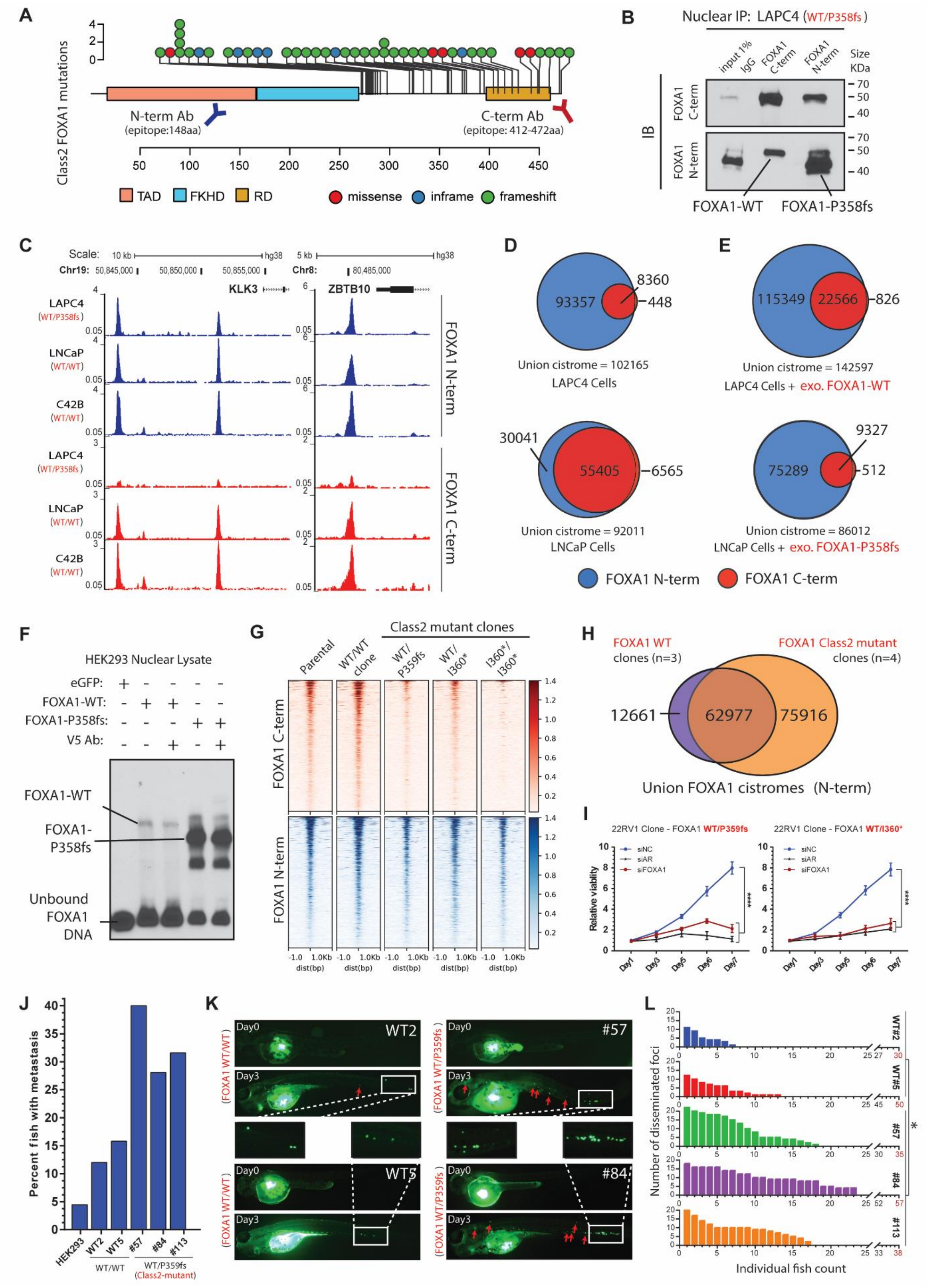

A
Class2 FOXA1 mutations
4
2
0
N-term Ab
(epitope:148aa)
C-term Ab
(epitope: 412-472aa)
50 100 150 200 250 300 350 400 450
TAD FKHD RD
missense inframe frameshift
B
Nuclear IP: LAPC4 (WT/P358fs)
input 1%
IgG
FOXA1 C-term
FOXA1 N-term
Size KDa
IB
FOXA1 C-term
FOXA1 N-term
70
50
40
FOXA1-WT
FOXA1-P358fs
C
Scale: 10 kb hg38
Chr19: 50,845,000 50,850,000 50,855,000
KLK3
5 kb hg38
Chr8: 80,485,000
ZBTB10
LAPC4 (WT/P358fs)
LNCaP (WT/WT)
C42B (WT/WT)
FOXA1 N-term
FOXA1 C-term
D
93357 8360 448
Union cistrome = 102165
LAPC4 Cells
30041 55405 6565
Union cistrome = 92011
LNCaP Cells
E
115349 22566 826
Union cistrome = 142597
LAPC4 Cells + exo. FOXA1-WT
75289 9327 512
Union cistrome = 86012
LNCaP Cells + exo. FOXA1-P358fs
FOXA1 N-term
FOXA1 C-term
F
HEK293 Nuclear Lysate
eGFP: + - - - -
FOXA1-WT: - + + - -
FOXA1-P358fs: - - - + +
V5 Ab: - - + - +
FOXA1-WT
FOXA1-P358fs
Unbound FOXA1 DNA
G
Class2 mutant clones
Parental
WT/WT clone
WT/P359fs
WT/I360*
I360*/I360*
FOXA1 C-term
FOXA1 N-term
-1.0 1.0Kb dist(bp)
H
FOXA1 WT clones (n=3)
FOXA1 Class2 mutant clones (n=4)
12661 62977 75916
Union FOXA1 cistromes (N-term)
I
22RV1 Clone - FOXA1 WT/P359fs
22RV1 Clone - FOXA1 WT/I360*
siNC
siAR
siFOXA1
Relative viability
Day1 Day3 Day5 Day6 Day7
****
J
Percent fish with metastasis
HEK293 WT2 WT5 #57 #84 #113
WT/WT
WT/P359fs (Class2-mutant)
K
Day0 Day3
WT2 WT5 #57 #84
(FOXA1 WT/WT)
(FOXA1 WT/P359fs)
L
Number of disseminated foci
Individual fish count
WT#2 WT#5 #57 #84 #113
*

Figure 2-3: Functional characterization of Class2 mutations of FOXA1. (A) Distribution of class2 mutations on the protein map of FOXA1. FOXA1 N-terminal and C-terminal antibodies with their proposed epitope sites are shown. (B) Co-immunoprecipitation and immunoblotting of FOXA1 using a N-terminal (N-term) and C-terminal (C-term) antibodies from LAPC4 nuclei. Species-matched IgG was used as control. (C) FOXA1 N-term and C-term ChIP-Seq normalized signal tracks from FOXA1 WT or class2 mutant PCa cells at canonical AR targets KLK3 and ZBTB10. (D and E) Overlaps between N-term and C-term FOXA1 cistromes in PCa cells at endogenous expression levels or (E) with exogenous overexpression of different FOXA1 variants. (F) EMSA with recombinant WT or P358fs mutant and KLK-enhancer element containing the FOXA1 consensus motif. (G) FOXA1 N-term and C-term ChIP-Seq read-density heatmaps in 22RV1 CRISPR clones with distinct class2 mutations. (H) Overlap between union N-term FOXA1 cistromes from WT (n=3) and class2-mutant (n=4) 22RV1 clones. (I) Growth curves of class2-mutant 22RV1 clones treated with non-targeting (siNC), AR or FOXA1 targeting siRNAs (n=6; ANOVA and Tukey’s test). (J) Frequency of metastasis in zebrafish embryos injected with HEK293 (negative control), WT or class2-mutant 22RV1 clones (n>=30/group) (K) Representative embryo images and, (L) disseminated foci counts (ANOVA; Tukey’s test) depicting the extent of metastases.

cells, both the WT and mutant FOXA1 proteins were robustly expressed, and, as expected, their expression diminished upon treatment with FOXA1-targeting siRNAs (**Supplementary Figure A1-12B,C**). Next, in a nuclear co-immunoprecipitation experiment, we confirmed that the N-terminal FOXA1 antibody precipitated both WT and truncated FOXA1 proteins, while the C-terminal antibody precipitated exclusively the WT protein (**Figure 2-3B**). Upon DHT stimulation, we also found that both FOXA1 variants interacted with the AR-signaling complex (**Supplementary Figure A1-12D**). However, most remarkably, in ChIP-Seq experiments in LAPC4 cells, we found that only the N-terminal antibody detected FOXA1 binding to DNA (**Figure 2-3C**). Even DHT-stimulated LAPC4 cells displayed similar results (**Supplementary Figure A1-13A**). In contrast, and as expected, in the WT FOXA1 LNCaP and C42B cells, the N-terminal and C-terminal FOXA1 cistromes considerably overlapped (**Figure 2-3D** and **Supplementary Figure A1-13B,C**). Thus, in LAPC4 cells, in presence of the P358fs (class2) mutant, WT FOXA1 does not interact with the chromatin.

To control for possible differences in expression, we repeated the ChIP-Seq experiments with viral overexpression of WT FOXA1 in LAPC4. Even with 13-fold overexpression of exogenous WT FOXA1, the endogenous class2 mutant retained its dominance in binding chromatin (**Figure 2-3E** and **Supplementary Figure A1-13D**). Conversely, overexpression of the P358fs mutant in LNCaP cells markedly attenuated the WT FOXA1 cistrome (**Figure 2-3E** and **Supplementary Figure A1-13E**). Next, we asked if cistromic dominance of the P358fs mutant could be due to a higher affinity for the FOXA1 DNA element. Thus, we performed electrophoretic mobility shift assays (EMSA) using the biotin-labeled KLK3 enhancer containing the consensus FOXA1 recognition motif 5'-GTAAACAA-3' (**Supplementary Figure A1-13F**). The recombinant mutant protein showed markedly stronger binding with DNA relative to WT FOXA1 (**Figure 2-3F** and **Supplementary Figure A1-14A**). The presence of super-shifted bands with the V5-tag antibody confirmed that the recombinant FOXA1 proteins bound to the shifted DNA (**Supplementary Figure A1-14B**). Similarly, an increase in binding affinity was noted for additional class2 mutants (**Supplementary Figure A1-14C**).

To explore class2 biology in isogenic models, using the CRISPR-Cas9 technology, we introduced frameshifting mutations in endogenous FOXA1 alleles in 22RV1 cells. We generated multiple heterozygous or homozygous mutant clones, some of which harbored class2 mutations seen in patients (e.g. P359fs, I360*, etc.; **Supplementary Figure A1-15A,B**). ChIP-seq experiments in these models further confirmed the cistromic dominance of distinct class2 mutants (**Figure 2-3G**). Further comparison of the union N-terminal cistromes from 3 WT and 4 class2 mutant clones revealed that class2 mutations induced a 2-fold expansion of the FOXA1 cistrome (**Figure 2-3H**). Notably, this was accompanied by a marked redistribution of AR on the chromatin (**Supplementary Figure A1-15C**). More importantly, knockdown of mutant FOXA1 or AR in these clones significantly attenuated proliferation (**Figure 2-3I**). This strongly suggests that class2 mutants activate a distinct transcriptional program that continues to be essential for the maintenance of advanced AR-dependent prostate cancer. LNCaP CRISPR clones that harbored a distinct set of mutations causing truncation closer to the forkhead domain also confirmed the functional essentiality of class2 mutations (**Supplementary Figure A1-15D**).

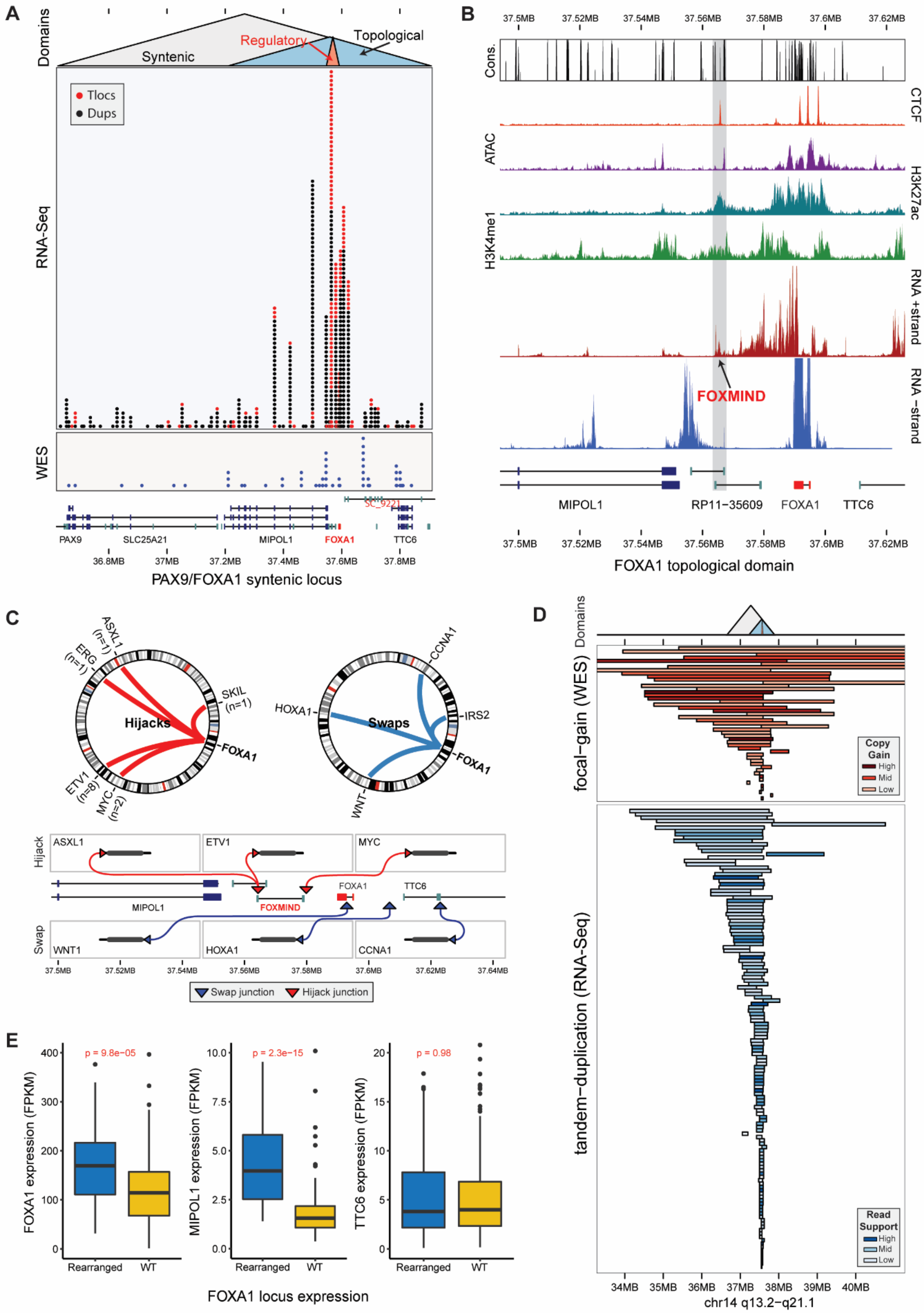
Next, we performed RNA-seq to identify transcriptional changes caused by class2 mutants. Here, we identified 129 up-regulated and 107 down-regulated genes (fold change>3, FDR<0.05) in mutant vs WT clones (**Supplementary Figure A1-16A**). Interestingly, the upregulated genes were significantly enriched for putative LEF1/TCF targets (**Supplementary Figure A1-16B**). LEF1/TCF protein complex is the primary transcriptional mediator of WNT-signaling [44]. This complex remains in an inactive state until bound by nuclear ß-Catenin [39,44]. Thus, we profiled the ß-Catenin expression in a larger panel of class2-mutant clones. Here, we found marked stabilization and accumulation of ß-Catenin in 12 distinct mutant vs 4 WT clones (**Supplementary Figure A1-16C**). Mutant clones also had higher expression of the mesenchymal marker vimentin. This strongly suggests that class2 mutants enable a distinct transcriptional program in collaboration with TFs of the WNT/ß-Catenin pathway.

WNT-signaling is implicated in promoting metastasis and therapy resistance in multiple cancers [39], including prostate cancer [5,45]. Thus, activation of WNT-signaling by class2 mutants is concordant with their acquisition in metastatic prostate cancer. Hence, we assessed the metastatic propensity of the mutant clones. In Boyden chamber assays, three distinct class2 clones showed 2- to 3-fold higher invasiveness compared to WT clones (**Supplementary Figure A1-16D**). More strikingly, in the in vivo zebrafish model [46], we detected metastasis more frequently in fish embryos injected with the mutant clones relative to WT clones, with cells physically invading through the yolk sac, intravasating into systemic vasculature and disseminating into different parts of the body (**Figure 2-3J,K**). Fish injected with the normal HEK293 cells showed a negligible incidence of metastasis (negative control; **Figure 2-3J**). Furthermore, fish injected with mutant clones had more aggressive metastasis with a significantly greater number of disseminated foci per fish (**Figure 2-3K,L**). Together, these functional data demonstrate that class2 mutants displace WT FOXA1 from the chromatin to activate a WNT/ß-Catenin-mediated transcriptional program that promotes metastasis.

Class 3 rearrangements drive FOXA1 overexpression

To understand the structural and functional consequences of *FOXA1* rearrangements, we mapped them onto chromatin domains and cis-regulatory elements. First, we noted that the *FOXA1* locus is highly syntenic (top panel, **Figure 2-4A**), with a linear structure preserved across the deuterostome superphylum [47]. In humans, the *PAX9/FOXA1* syntenic block comprises two topologically associating domains (TAD) organizing *PAX9* and *FOXA1* into separate chromatin interacting units (**Supplementary Figure A1-17**). Genes within these TADs had tissue-specific expression patterns, with *FOXA1* TAD robustly activated in the prostate (**Supplementary Figure A1-18A**). In prostate cancer, almost all breakpoints in the *FOXA1* locus clustered within the *FOXA1* TAD (**Figure 2-4A**). Most strikingly, among the *FOXA1* TAD-encompassing genes, the non-coding transcript *RP11-356O9.1* was expressed exclusively in the prostate (**Supplementary Figure A1-18A**). Upon closer inspection, we found that *RP11-356O9.1* was the spliced product of bi-directional transcription from ultra-highly conserved elements (UCEs), which were marked by active enhancer histone marks (**Figure 2-4B**) [47]. Expression of the dominant unspliced *RP11-356O9.1* (Methods) also most strongly correlated with *FOXA1* (rho=0.44, p=6.1e-08) and TTC6 expression levels (rho=0.68, p=4.4e-15) in patient samples (**Supplementary Figure A1-18B**). *FOXA1* and the transcribed UCEs were further circumscribed by CTCF binding-sites, indicative of a regulatory chromatin loop. Overall, these data implicate UCEs as a transcriptional topological anchor point for the *FOXA1* TAD [48], which we hereafter refer to as *FOXMIND* (FOXA1 Mastermind; **Figure 2-4B**).

This refined topological insight prompted us to examine the recurrent structural variation in terms of preservation/disruption of the putative *FOXA1* regulatory domain. Strikingly, we found that breakpoints from translocations were largely confined within a precise (50 kb) region between *FOXA1* and the 3’ UTR of *MIPOL1*, while breakpoint junctions from duplications had a broader distribution across the TAD (middle panel, **Figure 2-4A**). For translocations involving established cancer genes, we delineated two mechanistic patterns: 1) hijacking of the *FOXMIND* enhancer and 2) inserting upstream of the bi-directional *FOXA1* promoter **(Figure 2-4C**). The first pattern subsumes previously reported *ETV1* [49] and *SKIL* [50] fusions, a novel *ASXL1* fusion, and recurrent translocations with the *MYC* locus. The second pattern inserts an oncogene, such as *CCNA1, WNT1,* or *HOXA1*, upstream of the *FOXA1* promoter. Both mechanisms frequently result in the outlier expression of the translocated gene (**Supplementary Figure A1-19**). Next, we interrogated duplications from CNVs and RNA fusions to delineate consensus positioning of the focal rearrangements (Methods). We found the minimally amplified region to coincide with the *FOXMIND* regulatory domain, with a sharp boundary at the bi-directional *FOXA1* promoter (**Figure 2-4D**).



A
Domains
Syntenic
Regulatory
Topological
RNA-Seq
● Tlocs
● Dups
WES
PAX9
SLC25A21
MIPOL1
FOXA1
TTC6
36.8MB
37MB
37.2MB
37.4MB
37.6MB
37.8MB
PAX9/FOXA1 syntenic locus
B
37.5MB
37.52MB
37.54MB
37.56MB
37.58MB
37.6MB
37.62MB
Cons.
CTCF
ATAC
H3K27ac
H3K4me1
RNA +strand
RNA −strand
FOXMIND
MIPOL1
RP11−35609
FOXA1
TTC6
FOXA1 topological domain
C
ASXL1 (n=1)
ERG (n=1)
SKIL (n=1)
Hijacks
FOXA1
ETV1 (n=8)
MYC (n=2)
CCNA1
HOXA1
Swaps
IRS2
WNT
Hijack
ASXL1
ETV1
MYC
MIPOL1
FOXMIND
FOXA1
TTC6
Swap
WNT1
HOXA1
CCNA1
37.5MB
37.52MB
37.54MB
37.56MB
37.58MB
37.6MB
37.62MB
37.64MB
Swap junction
Hijack junction
D
Domains
focal-gain (WES)
Copy Gain
High
Mid
Low
tandem−duplication (RNA-Seq)
Read Support
High
Mid
Low
34MB
35MB
36MB
37MB
38MB
39MB
40MB
chr14 q13.2−q21.1
E
FOXA1 expression (FPKM)
p = 9.8e−05
0
100
200
300
400
Rearranged
WT
MIPOL1 expression (FPKM)
p = 2.3e−15
0.0
2.5
5.0
7.5
10.0
Rearranged
WT
TTC6 expression (FPKM)
p = 0.98
0
5
10
15
20
Rearranged
WT
FOXA1 locus expression

Figure 2-4: Genomic characterization of Class3 rearrangements of the FOXA1 locus. (A) Distribution of breakends (chimeric junctions or copy-number segment boundaries) in relation to the syntenic, topological, and regulatory domains of FOXA1. (B) Functional genomic characterization of the FOXA1 locus. Cons. - phastCons base-level conservation; ATAC - DNA accessibility signal; H3K27ac / H3K4me1 / CTCF - ChIP-seq normalized signal in prostate cell lines. Representative mCRPC RNA-seq coverage on the "+" and "-" strand. (C) Structural patterns of translocations. Hijack translocations occur between FOXMIND and FOXA1, swap translocations occur upstream of the FOXA1 promoter. (D) Inferred duplications within the FOXA1 locus based on RNA-seq (tandem breakends) and WES (copy-gains). (E) Transcriptional changes in protein-coding genes within the FOXA1 TAD (t-test).

FOXA1 is haploinsufficient in prostate development [1] and functionally essential in prostate cancer (**Supplementary Figure A1-1**), which prompted us to examine how copy-number in general, and focal duplications in particular, impact its expression. Strikingly, we found *FOXA1* levels to be poorly correlated with copy-number (**Supplementary Figure A1-20A**), but highly sensitive to focal structural variation (**Figure 2-4E**). Tandem duplications, ascertained at the RNA (**Figure 2-4E**) and DNA levels (**Supplementary Figure A1-20B**), significantly increased *FOXA1* and *MIPOL1*, but not *TTC6* expression. Surprisingly, translocations resulted in a decrease in *FOXA1* levels (**Supplementary Figure A1-20B**), despite a significant co-occurrence (odds-ratio=3.89, p=2.8e-6) with tandem-duplications (**Supplementary Figure A1-20C**). To explore this further, we carried out haplotype-resolved linked-read sequencing of MDA-prostate cancer-2B, a prostate cancer cell line with the *FOXMIND-ETV1* fusion. The oncogenic translocation was accompanied by focal tandem duplication of the non-translocated allele (**Supplementary Figure A1-21A**). Intriguingly, the translocated *FOXA1* allele was shut down resulting in monoallelic FOXA1 transcription (**Supplementary Figure A1-21B**), without net-loss of FOXA1 expression (266 FPKM, 95th percentile compared to mCRPC). This suggests that translocations result in a loss of *FOXA1* expression from the adjacent allele, which is overall "rescued" by activating tandem-duplications in the other allele. Based on these data, we propose a coalescent model wherein class3 structural variants re-position *FOXMIND* to drive the expression of *FOXA1* or other putative oncogenes.

Discussion

The *FAST*, *FURIOUS*, and *LOUD* aberrations of FOXA1

By leveraging an aggregate prostate cancer cohort of over 1500 cases, we were able to identify three previously undescribed structural classes of FOXA1 aberrations that diverged in genetic associations and oncogenic functions. Hence, we establish FOXA1 as a potent oncogene that is altered by gain-of-function genetic aberrations in AR-dependent prostate cancer. Overall, FOXA1 aberrations are observed in 34.6% of mCRPC (**Figure 2-5A**). Class1 mutants, recurrent at approximately 9%, originate in primary prostate cancer that lack other primary driver alterations5. Contrarily, class2 mutants recur at 4% and are detected at clonal levels only in metastatic prostate cancer. Finally, class3 genomic rearrangements are significantly enriched in metastatic disease and, within the limitations of RNA-seq and WES, are observed in 20-30% of cases. Notably, class3 tandem-duplications frequently co-occur with class2 mutations (odds-ratio=3.95, p=0.014).

Class1 FOXA1 mutations disrupt the Wing2 secondary structure and increase the transactivational ability of FOXA1 towards oncogenic AR-signaling (**Figure 2-5B**). Notably, in line with mutual exclusivity at the genetic level, class1 mutations and PI3K alterations transcriptionally converge to induce similar gene expression changes. This is a novel association and it further emphasizes the significance of PI3K-AKT-mTOR signaling in prostate cancer biology. In contrast, class2 aberrations truncate the C-terminus and impart cistromic dominance to potentiate WNT/ß-Catenin signaling that promotes metastasis in vivo (**Figure 2-5C**). This also attributes a unique role to the C-terminal domain of FOXA1 in hindering its interaction with chromatin. We demonstrate both class1 and class2 mutations to be neomorphic (i.e., result in a novel gain-of-function). Class3 rearrangements fall into two structural patterns: translocations that place putative oncogenes in the proximity of *FOXMIND* (i.e. the *FOXA1* enhancer), and duplications that preserve and amplify the *FOXMIND-FOXA1* regulatory domain (**Figure 2-5D**). Class3 rearrangements provide a mechanism to increase FOXA1 expression levels to supra-physiological levels (i.e. not observed in normal tissues). Given the unique traits, we propose to refer to these classes as ‘FAST’ (class1), ‘FURIOUS’ (class2), and ‘LOUD’ (class3).

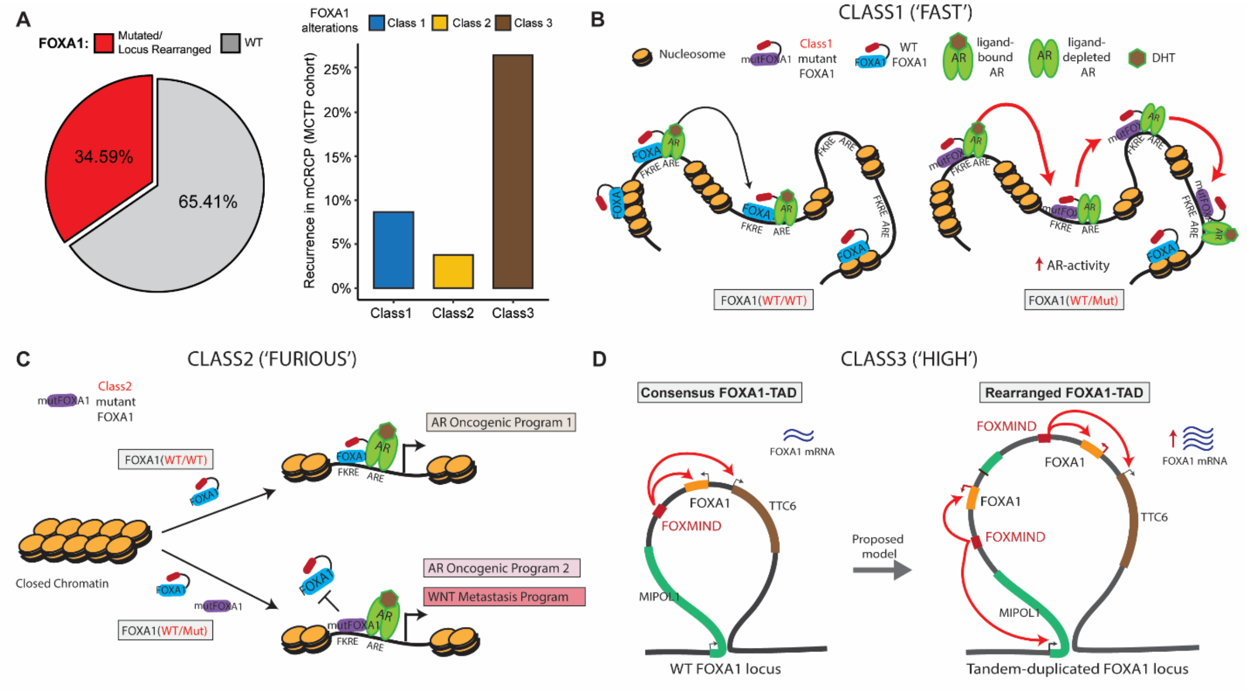


Figure 2-5: Summary of FOXA1 alterations and proposed models of function in prostate cancer progression. (A) Integrated (RNA-seq and WES) recurrence of distinct classes of FOXA1 alterations in mCRPC. (B) Class1 model: Wing2-disrupted FOXA1 shows increased chromatin mobility, resulting in transcriptional activation of AR targets and multiple oncogenic programs. (C) Class2 model: Truncated FOXA1 shows increased chromatin binding, resulting in AR-cistrome redistribution and increased WNT signaling. (D) Class3 model: Tandem-duplications within the FOXA1 TAD reposition FOXMIND to drive FOXA1 expression to supra-physiological levels.

Recent sequencing studies have uncovered activating alterations in FOXA1 regulatory regions in estrogen receptor (ER)-positive breast cancer, including hotspot mutations in the FOXA1 promoter [51] and cancer-predisposing SNPs at FOXA1 binding sites [52,53]. Here, we show that the FOXA1 locus is structurally rearranged in breast cancer as well. FOXA1 mutations are also observed in bladder cancer and AR-positive salivary gland tumors [7]. Thus, the three classes of FOXA1 alterations might be distinctly activating in other tumors with oncogenic AR or ER signaling.

In summary, our study presents genetic, mechanistic, and phenotypic evidence supporting FOXA1 mutations as activating driver alterations in AR-driven prostate cancer, and likely in other hormone-receptor related cancers. Notably, we propose a mechanism-based classification for FOXA1 alterations which includes increased FOXA1 mobility (Class1, 'Fast'), enhanced chromatin binding (Class2, 'Furious'), and rearrangement-mediated overexpression (Class3, 'Loud'). For the first time, we describe structural variation as an important mechanism of FOXA1 pathogenic function and delineate *FOXMIND* as a critical enhancer element recurrently altered in cancer. The high prevalence of *FOXA1* alterations (upwards of 35%), further emphasizes that mCRPC is a disease of aberrant transcription factor activity, already exemplified by both ETS fusions and AR alterations present in over half of mCRPC4. Considering the strong oncogenic effects of *FOXA1* alterations demonstrated in this study, co-targeting FOXA1 is an attractive therapeutic approach for mCRPC as well as other FOXA1-driven cancers.

Materials and methods

Cell culture

Most cell lines were originally purchased from the American Type Culture Collection (ATCC) and were cultured as per the standard ATCC protocols. LNCaP-AR and LAPC4 cells were gifts from Dr. Charles Sawyers' lab (Memorial Sloan-Kettering Cancer Center, New York, NY). Until otherwise stated, for all the experiments LNCaP, PNT2, LNCaP-AR, C42B, 22RV1, DU145, PC3 cells were grown in the RPMI 1640 medium (Gibco) and VCaP cells in the DMEM with Glutamax

(Gibco) medium supplemented with 10% Full Bovine Serum (FBS; Invitrogen). LAPC4 cells were grown in IMEM (Gibco) medium supplemented with 15%FBS and 1nM of R1881. Immortalized normal prostate cells: RWPE1 were grown in keratinocyte media with regular supplements (Lonza); PNT2 were grown in RPMI medium with 10%FBS. HEK293 cells were grown in DMEM (Gibco) medium with 10% FBS. All cells were grown in a humidified 5%CO2 incubator at 37C. All cell lines were biweekly tested to be free of mycoplasma contamination and genotyped every month at the University of Michigan Sequencing Core using Profiler Plus (Applied Biosystems) and compared with corresponding short tandem repeat (STR) profiles in the ATCC database to authenticate their identity in culture between passages and experiments.

Antibodies

For immunoblotting, the following antibodies were used: FOXA1_N-terminal (Cell Signaling Technologies: 58613S; Sigma-Aldrich: SAB2100835); FOXA1_C-terminal (ThermoFisher Scientific: PA5-27157; Abcam: ab23738); AR (Millipore: 06-680); LSD1 (Cell Signaling Technologies: 2139S); Vinculin (Sigma Aldrich: V9131); H3 (Cell Signaling Technologies: 3638S); GAPDH (Cell Signaling Technologies: 3683); B-Actin (A5316); B-Catenin (8480S); Vimentin(5741S) and V5-tag (R960-25).

For co-immunoprecipitation and ChIP-Seq experiments, the following antibodies were used: FOXA1_N-terminal (Cell Signaling Technologies: 58613S); FOXA1_C-terminal (ThermoFisher Scientific: PA5-27157); AR (Millipore: 06-680); V5-tag (R960-25).

Immunoblotting and nuclear co-immunoprecipitation

Cell lysates were prepared using the RIPA lysis buffer (ThermoFisher Scientific) and denatured in the complete NuPage 1X LDS/reducing agent buffer (Invitrogen) with 10 minutes of heating at 70C. 10-25ug of total protein was loaded per well, separated on 4–12% SDS polyacrylamide gels (Novex), and transferred onto a nitrocellulose membrane (Thermo Fisher Scientific) using a semi-dry transfer system (BioRad) at 25V for 1h. The membrane was incubated for 1 hour in blocking buffer (Tris-buffered saline, 0.1% Tween (TBS-T), 5% nonfat dry milk) and incubated overnight at 4°C with primary antibodies. Host species-matched secondary antibodies conjugated to horseradish peroxidase (HRP; BioRad) were used at 1/20,000 dilution to detect primary antibodies and blots were developed using enhanced chemiluminescence (ECL Prime, Thermo Fisher Scientific) following the manufacturer's protocol.

For nuclear co-immunoprecipitation assays, 8-10 million cells were fractionated to isolated intact nuclei using the NE-PER kit reagents (Thermo Fisher Scientific) and lysed in the complete IP lysis buffer (Thermo Fisher Scientific). Nuclear lysates were incubated for 2 hours at 4C with 30ul of magnetic Protein-G Dynabeads (Thermo Fisher Scientific) for pre-clearing. A fraction of the cleared lysate was saved as input and the remainder was incubated overnight (12-16 hours) with 10ug of target protein antibody at 4C with gentle mixing. The next day, 50ul of Dynabeads Protein-G beads were added to the lysate-antibody mixture and incubated for 2h at 4C. Beads were washed 3 times with IP buffer (150nM NaCl; Thermo Fisher Scientific) and directly boiled in 1X NuPage LDS/reducing agent buffer (ThermoFishner Scientific) to elute and denature the precipitated proteins. These samples were then immunoblotted as described above with the exception of using ProteinA-HRP secondary (GE HealthCare) antibody for detection.

RNA extraction and quantitative polymerase chain reaction

Total RNA was extracted using the miRNeasy Mini Kit (Qiagen), with the inclusion of an on-column genomic DNA digestion step using the RNase-free DNase kit (Qiagen), following the standard protocols. RNA was quantified using the NanoDrop 2000 Spectrophotometer (ThermoFisher Scientific) and 1ug of total RNA was used for complementary DNA (cDNA) synthesis using the SuperScript III Reverse Transcriptase enzyme (ThermoFisher Scientific) according to the manufacturer's instructions. 20ng of cDNA was inputted per polymerase chain reaction (PCR) using the FAST SYBR Green Universal Master Mix (ThermoFisher Scientific) and every sample was quantified in triplicates. Gene expression was calculated relative to GAPDH and HPRT1 (loading control) using the delta-delta Ct method and normalized to the control group for graphing. Quantitative PCR (qPCR) primers were designed using the Primer3Plus tool (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) and synthesized by Integrated DNA Technologies. Primers used in this study are listed below:

GAPDH: F, TGCACCACCAACTGCTTAGC; R, GGCATGGACTGTGGTCATGAG;

HPRT1: F, AGGCGAACCTCTCGGCTTTC; R, CTAATCACGACGCCAGGGCT;

B-Actin: F, AGGATGCAGAAGGAGATCACTG; R, AGTACTTGCGCTCAGGAGGAG;

AR: F, CAGTGGATGGGCTGAAAAAT; R, GGAGCTTGGTGAGCTGGTAG;

FOXA1-3': F, GAAGACTCCAGCCTCCTCAACTG; R, TGCCTTGAAGTCCAGCTTATGC;

FOXA1-5': F, CTACTACGCAGACACGCAGG; R, CCGCTCGTAGTCATGGTGTT.

siRNA-mediated gene knockdown

Cells were seeded in a 6-well plate at the density of 100,000-250,000 cells per well. After 12 hours, cells were transfected with 25nM of gene-targeting ON-TARGETplus SMARTpool siRNAs or non-targeting pool siRNAs as a negative control (Dharmacon) using the RNAiMAX reagent (Invitrogen) for two consecutive days, following manufacturer's instructions. Both total RNA and protein were extracted on day 3 (total 72h) to confirm efficient (>80%) knockdown of the target genes. For crystal violet staining, at day9 growth medium was aspirated and cells were first fixed with 4% formaldehyde solution, followed by 30-minute incubation in 0.5% crystal violet solution in 20% methanol and scanned. Catalogue numbers of siRNA SMARTpools (Dharmacon) used are: Non-targeting control (D-001810-10-05); AR (L-003400-00-0005); FOXA1 (L-010319-00-0005).

CRISPR sgRNA-mediated gene knockout

Cells were seeded in a 6-well plate at the density of 200,000-300,000 cells per well and infected with viral particles with lentiCRISPR-V2 plasmids coding either non-targeting (sgNC) or sgRNAs targeting the Exon1 or the Forkhead domain of FOXA1 (both ensuing in FOXA1 inactivation). This was followed by 3 days of puromycin selection after which proliferation assays were carried out as described below. The lentiCRISPR-V2 vector was a gift from Dr. Feng Zhang’s lab (Addgene plasmid # 52961). All sgRNAs were designed using the online Benchling tool (https://benchling.com/) and were synthesized as oligos from Integrated DNA Technologies. sgRNA sequences used in this study are: sgNC#1 5’-GTAGCGAACGTGTCCGGCGT-3’; sgNC#2; 5’-GACCGGAACGATCTCGCGTA-3’ sgFOXA1_Exon1: 5’-GTAGTAGCTGTTCCAGTCGC-3’; sgFOXA1_Forkhead: 5’-GCCGTTCTCGAACATGTTGC-3’.

Proliferation assays

For siRNA growth assays, cells were directly plated in a 96-well plate at the density of 2,500-8,000 per well and transfected with gene-specific or non-targeting siRNAs as described above on Day0 and Day1. Every treatment was carried out in six independent replicate wells. CellTiter-Glo reagent (Promega) was used to assess cell viability at multiple time points post-transfection following the manufacturer's protocol. Data was normalized to siNC-Day1 readings and plotted as relative cell viability to generate growth curves.

Alternatively, for CRISPR-sgRNA growth assays, cells were treated as described above for target gene inactivation and seeded into a 24-well plate at 20,000 cells/well density with 2 replicates per group. After 12 hours, plates were placed into the IncuCyte live-cell imaging machine (IncuCyte) set at the phase-contrast option to record cell confluence every 3 hours for up to 7-9 days. Similarly, for class1 growth assays (Figure 2-2h), stable doxycycline-inducible 22RV1 cells were grown in with 10% charcoal-stripped serum (CSS)-supplemented medium for 48 hours. Androgen starved cells were then seeded into 96-wells at 5000 cells/well density in 10%CSS medium with or without the addition of doxycycline (1ug/ml) to induce control or mutant protein expression (6 replicates/group). Once adherent, treated cells were placed in the IncuCyte live-cell imaging incubator (IncuCyte), set at the phase-contrast option, to record cell confluence every 3 hours for up to 7-9 days. For all IncuCyte assays, confluence measurements from all time points were normalized to the measurement at 0 hours and plotted as relative confluence to generate growth curves.

Cloning of representative FOXA1 mutants

WT FOXA1 coding sequence was purchased from Origene (Cat#: SC108256) and cloned into the pLenti6/V5 lentiviral vector (ThermoFisher Scientific) using the standard TOPO cloning protocol. Class1 missense mutations (I176M; H247Q and R261G) were engineered from the WT FOXA1 vector using the QuikChange II XL Site-Directed Mutagenesis Kit (Agilent Tech) as per the manufacturer's instructions. All point mutations were confirmed using Sanger sequencing through the University of Michigan Sequencing Core Facility. Edited plasmids were further transfected in HEK293 cells to confirm the robust expression of the mutant protein. For truncated class2 variants, the WT coding sequence up to the amino acid before the intended mutation was cloned. All FOXA1 variants had the V5-tag fused on the C-terminus. Also, select mutants were cloned into a doxycycline-inducible vector (pCW57.1) to generate stable lines. For FRAP assays, the pCW57.1 vector was edited to incorporate an in-frame GFP coding sequence at the C-terminal end.

Fluorescent recovery after photobleaching (FRAP) assay and data quantification

PNT2 cells were seeded in a 6 well plate at 200,000 cells/well density and transfected with 2ug of doxycycline-inducible vectors coding different FOXA1 variants fused to GFP on the C-terminal end. After 24 hours, cells were plated in the glass-bottom microwell dishes (MatTek: #P35G-1.5-14-C) in a phenol-free growth medium supplemented with doxycycline (1ug/ml). Cells were then incubated for 48 hours to allow for robust expression of the exogenous GFP-tagged protein and strong adherence to the glass surface. Microwell dishes were placed in a humidity control chamber set at 37C (Tokai-Hit) and mounted on the SP5 Inverted 2-Photon FLIM Confocal microscope (Leica). FRAP Wizard from the Leica Microsystems software suite was used to conduct and analyze FRAP experiments. Roughly half of the nucleus was photobleached using the Argon-laser at 488nm and 100% intensity for 20-30 iterative frames at 1.2 second intervals. Laser intensity was reduced to 1% for imaging post bleaching. Immediately after photobleaching, 2 consecutive images were collected at 1.2 second intervals followed by images taken at 10 seconds intervals for 60 frames (i.e. 10 minutes).

For data analyses, recovery of signal in the bleached half and loss of signal in the unbleached half were measured as average fluorescence intensities in at least 80% of the respective areas, excluding the immediate regions flanking the separating border. All intensity curves were generated from background-subtracted images. The fluorescence signal measured in a region-of-interest (ROI) was normalized to the signal prior to bleaching using the following formula[54]:

$$R = (It - Ibg)/(Io-Ibg)$$

Where ‘Io’ is the average intensity in the ROI before bleaching, ‘It’ is the average intensity in the ROI at any time-point post-bleaching, and ‘Ibg’ is the background fluorescence signal in a region outside of the cell nucleus. Raw recovery kinetic data from above were fitted with the best hyperbolic curves using the GraphPad Prism software and time to recover to 50% signal was calculated from the resulting best-fit equations.

Dual-luciferase AR reporter assay

HEK293 cells stably overexpressing the WT AR protein (i.e. HEK293-AR) were used for the AR reporter assays. HEK293-AR cells were seeded in a 12-well plate at 300,000 cell/well density and transfected with 2ug of the pLenti6/V5 vector coding different FOXA1 variants or GFP (control). After 8 hours, the medium was replaced with a 10%-CSS-supplemented phenol-free medium (androgen depleted) and cells were transfected with the AR-reporter Firefly luciferase or negative control constructs from the Cignal AR-Reporter(luc) kit (QIAGEN; Cat# CCS-1019L) as per the manufacturer’s instructions. Both constructs were premixed with constitutive Rinella luciferase vector as control. After 12 hours, cells were treated with different dosages of dihydrotestosterone (DHT) or enzalutamide (at 10uM dosage); and additional 24 hours later dual luciferase activity was recorded for every sample using the Dual-Glo Luciferase assay (Promega; E2980) and luminescence plate reader (Promega-GLOMAX-Multi Detection System). Each treatment condition had 4 independent replicates. Firefly luciferase signals were normalized with the matched Rinella luciferase signals to control for variable cell number and/or transfection efficiencies, and normalized signals were plotted relative to the negative control reporter constructs.

Electrophoretic mobility shift assay (EMSA)

HEK293 cells were plated in 10cm dishes at 1M/plate density and transfected with 10ug of the pLenti6/V5 vector coding GFP (control) or different FOXA1 variants. After 48 hours, cells were trypsinized and nuclear lysates were prepared using the NE-PER kit reagents (ThermoFisher Scientific). Immunoblots were run to confirm comparable expression of recombinant FOXA1 variants in 2ul (i.e. equal volume) of final nuclear lysates. Next, FOXA1 and AR ChIP-Seq data was used to identify the KLK3 enhancer element. 60bp of the KLK3 enhancer, centered at the FOXA1 consensus motif 5’-GTAAACAA-3’, was synthesized as single stranded oligos (IDT) and biotin-labeled using the Biotin 3’-End DNA labeling kit (ThermoFisher Scientific) and then annealed to generate a labeled double-stranded DNA duplex.

Binding reactions were carried out in 20ul volumes containing 2ul of the nuclear lysates, 50ng/uL poly(dI.dC), 1.25% glycerol, 0.025% Nonidet P-40 and 5mM MgCl2. 10fmol of biotin-labeled KLK3 enhancer probe was added at the very end with gentle mixing. Reactions were incubated for 1h at room temperature, size-separated on a 6% DNA retardation gels (100V for 1h; Invitrogen) in 0.5X TBE buffer, and transferred on the Biodyne Nylon membrane (0.45um; ThermoFisher Scientific) using a semi-dry system (BioRad). Transferred DNA was cross-linked to the membrane using the UV-light at 120mJ/cm2 for 1 minute. Biotin-labeled free and protein-bound DNA was detected using HRP-conjugated streptavidin (ThermoFisher Scientific) and developed using chemiluminescence according to the manufacturer's protocol.

Generation of CRISPR clones and stable lines

22RV1 or LNCaP cells were seeded in a 6-wells plate at 200,000 cells/well density and transiently transfected with 2.5ug of lentiCRISPR-V2 (Addgene: #52961) vector encoding the Cas9 protein and sgRNA cutting either at amino acid 271 (5'-GTCAAGTGCGAGAAGCAGCCG-3') or 359 (5'-GCCGGGCCCGGAGCTTATGGG-3') in Exon2 of FOXA1. Cells were treated with a non-targeting control sgRNA (5'-GACCGGAACGATCTCGCGTA-3') vector to generate isogenic WT clones. Transfected cells were selected with puromycin (Gibco) for 3-4 days and FACS-sorted as single cells into 96-well plates. Cells were maintained in 96-wells for 4-6 weeks with the replacement of the growth medium every 7 days to allow for the expansion of clonal lines. Clones that successfully seeded, were further expanded and genotyped for FOXA1 using Sanger sequencing and immunoblotting with the N-terminal FOXA1 antibody. Sequence and expression validated 22RV1 and LNCaP clones with distinct class2 mutations were used for growth, invasion and metastasis assays as described.

To generate stable cells, doxycycline-inducible vectors coding different FOXA1 variants or GFP (control) were packaged into viral particles through the University of Michigan Vector Core. prostate cancer cells were seeded in a 6-well plate at 100,000-250,000cells/well density and infected with 0.5ml of 10X viral titers. This was followed by 3-4 days of puromycin (Gibco) selection to generate stable lines.

Matrigel invasion assay

22RV1 CRISPR clones were grown in a 10% CSS-supplemented medium for 48 hours for androgen starvation. Special matrigel-coated invasion chambers were used that were additionally coated with a light-tight polyethylene terephthalate membrane to allow for fluorescent quantification of the invaded cells (Biocoat: 24-well format, #354166). 50,000 starved cells were resuspended in a serum-free medium and were added to each invasion chamber. 20% FBS-supplemented medium was added to the bottom wells to serve as a chemoattractant. After 12 hours, medium from the bottom well was aspirated and replaced with 2ug/ml Calcein-green AM dye (ThermoFisher Scientific; C3100MP) in 1X Hank's Balanced Salt Solution (Gibco) and incubated for 30 minutes at 37C. Invasion chambers were then placed in a fluorescent plate reader (Tecan-Infinite M1000 PRO) and the fluorescent signal from the invaded cells at the bottom was averaged from 16 distinct regions per chamber to determine the extent of invasion.

Chromatin immunoprecipitation (ChIP) with massively parallel DNA sequencing

ChIP experiments were carried out using the HighCell# ChIP-Protein G kit (Diagenode) as per the manufacturer’s protocol. Chromatin from 5M cells was used per ChIP reaction with 6.5ug of the target protein antibody. Briefly, cells were trypsinized and washed twice with 1XPBS, followed by cross-linking for 8 min in 1% formaldehyde solution. Crosslinking was terminated by the addition of 1/10 volume 1.25 M glycine for 5 min at room temperature followed by cell lysis and sonication (Bioruptor, Diagenode), resulting in an average chromatin fragment size of 200 bp. Fragmented chromatin was then used for immunoprecipitation using various antibodies with an overnight incubation at 4C. ChIP DNA was de-crosslinked and purified using the iPure Kit V2 (Diagenode) using the standard protocol. Purified DNA was then prepared for sequencing as per the manufacturer's instructions (Illumina). ChIP samples (1–10 ng) were converted to blunt-ended fragments using the T4 DNA polymerase, *E. coli* DNA polymerase I large fragment (Klenow polymerase), and T4 polynucleotide kinase (New England BioLabs (NEB)). A single A base was added to fragment ends by Klenow fragment (3′ to 5′ exo minus; NEB) followed by ligation of Illumina adaptors (Quick ligase, NEB). The adaptor-ligated DNA fragments were enriched by PCR using the Illumina Barcode primers and Phusion DNA polymerase (NEB). PCR products were size-selected using 3% NuSieve agarose gels (Lonza) followed by gel extraction using QIAEX II reagents (Qiagen). Libraries were quantified and quality checked using the Bioanalyzer 2100 (Agilent) and sequenced on the Illumina HiSeq 2000 Sequencer (125-nucleotide read length).

Zebrafish embryo metastasis experiment

All zebrafish experiments were carried out in designated University of Michigan Animal Facilities following standard IACUC guidelines. Wildtype AB^{TL} zebrafish were maintained in aquaria according to standard protocols[2]. Embryos were generated by natural pairwise mating and raised at 28.5°C on a 14h light/10h dark cycle in a 100 mm petri dish containing aquarium water with methylene blue to prevent fungal growth. Cell injections were carried out as described in this study[46]. Briefly, GFP-expressing normal (control) or cancer cells were resuspended in PBS at the concentration of $1x10^7$ cells/ml. 48 hours post-fertilization, wild-type embryos were dechorionated and anesthetized with 0.04 mg/ml tricaine. Approximately 10 nl (approx. 100 cancer cells) were microinjected into the perivitelline space using a borosilliac micropipette tip with filament. Embryos were returned to aquarium water and washed twice to remove tricaine, then moved to a 96 well plate with one embryo per well and kept at 35°C for the duration of the experiment. All embryos were imaged at 24 hours intervals to follow metastatic dissemination of injection cells. Water was changed daily to fresh aquarium water. More than 30 fish were injected for each condition (WT#2, n=30; WT#5, n=50; #57, n=35; #84, n=57; #113, n=38) and metastasis was visually assessed daily up to 5 days after injection. Embryos were either imaged directly in the 96 well plates or placed onto a concave glass slide to capture representative images using a fluorescent microscope (Olympus-IX71). For quantification, evidently distinct cell foci in the embryo body were counted 72 hours after the injections.

ChIP-seq data analysis

Paired-end 125bp reads were trimmed and aligned to the GRCh38 human reference using the STAR (version 2.4.0g1) aligner with splicing disabled, the resulting reads were filtered using samtools “samtools view -@ 8 -S -1 -F 384”. The resulting BAM file was sorted and duplicate marked using novosort and converted into bigwig files for visualization using “bedtools genomecov -bg -split -ibam” and “bedGraphToBigWig”. The coverage signal was normalized to total sequencing depth / 1e6 reads. Peak calling was performed using MACS2 with the following settings “macs2 callpeak --call-summits --verbose 3 -g hs -f BAM -n OUT --qvalue 0.05”. ChIP peak profile plots and read-density heatmaps were generated using deepTool2[55] and cistrome overlap analyses were carried out using the ChIPpeakAnno[56] package in R.

Utilized cohorts, data sets, and resources

This study leverages previously published public or restricted patient genetic data. Genetic calls for primary prostate cancer and breast cancer (BCa) were obtained from the Genomic Data Commons (GDC)[57] for the prostate cancer-PRAD[4] and BCa-BRCA[6,58] cohorts, respectively. Raw RNA-seq data (paired-end reads from unstranded polyA libraries) for those samples were downloaded from the GDC and processed with our standard Clinical RNA-seq Pipeline CRISPR/CODAC (see below). For the TCGA PRAD and BRCA cohorts, we downloaded mutational calls from multiple sources (GDC, cBio Portal, UCSC Xena) and additionally used the BAM-slicing tool to download sequence alignments from whole-exome sequencing libraries to the FOXA1 locus. We then used our internal pipeline (see below) to call SNVs and indels within FOXA1. We also used the downloaded aligned data for the manual review of FOXA1 mutation calls. Mutation calls for advanced primary and metastatic cases were obtained from the MSK-IMPACT cohort (downloaded from the cBio portal[59]). The main MCTP mCRPC cohort includes 360 cases reported previously (the location of all raw BAM files is provided in (Wu et al., 2018 in press), the 10 additional mCRPC cases included herein but not in Wu et al. are being included in the Database of Genotypes and Phenotypes (dbGaP): phs000673.v3.p1, and belong to a continuous sequencing program with the same IRB-approved protocol (MI-Oncoseq program, University of Michigan Clinical Sequencing Exploratory Research). The genetic sequencing data (WXS) for rapid autopsy cases is available from dbGaP: hs000554.v1.p1and phs000567.v1.p1. De-identified somatic mutation calls, RNA-seq fusion calls, processed/segmented copy-number data, and RNA-seq expression matrices across the full MCTP mCRPC 370 case-cohort is available on request from the authors.

Preparation of WES and RNA-seq libraries

Integrative clinical sequencing, comprising exome sequencing and polyA and/or capture RNA-seq, was performed using standard protocols in our Clinical Laboratory Improvement Amendments (CLIA) compliant sequencing lab. In brief, tumor genomic DNA and total RNA were purified from the same sample using the AllPrep DNA/RNA/miRNA kit (QIAGEN). Matched normal genomic DNA from blood, buccal swab, or saliva was isolated using the DNeasy

Blood & Tissue Kit (QIAGEN). RNA sequencing was performed by exome-capture transcriptome platform[60]. Exome libraries of matched pairs of tumor/normal DNAs were prepared as described before[61], using the Agilent SureSelect Human All Exon v4 platform (Agilent). All the samples were sequenced on the Illumina HiSeq 2000 or HiSeq 2500 (Illumina Inc) in paired-end mode. The primary base call files were converted into FASTQ sequence files using the bcl2fastq converter tool bcl2fastq-1.8.4 in the CASAVA 1.8 pipeline.

Analysis of whole-exome sequencing data

The FASTQ sequence files from whole-exome libraries were processed through an in-house pipeline constructed for analysis of paired tumor/normal data. The sequencing reads were aligned to the GRCh37 reference genome using Novoalign (version 3.02.08) (Novocraft) and converted into BAM files using SAMtools (version 0.1.19). Sorting, indexing, and duplicate marking of BAM files used Novosort (version 1.03.02). Mutation analysis was performed using freebayes (version 1.0.1) and pindel (version 0.2.5b9). Variants were annotated to RefSeq (via the UCSC genome browser, retrieved on 8/22/2016), as well as COSMIC v79, dbSNP v146, ExAC v0.3, and 1000 Genomes phase 3 databases using snpEff and snpSift (version 4.1g). SNVs and indels were called somatic if they were present with at least 6 variant reads and 5% allelic fraction in the tumor sample, and present at no more than 2% allelic fraction in the normal sample with at least 20X coverage; additionally, the ratio of variant allelic fractions between tumor and normal samples were required to be at least six in order to avoid sequencing and alignment artifacts at low allelic fractions. Minimum thresholds were increased for indels observed to be recurrent across a pool of hundreds of platform- and protocol-matched normal samples. Specifically, for each such indel, a logistic regression model was used to model variant and total read counts across the normal pool using PCR duplication rate as a covariate, and the results of this model were used to estimate a predicted number of variant reads (and therefore allelic fraction) for this indel in the sample of interest, treating the total observed coverage at this genomic position as fixed. The variant read count and allelic fraction thresholds were increased by these respective predicted values. This filter eliminates most recurrent indel artifacts without affecting our ability to detect variants in homopolymer regions from tumors exhibiting microsatellite instability. Germline variants were called using ten variant reads and 20% allelic fraction as minimum thresholds and were classified as rare if they had less than 1% observed population frequency in both the 1000 Genomes and ExAC databases. Exome data was analyzed for copy number aberrations and loss of heterozygosity by jointly segmenting B-allele frequencies and log2-transformed tumor/normal coverage ratios across targeted regions using the DNAcopy (version 1.48.0) implementation of the Circular Binary Segmentation algorithm. The Expectation-Maximization Algorithm was used to jointly estimate tumor purity and classify regions by copy number status. Additive adjustments were made to the log2-transformed coverage ratios to allow for the possibility of non-diploid tumor genomes; the adjustment resulting in the best fit to the data using minimum mean-squared error was chosen automatically and manually overridden if necessary.

Detection of copy-number breakpoints from Whole Exome Sequencing

The output of our clinical WES pipeline includes segmented copy-number data, inferred absolute copy-numbers and predicted parent-specific genotypes (e.g. AAB), detection of loss-of-heterozygosity (LOH), and detection of copy-neutral LOH (uniparental disomy). Together these data enable the detection of joint discontinuities in the copy-number profile (log-ratio and B-allele frequencies) at exon-level resolution. A subset of genomic rearrangements results in changes in copy-number or allelic shifts, and hence the presence of such discontinuities in paired tumor-normal WES data is strongly indicative of a somatic breakpoint. For example, 1 copy-gain will result in a segment with an increased log-ratio, and a corresponding zygosity deviation (see above). This segment will be discontinuous with adjacent segments, which will result in the call of a WES breakend (discontinuity) on either side of the copy-gain. The size of the breakend depends on the density of covered exons and, in general, the resolution is better in genic vs. intergenic regions. We assessed the presence of such breakpoints within the gene-dense and exon-dense FOXA1 locus. All copy-number breakends met statistical thresholds of the CBS algorithm (see above) at either the log-ratio or B-allele level.

Genetic characterization of mCRPC tumors samples at the pathway level

The co-occurrence or mutual exclusivity of FOXA1 aberrations with other previously described genetic events in prostate cancer has been carried out at the pathway level, but grouping putative functionally equivalent (and largely genetically mutually exclusive) events. All known types of ETS fusion (ERG, ETV1, FLI1, ETV4, ETV5) were considered as ETS-positive tumors, PI3K alterations included PTEN homozygous loss, PIK3CA activating mutations, and PIK3R1 inactivating mutations, AR pathway alterations included AR, NCOR1, NCOR2, and ZBTB16 mutations/deletions, but excluded AR amplifications / copy-gains. The KMT category included mutations in all recurrently mutated lysine methyltransferases. The WNT category included inactivating aberrations in APC and activating mutations in CTNNB1. DRD included cases with mutations in BRCA1, BRCA2, PALB2, ATM, all common mismatch repair genes, and CDK12.

Assessment of two-hit (biallelic) alterations

To assess the frequency of genetic inactivations of both alleles we integrated mutational, copy-number, and RNA-seq (fusion) data. A gene was considered having both alleles inactivated for any combination (pair) of the following events: copy-loss, mutation, truncating fusion, copy-number breakpoint, in addition to homozygous deletion of both copies and two independent mutations. Ambiguous cases were manually reviewed to increase the accuracy and ascertain whether both events e.g. copy-number breakpoint and gene fusion are likely independent events.

Unified mutation calling and variant classification of FOXA1

Mutation calls for FOXA1 obtained/downloaded from the GDC, TCGA flagship manuscripts[4,6], and our internal pipelines were lifted over to GRCh38 (using the Bioconductor package rtracklayer) and annotated with respect to the canonical RefSeq FOXA1 isoform. For TCGA samples/cases multiple call-sets were available and we manually reviewed all discrepancies in FOXA1 mutation calls resulting in a union call set with improved sensitivity and specificity. Mutational impact (consequence) was simplified into 3 categories: missense, inframe indel, and frameshift the latter category included stop-gain, stop-loss, and splice-site mutations. The resulting mutations were dichotomized into Class1 and Class2 based on their position relative to residue 275aa. Variant allele frequencies (VAF) were only available for TCGA and the in-house mCRPC cohorts.

Analysis of whole-genome sequencing data

The bcbio-nextgen pipeline version 1.0.3 was used for the initial steps of tumor whole-genome data analysis. Paired-end reads were aligned to the GRCh38 reference using BWA (bcbio default settings), and structural variant calling was done using LUMPY[62] (bcbio default settings), with the following post-filtering criteria: ‘‘(SR>=1 & PE>=1 & SU>=7) & (abs(SVLEN)>5e4) & DP<1000 & FILTER==’’PASS’’.’’ The following settings were chosen to minimize the number of expected germline variants: (FDR < 0.05 for germline status for both deletions and duplications), additionally common structural germline variants were filtered.

Analysis of 10X Genomics long-read sequencing data

High-molecular-weight (HMW) DNA from MDA-prostate cancer-2B and LNCaP cell lines was isolated and processed into linked-read NGS libraries per manufacturer's instructions (10X WGS v2 kit). The resulting paired-end sequencing data were sequenced on an Illumina Hi-Seq 2500 instrument and analyzed (demultiplexing, alignment, phasing, structural variant calls) using the longranger 2.2.1 pipeline with all default settings. The resulting libraries met all 10X-recommended QC parameters including molecule size, average phasing length, and sequencing coverage (~50X). Here, we focused on structural variant calls within the FOXA1 TAD and confirmed the presence of the previously reported FOXMIND-ETV1 fusions i.e. translocation for MDA-prostate cancer-2B, and balanced insertional translocation for LNCaP. Both cell lines were confirmed to harbor three copies of FOXA1 i.e. one translocated allele and two duplicated alleles.

RNA-seq data pre-processing and primary analysis

RNA-seq data processing, including quality control, read trimming, alignment, and expression quantification by read counting, was carried out as described previously[61], using our standard clinical RNA-seq pipeline ‘‘CRISP’’ (available at https://github.com/mcieslik-mctp/bootstrap-rnascape). The pipeline was run with default settings for paired-end RNA-seq data of at least 75bp. The only changes were made for unstranded transcriptome libraries sequenced at the Broad Institute and the TCGA/CCLE/CCLE cohorts, for which quantification using ‘‘featureCounts’’ (Liao et al., 2014) was used in unstranded mode ‘‘-s0.’’. The resulting counts were transformed into FPKMs using upper-quartile normalizations as implemented in EdgeR[63]. For mCRPC samples, FOXA1 expression estimates were adjusted by tumor-content estimated from WES (see above) given the highly prostate specific FOXA1 expression profile. For the quantification of

FOXMIND expression levels, a custom approach was necessary given the poor-annotation and unspliced nature of this transcript. First, we delineated regions of sense and antisense transcription from the FOXMIND ultra-conserved regulatory elements, chr14:37564150-37591250:+ and chr14:37547900-37567150:-, respectively. Next in order to make the expression estimates reliable in unstranded libraries we identified a region of significant overlap between the sense/antisense FOXMIND transcripts and FOXA1 and MIPOL1. These overlaps have been excluded from quantification, resulting in the following trimmed target regions: chr14:37564150-37589500, and chr14:37553500-37567150. Within those regions, the average base-level coverage normalized to sequencing depth was computed as an expression estimate.

Differential expression analyses

All differential expression analyses were done using limma R-package[64], with the default settings for the ‘‘voom’’[65], ‘‘lmFit,’’ ‘‘eBayes,’’ and ‘‘topTable’’ functions. The contrasts were designed as follows, to identify transcriptional signatures of Class1 mutants: 1) given the mutual exclusivity of the genotypes in primary and metastatic tumors, the overall MCTP mCRPC 370 cohort was partitioned into 4 groups: ETS/SPOP mutant tumors, Class 1 mutant tumors, Class 2 mutant tumors, tumors WT for ETS/SPOP/FOXA1. To avoid confounding effects, the Class2 and ETS/SPOP groups were excluded from Class1 transcriptional analyses. Next, the Class1 samples were contrasted with the WT samples with additional independent regressors for assay type (Capture vs polyA, as described previously), and mutational status (see above) for the following genes/pathways: PI3K, WNT, DRD, RB1, TP53. In other words, we constructed a design matrix with coefficients for Class 1 mutational status, in addition to coefficients for confounding variables and recurrent genetic heterogeneity. This allowed us to estimate the log fold-changes and adjusted p-values associated with FOXA1 mutations and other genotypes i.e. PI3K status in Supplementary

Figure A1-4. An analogous procedure was carried out for the primary class1 samples (TCGA) and for class2 mutations in mCRPC (MCTP), but given the lack of mutual-exclusivity between Class2 mutations and ETS/SPOP, only Class1 mutations were excluded.

Pathway and signature enrichment analyses

The Molecular Signature Database (MSigDB)[66] has been used as a source of gene sets comprising cancer hallmarks, molecular pathways, oncogenic signatures, and transcription factor targets. The enrichment of signatures was assessed using the parametric Random-Set method[67], and visualized using the GSEA enrichment statistic[68] and barcode plots. All p-values have been adjusted for multiple hypothesis testing using FDR correction. To identify putative transcription factors regulating differentially expressed genes, we used the transcription-factor prediction tool BART[37]. BART was run with all default settings and provided TF databases. We used voom/limma-based gene-level fold-changes as input to the algorithm.

Detections of structural variants from RNA-seq

The detection of chimeric RNAs (gene fusions, structural variants, circular RNAs, read-through events) was carried out using our in-house toolkit for the comprehensive detection of chimeric RNAs ‘‘CODAC’’ (available at https://github.com/mctp/codac), as introduced previously[61]. Briefly, three separate alignment passes (STAR 2.4.0g1) against the GRCh38 (hg38) reference with known splice-junctions provided by the (Gencode 27) are made for the purposes of expression quantification and fusion discovery. The first pass is a standard paired-end alignment followed by gene expression quantification. The second and third pass are for the purpose of gene fusion discovery and enable STAR’s chimeric alignment mode (chimSegmentMin: 10, chimJunctionOverhangMin: 1, alignIntronMax: 150000, chimScoreMin: 1). Fusion detection was carried out using CODAC with default parameters to balance sensitivity and specificity (annotation preset: balanced). CODAC uses MOTR v2 a custom reference transcriptome based on a subset of Gencode 27 (available with CODAC). Prediction of topology (inversion, duplication, deletion, translocation), and distance (adjacent – breakpoints in two directly adjacent loci, cytoband – breakpoints within the same cytoband based on UCSC genome browser, arm – breakpoints within the same chromosome arm). The high specificity of our pipeline has been assessed through Sanger sequencing [61]. To create fusion circos plots, we have color-coded the CODAC variants based on the inferred topology of the breakpoints. Unbiased discovery of recurrently rearranged loci has been carried out by breaking the genome into 1.5Mb windows with a step of 0.5Mb. For each window, the percentage of patients with at least one RNA breakend has been calculated. The resulting genomic windows were ranked and clustered by proximity for visualization. CODAC has the ability to make fusion calls independent of known transcriptome references/annotations and hence is capable of detecting fusions involving intergenic or poorly annotated regions.

Classification of FOXA1 locus genomic rearrangements

Structural variants within the *FOXA1* locus have been partitioned into two broad topological patterns: 1) translocations (including inversions and deletions involving distal loci on the same chromosome), and 2) focal duplications. The translocations have been further subdivided into Hijacking and Swapping events based on their position relative to *FOXMIND* (GRCh38: chr14:37564150-37591250) and *FOXA1*. Hijacking translocations position a translocation partner within the *FOXMIND-FOXA1* regulatory domain (defined as GRCh38: chr14:37547501-37592000, based on manual review of HI-C, CTCF, H3K4me1, H3K27ac, and evolutionary/syntenic data). Swapping translocations preserve the *FOXMIND-FOXA1* regulatory domain but insert the translocation partner upstream of the FOXA1 promoter, frequently "swapping-out" the *TTC6* gene. Notably, one isoform of the *TTC6* gene can be transcribed from the bi-directional *FOXA1* promoter. Focal duplications within the *FOXA1* locus have been derived from the CODAC structural-variant output file. Briefly, for each case independently, all RNA-seq fusion junctions annotated by CODAC as tandem-duplications and overlapping the *FOXA1* topological domain (GRCh38: chr14:37210001-37907919) have been collated and used to infer the minimal duplicated region (MDR). Since RNA-seq chimeric junctions are generally coinciding with splice junctions (limited resolution) and generally cannot be phased (ambiguous haplotype), the inference of MDRs makes the necessary and parsimonious assumption that overlapping tandem-duplications are due to a single somatic genetic event and not multiple independent events.

Chapter acknowledgements:

Contents of this chapter were previously published as part of the following manuscript in the journal *Nature*: Parolia A*, Cieslik M*, Chu SC, Xiao L, Ouchi T, Zhang Y, Wang X, Vats P, Cao X, Pitchiaya S, Su F, Wang R, Feng FY, Wu YM, Lonigro RJ, Robinson DR, & Chinnaiyan AM. Distinct structural classes of activating FOXA1 alterations in advanced prostate cancer. Nature. 2019;571: 413–418. However, the length of the narrative in this chapter has been substantially increased for the readers' benefit and better comprehension.

* These authors equally contributed to this work.

Author contributions:

A.P., M.C. and A.M.C. conceived and designed the study; A.P. performed all the experiments with assistance from L.X., T.O., X.W. and S.P. M.C. carried out bioinformatics analyses with assistance from A.P., Y.Z., R.J.L. and P.V. S.-C.C. and A.P. performed zebrafish in vivo experiments. F.S. and R.W. generated ChIP–seq and RNA-seq libraries. X.C. performed sequencing. F.Y.F.

provided genomic validation data. Y.-M.W. and D.R.R. coordinated clinical sequencing. A.P., M.C. and A.M.C. wrote the manuscript and organized the figures.

References:

1. Gao N, Ishii K, Mirosevich J, Kuwajima S, Oppenheimer SR, Roberts RL, et al. Forkhead box A1 regulates prostate ductal morphogenesis and promotes epithelial cell maturation. Development. 2005;132: 3431–3443.
2. Friedman JR, Kaestner KH. The Foxa family of transcription factors in development and metabolism. Cell Mol Life Sci. 2006;63: 2317–2328.
3. Javed A, Lteif A. Development of the human breast. Semin Plast Surg. 2013;27: 5–12.
4. Abeshouse A, Ahn J, Akbani R, Ally A, Amin S, Andry CD, et al. The Molecular Taxonomy of Primary Prostate Cancer. Cell. 2015;163: 1011–1025.
5. Robinson D, Van Allen EM, Wu Y-M, Schultz N, Lonigro RJ, Mosquera J-M, et al. Integrative clinical genomics of advanced prostate cancer. Cell. 2015;161: 1215–1228.
6. Ciriello G, Gatza ML, Beck AH, Wilkerson MD, Rhie SK, Pastore A, et al. Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. Cell. 2015;163: 506–519.
7. Dalin MG, Desrichard A, Katabi N, Makarov V, Walsh LA, Lee K-W, et al. Comprehensive Molecular Characterization of Salivary Duct Carcinoma Reveals Actionable Targets and Similarity to Apocrine Breast Cancer. Clin Cancer Res. 2016;22: 4623–4633.
8. Gao N, Zhang J, Rao MA, Case TC, Mirosevich J, Wang Y, et al. The role of hepatocyte nuclear factor-3 alpha (Forkhead Box A1) and androgen receptor in transcriptional regulation of prostatic genes. Mol Endocrinol. 2003;17: 1484–1507.
9. Wang Q, Li W, Liu XS, Carroll JS, Jänne OA, Keeton EK, et al. A hierarchical network of transcription factors governs androgen receptor-dependent prostate cancer growth. Mol Cell. 2007;27: 380–392.
10. Cirillo LA, Lin FR, Cuesta I, Friedman D, Jarnik M, Zaret KS. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. Mol Cell. 2002;9: 279–289.
11. Iwafuchi-Doi M, Donahue G, Kakumanu A, Watts JA, Mahony S, Pugh BF, et al. The Pioneer Transcription Factor FoxA Maintains an Accessible Nucleosome Configuration at Enhancers for Tissue-Specific Gene Activation. Mol Cell. 2016;62: 79–91.
12. Clark KL, Halay ED, Lai E, Burley SK. Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. Nature. 1993;364: 412–420.
13. Bednar J, Horowitz RA, Grigoryev SA, Carruthers LM, Hansen JC, Koster AJ, et al. Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. Proc Natl Acad Sci U S A. 1998;95: 14173–14178.
14. Gibbs EB, Kriwacki RW. Linker histones as liquid-like glue for chromatin. Proc Natl Acad Sci U S A. 2018;115: 11868–11870.
15. Fournier M, Bourriquen G, Lamaze FC, Côté MC, Fournier É, Joly-Beauparlant C, et al. FOXA and master transcription factors recruit Mediator and Cohesin to the core transcriptional regulatory circuitry of cancer cells. Sci Rep. 2016;6: 34962.
16. Iwafuchi-Doi M, Zaret KS. Pioneer transcription factors in cell reprogramming. Genes Dev.

2014;28: 2679–2692.

17. Lupien M, Eeckhoute J, Meyer CA, Wang Q, Zhang Y, Li W, et al. FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. Cell. 2008;132: 958–970.
18. Wang Q, Li W, Zhang Y, Yuan X, Xu K, Yu J, et al. Androgen receptor regulates a distinct transcription program in androgen-independent prostate cancer. Cell. 2009;138: 245–256.
19. Pomerantz MM, Li F, Takeda DY, Lenci R, Chonkar A, Chabot M, et al. The androgen receptor cistrome is extensively reprogrammed in human prostate tumorigenesis. Nat Genet. 2015;47: 1346–1351.
20. Robinson JLL, Hickey TE, Warren AY, Vowler SL, Carroll T, Lamb AD, et al. Elevated levels of FOXA1 facilitate androgen receptor chromatin binding resulting in a CRPC-like phenotype. Oncogene. 2014;33: 5666–5674.
21. Jain RK, Mehta RJ, Nakshatri H, Idrees MT, Badve SS. High-level expression of forkhead-box protein A1 in metastatic prostate cancer. Histopathology. 2011;58: 766–772.
22. Gerhardt J, Montani M, Wild P, Beer M, Huber F, Hermanns T, et al. FOXA1 promotes tumor progression in prostate cancer and represents a novel hallmark of castration-resistant prostate cancer. Am J Pathol. 2012;180: 848–861.
23. Ni M, Chen Y, Lim E, Wimberly H, Bailey ST, Imai Y, et al. Targeting androgen receptor in estrogen receptor-negative breast cancer. Cancer Cell. 2011;20: 119–131.
24. Robinson JLL, Macarthur S, Ross-Innes CS, Tilley WD, Neal DE, Mills IG, et al. Androgen receptor driven transcription in molecular apocrine breast cancer is mediated by FoxA1. EMBO J. 2011;30: 3019–3027.
25. Barbieri CE, Baca SC, Lawrence MS, Demichelis F, Blattner M, Theurillat J-P, et al. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. Nat Genet. 2012;44: 685–689.
26. Yang YA, Yu J. Current perspectives on FOXA1 regulation of androgen receptor signaling and prostate cancer. Genes Dis. 2015;2: 144–151.
27. Jin H-J, Zhao JC, Ogden I, Bergan RC, Yu J. Androgen receptor-independent function of FoxA1 in prostate cancer metastasis. Cancer Res. 2013;73: 3725–3736.
28. Jin H-J, Zhao JC, Wu L, Kim J, Yu J. Cooperativity and equilibrium with FOXA1 define the androgen receptor transcriptional program. Nat Commun. 2014;5: 3972.
29. Zehir A, Benayed R, Shah RH, Syed A, Middha S, Kim HR, et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. Nat Med. 2017;23: 703–713.
30. Grasso CS, Wu Y-M, Robinson DR, Cao X, Dhanasekaran SM, Khan AP, et al. The mutational landscape of lethal castration-resistant prostate cancer. Nature. 2012;487: 239–243.
31. Nag A, Smith RG. Amplification, rearrangement, and elevated expression of c-myc in the human prostatic carcinoma cell line LNCaP. Prostate. 1989;15: 115–122.
32. Jenkins RB, Qian J, Lieber MM, Bostwick DG. Detection of c-myc oncogene amplification and chromosomal anomalies in metastatic prostatic carcinoma by fluorescence in situ hybridization. Cancer Res. 1997;57: 524–531.
33. Katoh M, Katoh M. Human FOX gene family (Review). Int J Oncol. 2004;25: 1495–1500.
34. Gao J, Chang MT, Johnsen HC, Gao SP, Sylvester BE, Sumer SO, et al. 3D clusters of somatic mutations in cancer reveal numerous rare mutations as functional targets. Genome Med. 2017;9: 4.

35. Li J, Dantas Machado AC, Guo M, Sagendorf JM, Zhou Z, Jiang L, et al. Structure of the Forkhead Domain of FOXA2 Bound to a Complete DNA Consensus Site. Biochemistry. 2017;56: 3745–3753.
36. Sekiya T, Muthurajan UM, Luger K, Tulin AV, Zaret KS. Nucleosome-binding affinity as a primary determinant of the nuclear mobility of the pioneer transcription factor FoxA. Genes Dev. 2009;23: 804–809.
37. Wang Z, Civelek M, Miller CL, Sheffield NC, Guertin MJ, Zang C. BART: a transcription factor prediction tool with query gene sets or epigenomic profiles. Bioinformatics. 2018. doi:10.1093/bioinformatics/bty194
38. Zheng D, Decker KF, Zhou T, Chen J, Qi Z, Jacobs K, et al. Role of WNT7B-induced Noncanonical Pathway in Advanced Prostate Cancer. Mol Cancer Res. 2013;11: 482–493.
39. Zhan T, Rindtorff N, Boutros M. Wnt signaling in cancer. Oncogene. 2017;36: 1461–1473.
40. Ho LH, Taylor R, Dorstyn L, Cakouros D, Bouillet P, Kumar S. A tumor suppressor function for caspase-2. Proc Natl Acad Sci U S A. 2009;106: 5336–5341.
41. Kumar S. Caspase 2 in apoptosis, the DNA damage response and tumour suppression: enigma no more? Nat Rev Cancer. 2009;9: 897.
42. Signoretti S, Waltregny D, Dilks J, Isaac B, Lin D, Garraway L, et al. p63 is a prostate basal cell marker and is required for prostate development. Am J Pathol. 2000;157: 1769–1775.
43. Mu P, Zhang Z, Benelli M, Karthaus WR, Hoover E, Chen C-C, et al. SOX2 promotes lineage plasticity and antiandrogen resistance in TP53- and RB1-deficient prostate cancer. Science. 2017;355: 84–88.
44. Behrens J, von Kries JP, Kühl M, Bruhn L, Wedlich D, Grosschedl R, et al. Functional interaction of beta-catenin with the transcription factor LEF-1. Nature. 1996;382: 638–642.
45. Miyamoto DT, Zheng Y, Wittner BS, Lee RJ, Zhu H, Broderick KT, et al. RNA-Seq of single prostate CTCs implicates noncanonical Wnt signaling in antiandrogen resistance. Science. 2015;349: 1351–1356.
46. Teng Y, Xie X, Walker S, White DT, Mumm JS, Cowell JK. Evaluating human cancer cell metastasis in zebrafish. BMC Cancer. 2013;13: 453.
47. Wang W, Zhong J, Su B, Zhou Y, Wang Y-Q. Comparison of Pax1/9 locus reveals 500-Myr-old syntenic block and evolutionary conserved noncoding regions. Mol Biol Evol. 2007;24: 784–791.
48. Amaral PP, Leonardi T, Han N, Viré E, Gascoigne DK, Arias-Carrasco R, et al. Genomic positional conservation identifies topological anchor point RNAs linked to developmental loci. Genome Biol. 2018;19: 32.
49. Tomlins SA, Laxman B, Dhanasekaran SM, Helgeson BE, Cao X, Morris DS, et al. Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. Nature. 2007;448: 595.
50. Annala M, Kivinummi K, Tuominen J, Karakurt S, Granberg K, Latonen L, et al. Recurrent SKIL-activating rearrangements in ETS-negative prostate cancer. Oncotarget. 2015;6: 6235–6250.
51. Rheinbay E, Parasuraman P, Grimsby J, Tiao G, Engreitz JM, Kim J, et al. Recurrent and functional regulatory mutations in breast cancer. Nature. 2017;547: 55–60.
52. Cowper-Sal lari R, Zhang X, Wright JB, Bailey SD, Cole MD, Eeckhoute J, et al. Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression. Nat Genet. 2012;44: 1191–1198.
53. Zhang X, Cowper-Sal lari R, Bailey SD, Moore JH, Lupien M. Integrative functional

genomics identifies an enhancer looping to the SOX9 gene disrupted by the 17q24.3 prostate cancer risk locus. Genome Res. 2012;22: 1437–1446.

54. Phair RD, Scaffidi P, Elbi C, Vecerová J, Dey A, Ozato K, et al. Global nature of dynamic protein-chromatin interactions in vivo: three-dimensional genome scanning and dynamic interaction networks of chromatin proteins. Mol Cell Biol. 2004;24: 6393–6402.
55. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 2016;44: W160–5.
56. Zhu LJ, Gazin C, Lawson ND, Pagès H, Lin SM, Lapointe DS, et al. ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. BMC Bioinformatics. 2010;11: 237.
57. Wilson S, Fitzsimons M, Ferguson M, Heath A, Jensen M, Miller J, et al. Developing Cancer Informatics Applications and Tools Using the NCI Genomic Data Commons API. Cancer Res. 2017;77: e15–e18.
58. Network TCGA. Comprehensive molecular portraits of human breast tumours. Nature. 2012;490: 61–70.
59. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012;2: 401–404.
60. Cieslik M, Chugh R, Wu Y-M, Wu M, Brennan C, Lonigro R, et al. The use of exome capture RNA-seq for highly degraded RNA with application to clinical cancer sequencing. Genome Res. 2015;25: 1372–1381.
61. Robinson DR, Wu Y-M, Lonigro RJ, Vats P, Cobain E, Everett J, et al. Integrative clinical genomics of metastatic cancer. Nature. 2017;548: 297–303.
62. Layer RM, Chiang C, Quinlan AR, Hall IM. LUMPY: a probabilistic framework for structural variant discovery. Genome Biol. 2014;15: R84.
63. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26: 139–140.
64. Smyth GK. limma: Linear Models for Microarray Data. Bioinformatics and Computational Biology Solutions Using R and Bioconductor. Springer, New York, NY; 2005. pp. 397–420.
65. Law CW, Chen Y, Shi W, Smyth GK. Voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 2014;15: R29.
66. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database Hallmark Gene Set Collection. cels. 2015;1: 417–425.
67. Newton MA, Quintana FA, Boon JAD, Sengupta S, Ahlquist P. Random-Set Methods Identify Distinct Aspects of the Enrichment Signal in Gene-Set Analysis. Ann Appl Stat. 2007;1: 85–106.
68. Subramanian A, Kuehn H, Gould J, Tamayo P, Mesirov JP. GSEA-P: a desktop application for Gene Set Enrichment Analysis. Bioinformatics. 2007;23: 3251–3253.

Chapter 3

Targeting Enhancer Addiction in Cancer by Impeding Chromatin Accessibility[2]

Abstract

The switch/sucrose non-fermentable (SWI/SNF) complex plays a crucial role in chromatin remodeling and is recurrently altered in over 20% of human cancers. Here, we developed a proteolysis targeting chimera (PROTAC) degrader of ATPase subunits of the SWI/SNF complex, SMARCA2 and SMARCA4. Intriguingly, we found androgen receptor (AR)/forkhead box A1 (FOXA1)-positive prostate cancer and MYC-driven multiple myeloma cell lines to be exquisitely sensitive to dual SMARCA2 and SMARCA4 degradation relative to benign prostate as well as other cancer cell lines, including those with inactivating SMARCA4 mutations. Mechanistically, SWI/SNF inhibition rapidly compacts the cis-regulatory elements that are bound and activated by transcription factors that drive cancer proliferation, namely AR, FOXA1, ERG, and MYC. This ensues in chromatin untethering of these oncogenic drivers, chemical decommissioning of their core enhancer circuitry, and attenuation of downstream gene programs. Furthermore, we found SWI/SNF inhibition to disrupt super-enhancer and promoter DNA looping interactions that wire supra-physiologic expression of the AR, FOXA1, and MYC oncogenes, thereby tempering their

---

[2] Contents of this chapter were summarized in a scientific manuscript that is presently under review in a reputable scientific journal. Specific author contributions are included in chapter acknowledgements.

expression in cancer cells. Monotherapy with the SMARCA2/4 degrader induced potent inhibition of tumor growth in cell line-derived xenograft models of prostate cancer, as well as multiple myeloma, and remarkably synergized with AR antagonists, inducing disease remission in models of castration-resistant prostate cancer. We also found the combinatorial treatment to significantly inhibit the growth of enzalutamide-resistant disease using in vitro as well as patient-derived xenograft models. Notably, no major toxicities were seen in mice upon prolonged treatment with the SMARCA2/4 degrader, including no indications of thrombocytopenia, gastrointestinal goblet cell depletion, or germ cell degeneration. Taken together, these results suggest that impeding enhancer accessibility through SWI/SNF ATPase inactivation represents a novel therapeutic approach in enhancer-addicted human cancers.

Introduction

In mammalian cells, DNA is wrapped around histone octamers—collectively referred to as nucleosomes—which form a physical barrier to all DNA-based processes [1,2]. Thus, gene expression is regulated by modifying the physical accessibility of the DNA through nucleosomal remodeling and, when in an accessible state, through binding of the transcription factor machinery [3,4]. In this regulatory context, non-coding genomic elements called enhancers have emerged as central genomic hubs that serve as integrative platforms for transcription factor binding and activation of lineage-specific gene programs [5–7]. The enhancer elements can lie within untranslated or distal intergenic regions and make looping interactions with their target gene promoters to potentiate RNA polymerase II-mediated gene transcription [8,9].

Notably, in human cancers, genetic alterations invariably lead to an aberrant transcriptional state that is often wired through expansion and extensive remodeling of the enhancer landscape [10–

13]. This includes *de novo* commissioning of new enhancers (also known as neo-enhancers) by reprogramming of pioneer factor cistromes [14], enhancer hijacking via structural rearrangements [15,16], and/or abnormal enhancer–promoter interactions via alterations in chromatin topology [17]—all to enable hyper-expression of driver oncogenes. This has garnered immense interest in therapeutically targeting aberrant enhancer function in human cancers. However, the molecular machinery involved in enhancer maintenance and/or activation in cancer cells remains poorly characterized and, thus, unexplored for therapeutic targetability.

Recent exome and genome-wide sequencing studies have uncovered frequent alterations in genes encoding constituent subunits of the switch/sucrose non-fermentable (SWI/SNF, also known as BAF) complex in over 20% of human cancers [18,19]. SWI/SNF is a multi-subunit chromatin remodeling complex that uses energy from ATP hydrolysis to reposition or eject nucleosomes at non-coding regulatory elements, thereby enabling free DNA access for the transcriptional machinery[20]. In SWI/SNF-mutant tumors, the residual complex is thought to enable oncogenic transcriptional programs and is speculated to be a viable therapeutic target [21–23], yet, hitherto, no studies have comprehensively assessed the therapeutic efficacy of SWI/SNF inhibition across a wide spectrum of human cancers. To facilitate this evaluation, we developed and characterized a highly-selective proteolysis targeting chimera (PROTAC) degrader of the SWI/SNF ATPase subunits SMARCA2 (also known as BRM) and SMARCA4 (also known as BRG1) that are required for the nucleosomal remodeling functions of SWI/SNF family complexes.

Surprisingly, we found transcription factor-addicted cancers (e.g., prostate cancer and multiple myeloma) to be exquisitely and preferentially sensitive to SWI/SNF inhibition, which triggered a rapid and specific loss of physical accessibility at enhancer elements, thereby disrupting enhancer-wired oncogenic gene programs. To the best of our knowledge, this study is the first preclinical proof of concept that targeted obstruction of chromatin accessibility at enhancer elements can be a potent therapeutic strategy in transcriptionally-addicted tumors. Our degrader compound will serve as an important pharmacologic tool in mechanistic interrogation of frequent SWI/SNF alterations in preclinical cancer models and for evaluating the therapeutic benefit of SWI/SNF inhibition in clinical trials.

Results

AU-15330 is a selective and highly-potent PROTAC degrader of the SWI/SNF ATPases

Here, we developed the PROTAC degrader AU-15330 that is chemically composed of a bait moiety that specifically binds to the bromodomain in SMARCA2 and SMARCA4, and is covalently linked to a ligand for the von Hippel-Lindau (VHL) ubiquitin ligase (**Figure 3-1A, Supplementary Figure A2-1A**). AU-15330 also binds to the bromodomain of PBRM1, which is a secondary module component and relies on the ATPase module for its assembly onto the core SWI/SNF complex [24]. Treatment of several human-derived cell lines with AU-15330 led to time and dose-dependent degradations of SMARCA2, SMARCA4, and PBRM1—with the majority of the proteins degraded within 1-2 h of treatment at 1 μM (**Figure 3-1B**). Mass spectrometry-based global proteomics further confirmed SMARCA2, SMARCA4, and PBRM1 as the only significantly down-regulated proteins after 4 h of AU-15330 treatment (**Supplementary Figure A2-1B**). Remarkably, we detected no change in the abundance of other bromodomain-containing proteins that were detectable in the mass spectrometry data, even at the 1 μM AU-15330 dosage (**Figure 3-1C**). Recent studies have shown that the SWI/SNF complex assembles in a modular fashion, with the ATPase module being the last to bind to the SMARCC1 (also known as BAF155)-containing core SWI/SNF complex [24]. In accordance with these findings, SMARCC1 nuclear immunoprecipitation followed by mass spectrometry showed no changes in the integrity of the core and secondary modules upon AU-15330 treatment, but both sub-complexes were detached from subunits belonging to the ATPase module (**Figure 3-1D**).

Having credentialed the specificity of AU-15330 for its direct targets, we next set out to phenotypically characterize it using a panel of over 65 normal and cancer cell lines from 14 different lineages. Here, we found androgen receptor (AR)/forkhead box A1 (FOXA1)-driven prostate cancer cells to be exquisitely and preferentially sensitive to SWI/SNF inhibition, with half-maximal inhibitory concentrations (IC_{50}) approaching single-digit, nanomolar values (all IC_{50} values < 100 nM; **Figure 3-1E**, **Supplementary Figure A2-1C, D and Supplementary Table 1**). In the same assay, AR/FOXA1-negative prostate cancer cells showed moderate sensitivity (IC_{50} between 100-400 nM), while normal/non-neoplastic prostate cells were resistant ($IC_{50} > 1000$ nM) to SWI/SNF inhibition. Intriguingly, AR/FOXA1-positive prostate cancer cells were even more sensitive than SMARCA4-null cancer cell lines, namely SK-MEL-5, Capan-2, and H-838 (**Figure 3-1E**). This strongly suggests that the SWI/SNF complex plays a key oncogenic role in the survival of AR/FOXA1-positive prostate cancers, which is in line with the absence of recurrent deleterious SWI/SNF mutations in this cancer type [25,26]. We also found MYC-driven multiple myeloma cells and AR and estrogen receptor-positive breast cancer cells (WA-72-P and WA-72-As) to be sensitive to AU-15330 treatment (**Figure 3-1E, S1E**).

In several prostate cancer cell lines, we detected a robust expression of both the SWI/SNF ATPases, which were degraded in a dose-dependent manner within 1-2 h of AU-15330 treatment (**Supplementary Figure A2-2A, B**). Concordantly, we found AU-15330 to markedly attenuate the growth of these cells and induce apoptotic cell death, while normal/non-neoplastic prostate cells showed no anti-proliferative effect at parallel dosages (**Figure 3-1F, Supplementary Figure**

A2-1C and A2-2C). Competition of AU-15330 using a free VHL ligand further reversed degradation of SWI/SNF targets (**Supplementary Figure A2-2D**) as well as rescued the growth inhibition seen in prostate cancer cells, with both effects being dose-dependent (**Supplementary Figure A2-2E**). We next employed a genetic approach to verify that AU-15330-mediated cytotoxicity stems from direct degradation of both SWI/SNF ATPases. Using CRISPR/Cas9, we knocked out either SMARCA2 or SMARCA4 individually or together in LNCaP cells. We found that knock-out of either ATPase alone had moderate to no effect on cell growth, while the double knock-out of both SMARCA2 and SMARCA4 led to a significant anti-proliferative effect comparable to that with AU-15330 treatment (**Supplementary Figure A2-2F**). This demonstrates that concurrent inhibition of both ATPases is essential for AU-15330-mediated cytotoxicity in prostate cancer cells. Altogether, this data shows that AU-15330 is a high-fidelity, VHL-dependent degrader of essential SWI/SNF ATPases, namely SMARCA2 and SMARCA4, and that complete inhibition of SWI/SNF activity triggers apoptotic cell death in AR/FOXA1-dependent prostate cancer cells.

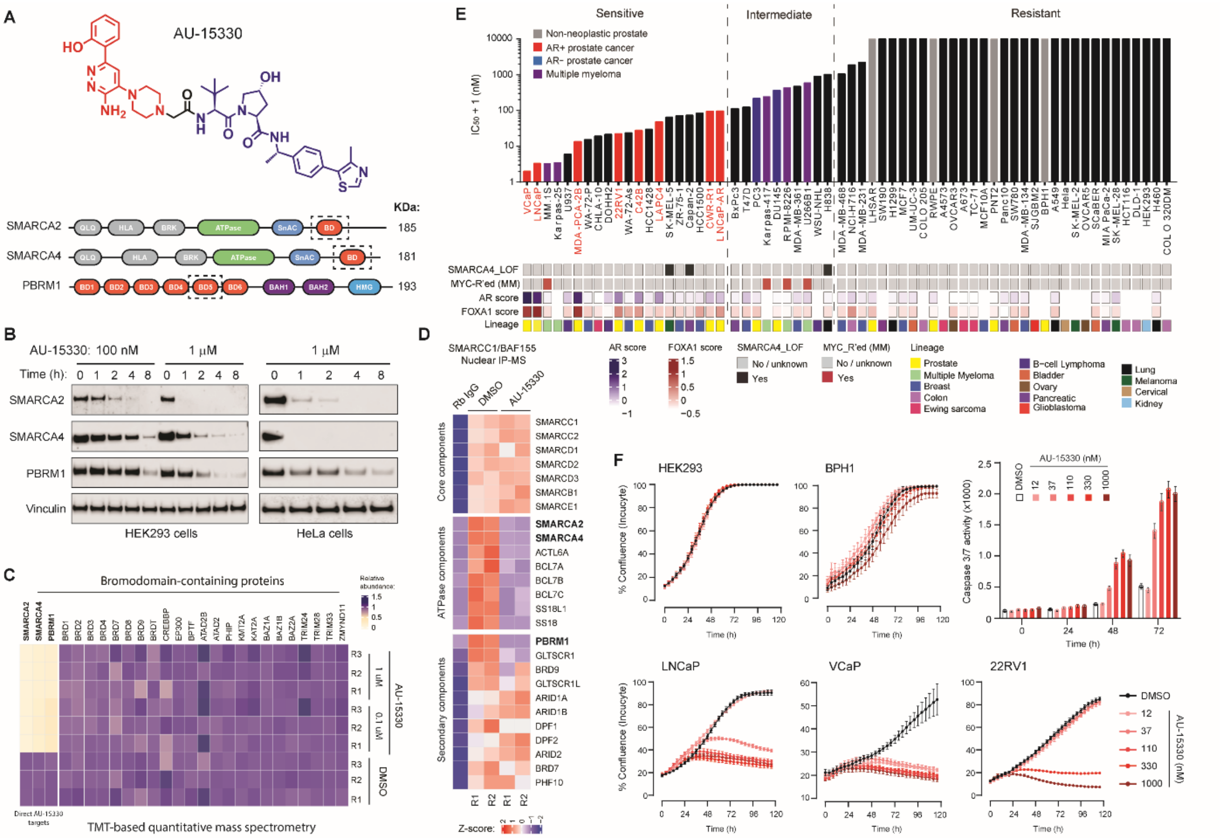


Figure 3-1: AU-15330 is a specific degrader of SWI/SNF ATPases with preferential cytotoxicity in enhancer-binding transcription factor-driven cancers. (A) Chemical structure of AU-15330 and a schematic illustration of SMARCA2, SMARCA4, and PBRM1 domain architecture. AU-15330 targeted bromodomains (BD) are boxed. KDa, kilo-dalton. (B) Immunoblots for SMARCA2, SMARCA4, and PBRM1 proteins from whole-cell lysates upon treatment of HEK293 and HeLa cells with AU-15330 at increasing concentrations or time durations. Vinculin serves as a loading control. (C) Heat map of relative abundance of several bromodomain-containing proteins detectable via Tandem Mass Tag (TMT)-based quantitative mass spectrometry upon AU-15330 treatment (1 µM) for 4 h. DMSO, dimethyl sulfoxide (vehicle). (D) Heat map of mammalian SWI/SNF (BAF) complex subunits split into three constituent modules detected in SMARCC1 (also known as BAF155) nuclear co-immunoprecipitation followed by mass spectrometry (IP-MS). Normal rabbit (Rb) IgG was used as control. Direct AU-15330 targets are shown in bold. (E) Half-maximal inhibitory concentrations (IC50) of AU-15330 in a panel of human-derived cancer or normal cell lines after 5 days of treatment. Cell lines with known SMARCA4 loss-of-function (LOF) alterations and multiple myeloma (MM) cell lines with MYC rearrangements (MYC R'ed) are identified in the panel below. Androgen receptor (AR) and forkhead box A1 (FOXA1) scores quantify their transcriptional activities using cognate multi-gene signatures (see Methods). AR+, AR-positive; AR-, AR-negative. (F) Growth curves (Incucyte) of multiple non-neoplastic or prostate cancer cells upon treatment with increasing concentrations of AU-15330. Top, rightmost panel shows real-time Incucyte-based assessment of apoptotic signal in LNCaP cells after treatment with DMSO (solvent control) or increasing concentrations of AU-15330.

SWI/SNF inhibition physically compacts core-enhancers and attenuate oncogenic gene programs

As the SWI/SNF complex actively remodels nucleosomal packaging of the DNA, we next profiled the effect of AU-15330 treatment on physical chromatin accessibility using the assay for transposase-accessible chromatin followed by sequencing (ATAC-seq) [27]. We detected a rapid and complete loss in chromatin accessibility at over 30,000 sites in VCaP cells in as little as 4 h of AU-15330 treatment (**Figure 3-2A, Supplementary Figure A2-3A**). By contrast, approximately 25,000 genomic sites showed little to no change in nucleosomal density (**Supplementary Figure A2-3B**). Such profound changes in chromatin accessibility were not observed upon treatment with a BRD4 degrader (ZBC-260 [28], **Figure 3-2A, Supplementary Figure A2-3A, B**), suggesting a distinct mechanism of action for SWI/SNF inhibition-triggered cytotoxicity. In our genetic models using CRISPR/Cas9 and shRNA-mediated target inhibition (**Supplementary Figure A2-3C**), consistent with the cellular phenotype, we detected marked chromatin compaction only upon concurrent loss of both SWI/SNF ATPases (**Supplementary Figure A2-3D**). Over 90% of the AU-15330-compacted sites were within distal regulatory genomic regions that were enriched for enhancers, while the retained sites were predominantly within gene promoters (**Figure 3-2B**). *De novo* motif analysis of AU-15330-compacted sites identified DNA binding elements for major oncogenic transcription factors in prostate cancer, including AR, FOXA1, HOXB13, and ERG (**Figure 3-2C**). Binding analysis for the regulation of transcription (BART; [29]) using a compendium of chromatin immunoprecipitation followed by sequencing (ChIP-seq) data further revealed the binding of AR and FOXA1 to be most significantly enriched within these regulatory sites (**Supplementary Figure A2-3E**). As expected, the retained promoter sites showed enrichment for Pol II and E2F motifs (**Supplementary Figure A2-3F**). Interrogation of chromatin changes in LNCaP cells upon AU-15330 treatment reproduced the above findings, wherein AU-15330-compacted genomic sites were predominantly within distal regulatory regions that were enriched for the consensus binding motifs of bona fide oncogenic drivers, namely AR, FOXA1, HOXB13, and ETS (**Supplementary Figure A2-4A-C**).

Thus, we next set out to evaluate the effect of SWI/SNF inhibition on chromatin interaction of oncogenic transcription factors. Strikingly, 6 h after AU-15330 treatment, ChIP-seq analysis revealed a marked attenuation of chromatin binding of AR, FOXA1, and ERG in VCaP cells (**Figure 3-2D, Supplementary Figure A2-4D**). This was accompanied by parallel depletion of enhancer-associated histone 3 lysine 27 acetylation (H3K27Ac) mark at the nucleosomes flanking the AU-15330-compacted sites by the 24 h time point (**Figure 3-2D**), as well as global decreases in H3K27Ac levels (**Supplementary Figure A2-4E**). AU-15330 treatment in LNCaP cells entirely reproduced the genome-wide attenuation of AR and FOXA1 binding at their enhancer elements, impeding chemical activation of these regulatory sites (**Supplementary Figure A2-4E-G**). Further analysis of the ChIP-seq data revealed a marked overlap in the binding sites of AR, FOXA1, ERG [30], and SMARCC1 [30] in VCaP cells, with the majority of the shared binding sites lying within non-coding H3K27Ac-enriched regions (**Supplementary Figure A2-5A-C**). Concordantly, we identified multiple core SWI/SNF components in the mass spectrometry-based datasets of AR, FOXA1, and ERG interactomes [30,31] (**Supplementary Figure A2-5D**), which we were able to confirm via reciprocal co-immunoprecipitation assays in AR-positive prostate cancer cells (**Supplementary Figure A2-5E**). This data strongly suggests that the SWI/SNF complex is a common chromatin cofactor of the central transcriptional machinery in prostate cancer cells and that its ATPase activity is essential to retain enhancers in an accessible, chemically active confirmation. Accordingly, in two independent prostate cancer cell lines, AU-15330 triggered a marked loss in AR and FOXA1 binding and H3K27Ac abundance at bona fide enhancers of classical AR target genes, namely *KLK3* and *KLK2* (**Figure 3-2E**). Serving as an important control, we saw no major defect in chromatin binding of CTCF at the insulator sites in either VCaP or LNCaP AU-15330-treated cells (**Supplementary Figure A2-6A-D**).

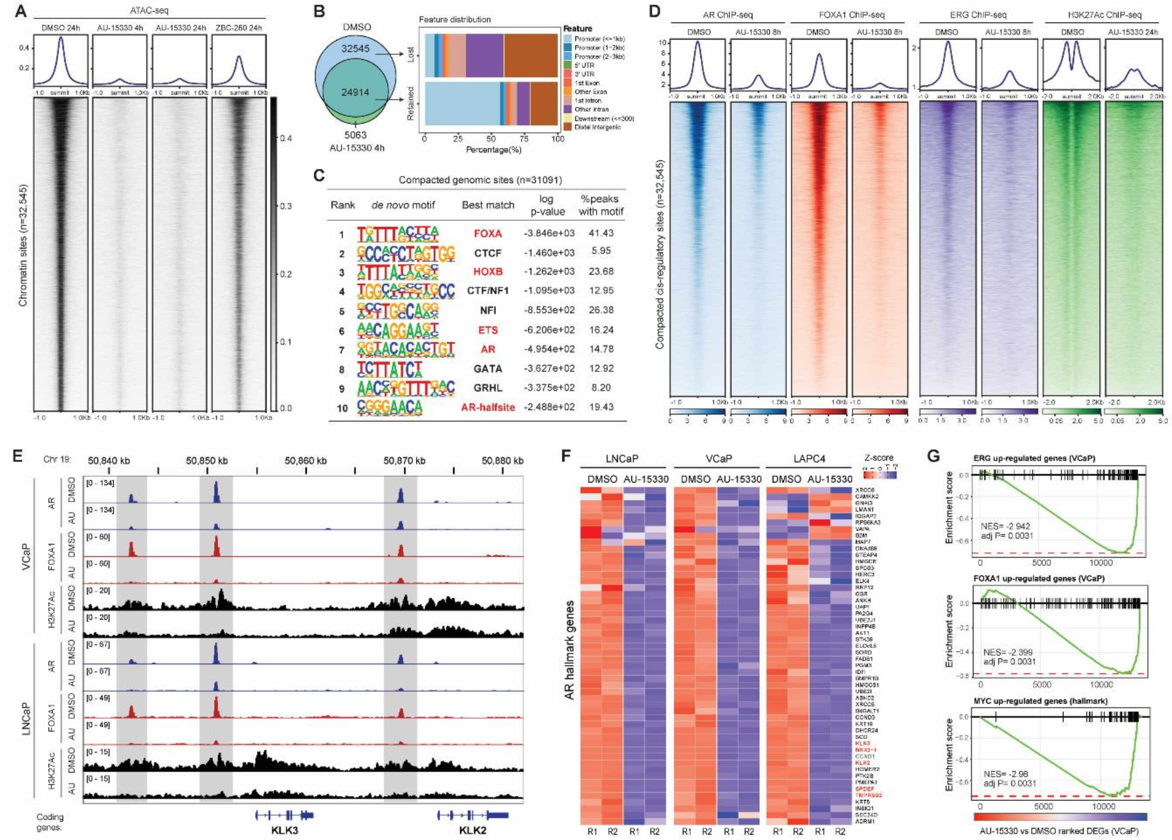


Figure 3-2: SWI/SNF ATPase inhibition disrupts physical chromatin accessibility at the core-enhancer circuitry to disable oncogenic transcriptional programs. (A) ATAC-seq read-density heat maps from VCaP cells treated with DMSO (solvent control), AU-15330, or ZBC-260 (a BRD4 degrader) for indicated durations (n=2 biological replicates per condition). (B) Genome-wide changes in chromatin accessibility upon AU-15330 treatment for 4 h in VCaP cells along with genomic annotation of sites that lose physical accessibility (i.e. lost) or remain unaltered (i.e. retained). (C) Top ten *de novo* motifs (ranked by p-value) enriched within AU-15330-compacted genomic sites (HOMER, hypergeometric test) in VCaP cells. (D) ChIP-seq read-density heat maps for AR, FOXA1, ERG, and H3K27Ac at the AU-15330-compacted genomic sites in VCaP cells after treatment with DMSO (solvent control) or AU-15330 (at 1 μM) for indicated time durations and stimulation with R1881, a synthetic androgen (at 1 nM for 3 h). (E) ChIP-seq read-signal tracks for AR, FOXA1, and H3K27Ac within the *KLK2/3* gene locus in R1881-stimulated VCaP and LNCaP cells with or without AU-15330 (AU) treatment. (F) RNA-seq heat maps for classical AR target genes in LNCaP, VCaP, and LAPC4 prostate cancer cells with or without AU-15330 treatment for 24 h. Well-characterized AR target genes are shown in red. (G) Gene set enrichment analysis plots for the ERG, FOXA1, and MYC regulated genes using the fold change rank-ordered gene signature from AU-15330-treated (at 1 nM for 24 h) VCaP cells. NES, net enrichment score; adj P, adjusted p-value; and DEGs, differentially expressed genes.

Next, global transcriptomic profiling using RNA sequencing (RNA-seq) revealed a significant down-regulation of AR and FOXA1-regulated genes in three independent prostate cancer cell lines (**Figure 3-2F, G, Supplementary Figure A2-7A**), as well as ERG-regulated transcripts in ERG fusion-positive VCaP cells (**Figure 3-2G**). In all prostate cancer cells, we also noticed a remarkable loss in the expression of MYC target genes upon AU-15330 treatment (**Figure 3-2G, Supplementary Figure A2-7A**). Intriguingly, we found the global AU-15330 gene signature to be remarkably concordant with transcriptional changes associated with ARID1A loss in VCaP cells (**Supplementary Figure A2-7B**). As ARID1A is an exclusive component of the classical SWI/SNF (also known as cBAF) complex [24], we interrogated the specific role of polybromo-containing and non-classical SWI/SNF (also known as pBAF and ncBAF, respectively) complexes in mediating the AU-15330 transcriptional phenotype. To this end, we treated prostate cancer cells with a BRD9 degrader (a specific ncBAF component; dBRD9 [32]), a dual BRD7 and BRD9 degrader (BRD7 is a specific pBAF component; dBRD7/9 [33]), or AU-15330 at equal dosages and compared changes in gene expression. Here, neither BRD7 nor BRD9 inhibition alone was able to attenuate the expression of canonical AR, FOXA1, and ERG target genes or the *MYC* gene to an extent comparable to that upon AU-15330 treatment (**Supplementary Figure A2-7C, D**). This strongly suggests that the classical SWI/SNF complex is the primary cofactor of enhancer-binding transcription factors and is essential for enabling their oncogenic gene programs. Thus, SWI/SNF inhibition specifically abolishes enhancer activity due to the loss of physical chromatin accessibility and binding of oncogenic transcription factors in prostate cancer cells.

SWI/SNF inactivation disrupts super-enhancer and promoter interactions and tempers supra-physiologic expression of driver oncogenes

Recent genomic studies have revealed the expression of *AR*, *MYC*, and *FOXA1* genes themselves to be amplified in advanced prostate cancer via copy-amplification and/or enhancer duplication [16,34,35]. Thus, we profiled changes in the expression of these driver oncogenes after treatment with AU-15330. Remarkably, we found AU-15330 treatment to decrease expression of *AR*, *FOXA1*, *MYC*, and *TMPRSS2-ERG* genes to 40-60% of their initial expression by 24 h (**Figure 3-3A, Supplementary Figure A2-8A**), which was also evident at the protein level (**Figure 3-3B**). Importantly, a more severe attenuation of expression of these transcripts was noted upon BRD4 degradation (**Figure 3-3A, Supplementary Figure A2-8A**), again suggesting a distinct mechanism of action for AU-15330 and likely explaining the dose-limiting toxicites associated with BRD4 inhibitors (see below for discussion). In our genetic models, we found only concurrent inhibition of both ATPases to recapitulate the transcriptional phenotype of AU-15330 at both RNA and protein levels (**Supplementary Figure A2-8B**). Notably, we also found SWI/SNF inhibition to down-regulate the expression of other oncogenic drivers such as *EZH2, CCND1, SPDEF,* and *GATA2* in prostate cancer cells (**Supplementary Figure A2-8A**).

The hyper-expression of oncogenes like *AR, FOXA1,* and *MYC* in cancer has been shown to be wired through looping interactions with multi-enhancer clusters, often referred to as super-enhancers [36]. Concordantly, an unbiased annotation of super-enhancers in VCaP cells using H3K27Ac ChIP-seq identified several such regulatory clusters in cis-proximity of the *AR, MYC,* and *TMPRSS2-ERG* genes (**Figure 3-3C**). Remarkably, AU-15330 treatment led to immediate compaction (within 4 h) of these sites, as well as a depletion of H3K27Ac at the flanking nucleosomes (**Figure 3-3D**), with a parallel loss in the binding of both AR and FOXA1 at the super-enhancer clusters (**Supplementary Figure A2-8C**). Next, to experimentally confirm the interaction of super-enhancers with their target gene promoters, we performed H3K4me3 (an active promoter mark) and H3K27Ac Hi-C coupled with ChIP-seq (HiChIP-seq) assay and quantified changes in these interactions upon treatment with AU-15330. In line with the physical and chemical decommissioning of enhancer elements, we found SWI/SNF inhibition to completely disrupt the three-dimensional looping interactions of bona fide, *cis*-coded enhancers with the *AR* gene promoter (**Figure 3-3E, F**). Notably, recent clinical studies have reported focal amplification of the ChrX:66.9Mb *AR* enhancer in castration-resistant prostate cancer to drive resistance to second-generation AR inhibitors by amplifying the expression of the *AR* gene [34,35]. Similar attenuation of enhancer–promoter interactions was detected in the H3K27Ac HiChIP-seq data at the *FOXA1* locus (**Supplementary Figure A2-8D**)—which is also recurrently rearranged to duplicate or hijack *FOXA1* enhancers in advanced prostate cancer [16]. A global aggregate of HiChIP-seq signal for interaction between enhancers and gene promoters showed a marked attenuation of contact strength and/or frequency merely 4 h after AU-15330 treatment (**Figure 3-3G**). Importantly, consistent with no change in the CTCF cistrome, in the CTCF HiChIP-seq data, we found no change in the looping interactions between insulator elements (**Figure 3-3H, Supplementary Figure A2-8E**). Altogether, this data suggests that SWI/SNF inhibition leads to a genome-wide collapse of the AR, FOXA1, ERG, and MYC-activated core enhancer circuitry in prostate cancer cells, while having little to no effect on the CTCF-instructed topological chromatin architecture. Thus, AU-15330-triggered SWI/SNF inhibition reverses the hyper-expression of *AR*, *FOXA1*, *MYC,* and *ERG* genes themselves via disruption of their promoter interaction with cis-coded super-enhancers, in addition to obstructing their chromatin binding and downstream transcriptional programs.

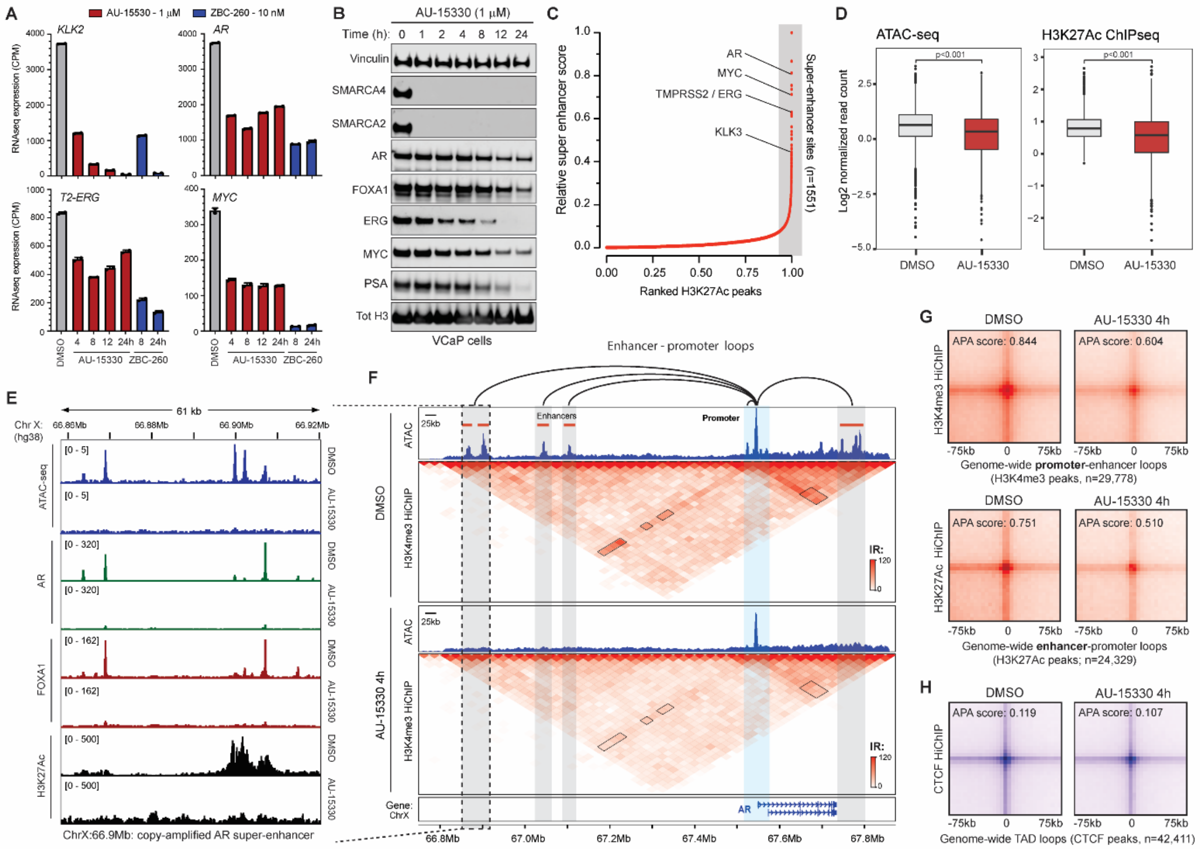
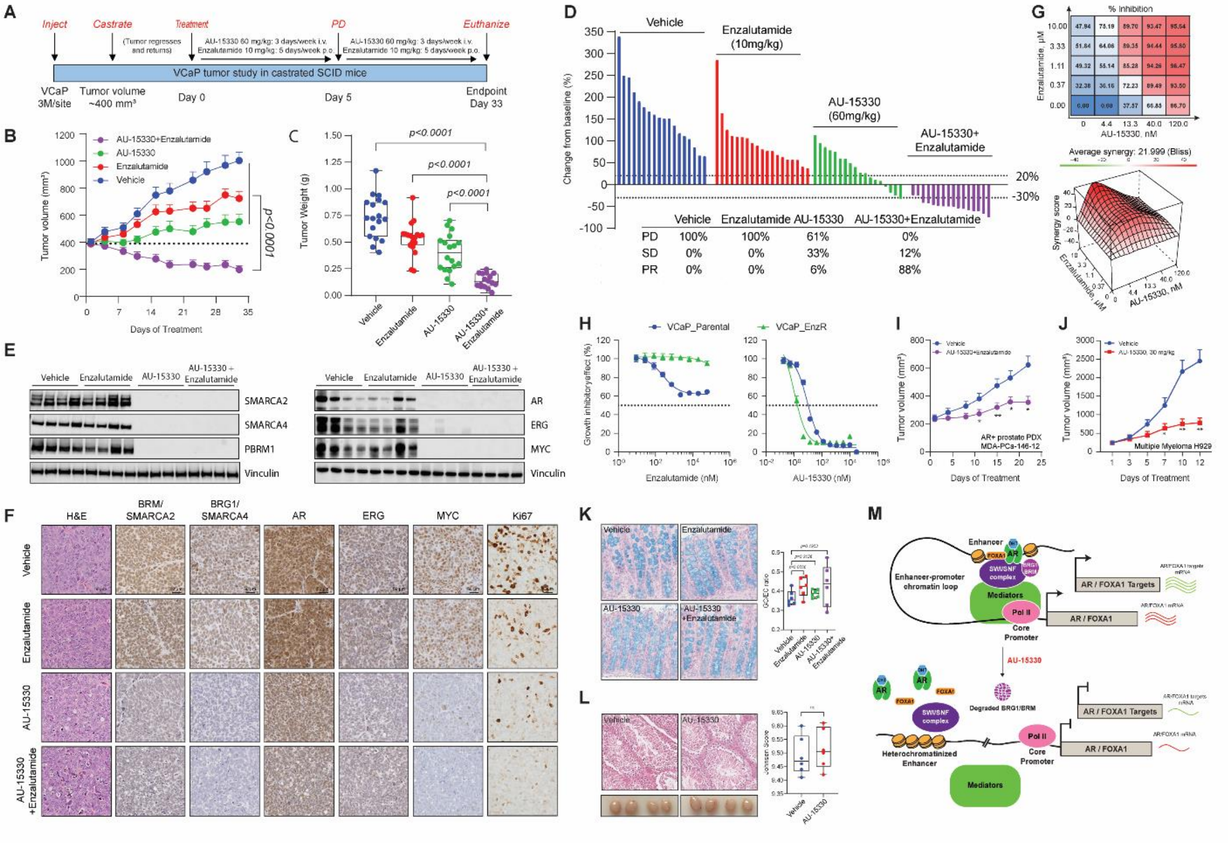


Figure 3-3: SWI/SNF ATPase inhibition disrupts enhancer-promoter loops to temper supra-physiologic expression of driver oncogenes. (A) Expression (qPCR) of indicated mRNA transcripts upon treatment of VCaP cells with AU-15330 (at 1 μM) or ZBC-260 (at 10 nM; a BRD4 degrader) for indicated time durations. CPM, counts per million reads mapped. (B) Immunoblots of indicated proteins in VCaP cells treated with DMSO for 24 h or AU-15330 at 1 μM for increasing time durations. (C) H3K27Ac ChIP-seq signal rank-ordered list of super-enhancers in VCaP cells with select cis-coded driver oncogenes denoted (HOMER). (D) Normalized read density of ATAC-seq or H3K27Ac ChIP-seq at the super-enhancer sites in VCaP cells treated with DMSO or AU-15330 (at 1 μM) for 6 h (for ATAC) or 24 h (for H3K27Ac ChIP-seq). (E) ATAC-seq and AR, FOXA1, and H3K27Ac ChIP-seq read-signal tracks within the recurrently amplified *AR* super-enhancer elements in VCaP cells with or without AU-15330 (1 μM) treatment. Mb, megabasepair; hg38, human genome reference consortium build 38. (F) H3K4me3 HiChIP-seq heat maps within the *AR* gene locus in VCaP cells plus/minus treatment with AU-15330 (1 μM) for 4 h (bin size = 25Kb). ATAC-seq read-density tracks from the same treatment conditions are overlaid. Grey highlights mark enhancers, while the blue highlight marks the *AR* promoter. Loops indicate read-supported cis-interactions within the locus. IR, interaction reads (G) Aggregate peak analysis (APA) plots for H3K4me3 (active promoter mark) and H3K27Ac (active enhancer mark) HiChIP-seq data for all possible interactions between putative enhancers and gene promoters in VCaP cells plus/minus treatment with AU-15330 (1 μM) for 4 h. (H) Aggregate peak analysis plots for CTCF HiChIP-seq data for all possible interactions between CTCF-bound insulator elements in VCaP cells plus/minus treatment with AU-15330 (1 μM) for 4 h. TAD, topologically associating domain.

SWI/SNF inactivation shows strong antitumor potency in preclinical cancer models and synergizes with existing therapeutics

Next, we set out to pharmacologically characterize AU-15330 using preclinical animal models of advanced prostate cancer. As AU-15330 binds to and degrades mouse orthologs of its SWI/SNF targets (**Supplementary Figure A2-9A**), we first carried out a comprehensive assessment of AU-15330 tolerability using different dosages as well as dosing patterns in immuno-competent, non-tumor bearing mice (**Supplementary Figure A2-9B**). Notably, AU-15330 was robustly detected in plasma for several hours after intravenous administration (**Supplementary Figure A2-9C**) and triggered a marked loss in SMARCA2 and SMARCA4 proteins in normal mouse tissues within 5 days of treatment (**Supplementary Figure A2-9D**). Yet, AU-15330 treatment over a period of 2 weeks in mice resulted in no major changes in either whole body weights (**Supplementary Figure A2-9E**) or weights of essential visceral organs (**Supplementary Figure A2-9F**). Furthermore, we found no reduction in the levels of white or red blood cells or blood platelets (**Supplementary Figure A2-9G**), acute depletion of the latter being previously reported for the first-generation of BRD4 inhibitors [37,38].

Given the absence of toxicities to normal cells, we next set out to assess tumor-specific cytotoxicity of AU-15330 using xenograft mouse models. We first employed the castration-resistance VCaP-derived prostate cancer model, termed VCaP-Castrate. To generate this model, VCaP cells were subcutaneously grafted in immunocompromised mice that were castrated after 2 weeks of tumor growth to induce disease remission. This was eventually followed by tumor re-growth in the androgen-depleted conditions, generating the aggressive, castration-resistant tumor which we termed VCaP-CRPC (**Figure 3-4A**). As expected, treatment of castrated male mice bearing the VCaP-CRPC xenografts with enzalutamide (a second-generation AR antagonist) alone showed moderate anti-tumor efficacy; however, treatment with AU-15330 alone led to potent inhibition of tumor growth (**Figure 3-4B, C**), triggering disease regression in some of the treated animals (**Figure 3-4D**). Remarkably, we found treatment with the combinatorial regimen of AU-15330 and enzalutamide to induce the most potent growth-inhibitory effect, with tumor regression in all of the treated animals (**Figure 3-4B-D**). Molecular interrogation of the treated tumors showed robust down-regulation of SWI/SNF targets as well as AR, ERG, and MYC proteins in xenografts within 5 days of AU-15330 treatment, both when administered alone or in combination with enzalutamide (**Figure 3-4E, F**). Concordantly, we found the strongest loss of Ki67 expression in tumors from the combination arm, with clear histologic evidence of cell death (**Figure 3-4F, Supplementary Figure A2-11F**). AU-15330 monotherapy also strongly inhibited the growth of the H929 cell line-derived multiple myeloma xenografts in mice (**Supplementary Figure A2-11G**).



	Vehicle	Enzalutamide	AU-15330	AU-15330+Enzalutamide
PD	100%	100%	61%	0%
SD	0%	0%	33%	12%
PR	0%	0%	6%	88%

Enzalutamide, μM	0	4.4	13.3	40.0	120.0
10.00	47.94	73.19	89.70	93.47	95.54
3.33	51.84	64.06	89.35	94.44	95.80
1.11	49.32	55.14	85.28	94.26	96.47
0.37	32.38	36.16	72.23	89.49	93.50
0.00	0.00	0.00	37.57	66.85	86.70

Figure 3-4: A SWI/SNF ATPase degrader inhibits tumor growth in preclinical models of castration-resistant prostate cancer and synergizes with enzalutamide. (A) Schematic outlining the AU-15330 *in vivo* efficacy study using the VCaP-CRPC (castration-resistant prostate cancer) xenograft model. VCaP-xenograft bearing male mice were castrated and, upon tumor regrowth, randomized into various treatment arms that were administered vehicle, enzalutamide, AU-15330, or the combination of AU-15330+enzalutamide at indicated concentrations. PD, pharmacodynamics. (B) Tumor volume measurement showing the effect of AU-15330 alone or in combination with enzalutamide. Caliper measurements were taken twice a week. Mean tumor volume ± standard error of the mean (S.E.M.) is shown. (Vehicle: n=18; AU-15330: n=20; enzalutamide: n=18; AU-15330+enzalutamide: n=16.) (C) Individual tumor weights from different treatment groups with *p* values are shown. (Vehicle: n=18; AU-15330: n=20; enzalutamide: n=18; AU-15330+enzalutamide: n=16.) (D) Waterfall plot depicting the change in tumor volume after 10 days of treatment in all treatment arms of the *in vivo* study. The Response Evaluation Criteria in Solid Tumors (RECIST) were used to stratify tumors into progressive disease (PD), stable disease (SD), or partial response (PR). PD: at least a 20% increase in tumor size; SD: an increase of <20% to a decrease of <30% in tumor size; PR: at least a 30% decrease in tumor size. (E) Immunoblots of direct AU-15330 targets (SMARCA2, SMARCA4, and PBRM1) and oncogenic transcription factors (AR, ERG, and MYC) in the whole cell lysate from VCaP-CRPC xenografts from all treatment arms (n=4 tumors/arm) after 5 days of *in vivo* treatment. Vinculin serves as a loading control. (F) Representative H&E and immunohistochemistry images showing expression of direct AU-15330 targets, driver transcription factors, or Ki67 in VCaP-CRPC xenografts tissues from all treatment arms (n=2 tumor/arm). (G) *Top panel*, Assessment of VCaP cell viability (Celltiter-Glo) upon treatment with increasing concentrations of AU-15330 in parallel with enzalutamide (percent viabilities are an average from 3 biological replicates). These measurements were used to carry out a Bliss index-based assessment of drug synergism, with net synergism scores plotted in the panel below. Red peaks depict the extent of drug synergy. (H) Dose-response curve of VCaP_Parental and VCaP_EnzR (enzalutamide-resistant) cells treated with enzalutamide or AU-15330. Data represent mean ± S.E.M. (n=6) from one of three independent experiments. (I) Tumor volume measurement showing the effect of the combination of AU-15330 and enzalutamide in the inherently enzalutamide-resistant MDA-PCa-146-12 patient-derived prostate cancer xenografts (PDX) (n=20/arm). Caliper measurements were taken twice a week. Mean tumor volume ± S.E.M. is shown. (J) Tumor volume measurement showing the effect of AU-15330 in H929-derived multiple

myeloma xenografts (n=5/arm). Caliper measurements were taken twice a week. Mean tumor volume ± S.E.M. is shown. (Vehicle: n=5 and AU-15330: n=5). (K) Representative images of Alcian blue staining from the large intestinal tract harvested at the efficacy study endpoint from animals from all treatment groups (n=2/group). Quantification of goblet to epithelial cell densities in the colon are plotted in the right panel. (L) Representative H&E of the testis gland harvested from DMSO or AU-15330-treated intact male mice after 21 days of *in vivo* treatment. Quantification of germ cell density and maturation carried out using the Johnsen scoring system is plotted in the right panel. Gross images of the testis glands are shown in the panel below. (M) Schematic depicting the mechanism of action for AU-15330-triggered cytotoxicity in AR/FOXA1-signaling-driven prostate cancer or MYC-signaling-driven multiple myeloma. SWI/SNF ATPase inhibition via concurrent SMARCA2/4 degradation induces a rapid, complete, and targeted loss in chromatin accessibility at the core-enhancer circuitry of AR, FOXA1, ERG, and MYC, thereby attenuating their cancer-promoting transcriptional programs as well as tempering the enhancer-wired supra-physiologic expression of driver oncogenes.

Given the compounded tumor inhibitory effect in the combination arm, we carried out a comprehensive *in vitro* evaluation of drug synergism between AU-15330 and enzalutamide. Using the Bliss Independence method to assess the combined efficacy of drugs [39], we found AU-15330 and enzalutamide to show remarkable synergism (synergy scores of 21.99 for VCaP, 17.06 for LNCaP, and 19.29 for C4-2B) in prostate cancer cell lines harboring distinct therapy-induced AR alterations (**Figure 3-4G, Supplementary Figure A2-10**). Consistently, we found AU-15330 to be similarly potent in enzalutamide-resistant derivatives of VCaP and LNCaP cell lines (**Figure 3-4H, Supplementary Figure A2-11H**) [40]. The combinatorial regimen also markedly inhibited tumor growth in a patient-derived xenograft model that is inherently resistant to enzalutamide, namely MDA-PCa-146-12 (**Figure 3-4I-J, Supplementary Figure A2-12A-C**). Notably, in all arms of these efficacy studies, we found no gross changes in animal body weights or histologic evidence of any toxicity in the normal glands, despite the degradation of mouse SMARCA2 and SMARCA4 proteins upon AU-15330 treatment (**Supplementary Figure A2-11E**). This includes no sign of goblet cell depletion in the gastrointestinal tract (**Figure 3-4K**), no defect in germ cell maturation (**Figure 3-4L**), and no testicular atrophy (bottom panel, **Figure 3-4L**) in AU-15330 treated animals—all being major toxicities previously reported for the BRD4 class of therapeutics [37,38].

Discussion

Modern genomic catalogs of human cancers show a high recurrence of somatic alterations in non-coding regulatory regions and chromatin modifier genes [41]. A recent landmark study also profiled the chromatin accessibility landscape across human cancers, characterizing their functional role in wiring aberrant transcriptional programs [13]. This has prompted mechanistic interrogations into how structural chromatin changes enable carcinogenesis and if it engenders specific molecular vulnerabilities that can be therapeutically exploited. In this study, we developed and characterized a novel PROTAC degrader of essential ATPase subunits of the SWI/SNF complex, which is recurrently altered in over 20% of human cancers [18,19,42]. Using our pharmacologic agent, we carry out a comprehensive cytotoxicity assessment of SWI/SNF ATPase inactivation in a panel of 65 cell lines from 14 distinct lineages. Surprisingly, even more profoundly than SMARCA4-mutated cancers, we found enhancer-binding transcription factor-driven prostate and plasma cell tumors to acutely rely on the ATP-dependent chromatin remodeling activity of the SWI/SNF complex for growth and survival. Interestingly, neither of the AU-15330-sensitive tumor types show a high recurrence of coding alterations in SWI/SNF subunit genes, such as ARID1A/B, SMARCA4, etc. [24–26]. This positions the wild-type SWI/SNF complex as a requisite cofactor of the oncogenic enhanceosome complexes that, using energy from ATP hydrolysis, displaces nucleosomes at enhancer elements to enable hyper-proliferative gene programs in cancer cells.

Several mechanistic studies have suggested a neomorphic role for the residual complex in SWI/SNF-mutant tumors, prompting the development of innovative strategies to target the aberrant SWI/SNF complex [21,22,43]. For instance, targeting specific molecular vulnerabilities associated with SMARCA4 loss in non-small cell lung cancer has been explored [44].

Interestingly, using AU-15330, we found simultaneous degradation of both SWI/SNF ATPases to induce a potent cytotoxic response in tumors that harbor mutations in transcriptional regulators, in addition to inhibiting the growth of SMARCA4-null cancers. Using genetic techniques, we confirmed the anti-tumor efficacy of AU-15330 to indeed stem from the concurrent loss of both ATPases. Notably, AU-15330 treatment had no effect in several models of lung or pancreatic cancer that are addicted to aberrant kinase signaling (e.g., NCI-H460, MIA PaCa-2 ; **Figure 3-1E**). Furthermore, SWI/SNF ATPase inactivation was well-tolerated in adult mice, showing no major toxicities to the physiologic functions of essential organs. Altogether, these findings demonstrate that dual inhibition of SWI/SNF ATPases is a viable and potent therapeutic approach in enhancer-addicted as well as SWI/SNF-aberrant tumors, subject to the assessment of its efficacy and tolerability in human clinical trials.

Lastly, from a translational standpoint, SWI/SNF degraders are likely to synergize with the existing lines of therapy that target transcription factor oncogenes, such as the AR-targeted drugs enzalutamide and apalutamide in prostate cancer. Complementing the direct antagonistic effects of these drugs, simultaneous degradation of SWI/SNF ATPases would physically impede access of the oncogenic transcriptional machinery (e.g., AR/FOXA1 signaling complex) to the enhancer landscape as well as disrupt enhancer-wired, supra-physiologic expression of the transcription factor oncogenes themselves (e.g., *AR*, *FOXA1*, *ERG*, *MYC*, etc.). This will establish a new treatment paradigm for enhancer-addicted cancers, wherein transcription factor-targeted therapies can be synergistically combined with SWI/SNF degraders either, to 1) derive a more potent anti-tumor response in the primary disease or 2) extort response in therapy-resistant, advanced tumors (e.g., castration and/or enzalutamide-resistant prostate cancer). Thus, further development and clinical evaluation of SWI/SNF-targeted therapeutics hold great promise towards preventing or delaying the emergence of resistance to transcription factor-targeted therapies, which is currently a major challenge in the clinic [45,46]. In line with this, a recent study implicated an altered SWI/SNF complex in enabling the transition of prostate adenocarcinoma to a neuroendocrine disease [47]. AU-15330 thus might be potent even in the therapy-emergent AR-independent prostate cancer, which is consistent with AR-negative prostate cancer cells being moderately sensitive to AU-15330 in our pan-cancer cytotoxicity screen (**Figure 3-1E**).

In summary, here we report AU-15330 as a novel, highly-specific PROTAC degrader of SWI/SNF ATPase components, SMARCA2, SMARCA4, and PBRM1, that shows preferential cytotoxicity in transcription factor-addicted cancers at low nanomolar concentrations. Our study identifies the SWI/SNF complex as a novel transcriptional dependency in AR/FOXA1-driven prostate cancers and, by extension, MYC-driven multiple myelomas. Mechanistically, we show its inhibition to induce a rapid, complete, and targeted loss in chromatin accessibility at the core-enhancer circuitry of AR, FOXA1, MYC, and ERG, thereby attenuating their cancer-promoting transcriptional programs as well as tempering the enhancer-wired supra-physiologic expression of driver oncogenes (**Figure 3-4M**). This work is consistent with prior reports suggesting an oncogenic role of the SWI/SNF complex in prostate cancer [30,47–50]; as well as, for the first time, implicates the SWI/SNF complex as an essential chromatin regulator in multiple myeloma. To our knowledge, this is the first study to demonstrate that physical chromatin accessibility can be modulated at non-coding regulatory elements as a novel therapeutic strategy in cancer treatment.

SWI/SNF ATPase degraders add to the growing arsenal of chromatin-targeted therapeutics for directly combating enhancer-addiction in human cancers, warranting their safety and efficacy assessments in clinical trials.

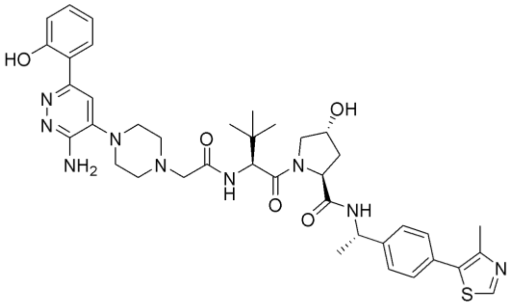
Materials and methods

Chemical structure and synthesis of AU-15330

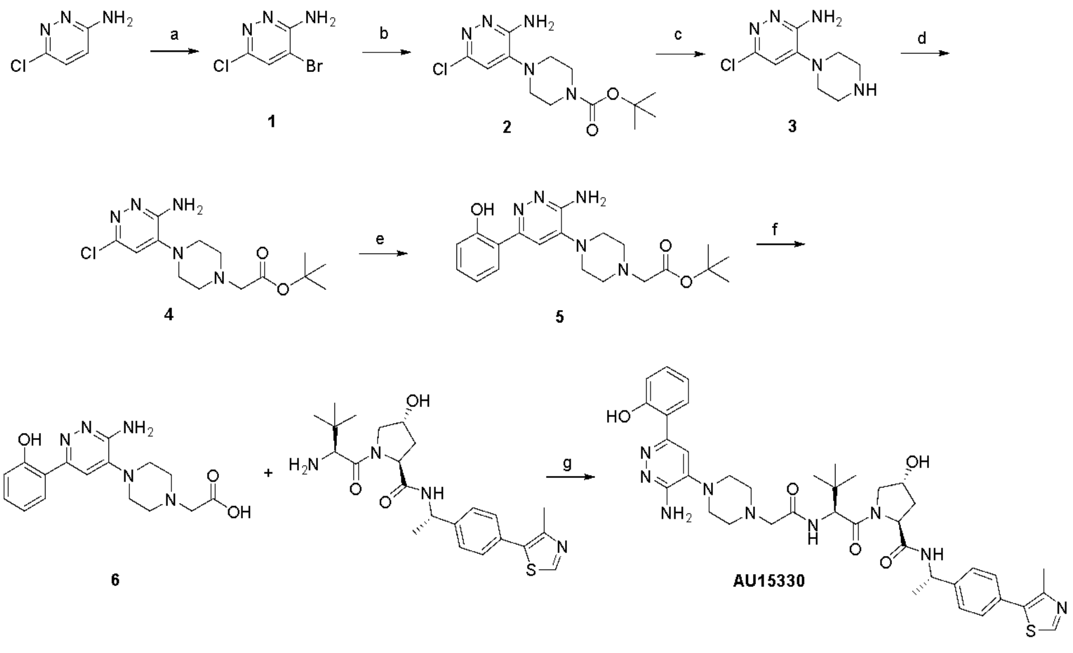
General Information: All chemicals and solvents were obtained from commercial suppliers and used without further purification. Purification was performed using combi-flash Nextgen300. All reactions were monitored by TLC, using silica gel plates with fluorescence F254 and UV light visualization. ^{1}H NMR spectra was recorded on a Varian Mercury Plus at 400 MHz and ^{13}C NMR spectra was recorded on a JEOL-ECZ-400S spectrometer at 100 MHz. Coupling constants (J) are expressed in hertz (Hz). Chemical shifts (δ) of NMR are reported in parts per million (ppm) units relative to internal control (TMS). Signal splitting patterns are described as singlet (s), doublet (d), triplet (t), quartet (q), multiplet (m), broad (br) or a combination thereof. The low resolution of ESI-MS was recorded on an Agilent-6120 and the high-resolution mass (resolution-70000) for the compound was generated using Q-Exactive Plus orbitrap system, Thermo Scientific, US using electrospray ionization (ESI). HPLC was recorded Waters 2696, the column used was a YMC Triart C-18 EXRS (150*4.6) mm 5µm using 0.01M Ammonium acetate in (Aq); Mobile phase-B:ACN 100%; Method -T/%B: 0/10, 2/10, 5/85, 13/85, 14/10, 15/10 method and Flow rate: 1.0 ml/min.

Abbreviations used: DMSO for dimethyl sulfoxide, DIPEA for *N*, *N*-diisopropylethylamine, MeOH for methanol, DMF for *N*, *N*-dimethylformamide, HATU for 1- [bis(dimethylamino) methylene]-1*H*-1,2,3-triazolo[4,5-*b*]pyridinium 3-oxid hexafluorophosphate, DCM for dichloromethane, Pd(dppf)Cl$_2$ for [1,1′-bis(diphenylphosphino)ferrocene]dichloro-palladium(II).

Final chemical structure of AU-15330:

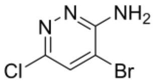


Synthesis:



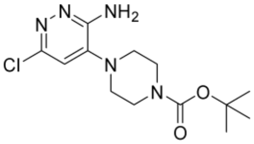
Reagents and conditions: a) Br$_2$, NaHCO$_3$, MeOH, 0 °C - RT, 16 h; b) DMF, 90 °C, 16h; c) 4M HCl in 1,4-dioxane, DCM, 0 °C - RT, 16h; d) DIPEA, DMF, 60 °C, 16h; e) Pd(dppf)Cl$_2$•DCM, K$_2$CO$_3$, dioxane, water, sealed tube, 120 °C, 6 h; f) 4 M HCl in 1,4-dioxane, DCM, 0 °C - RT, 16h; g) HATU, DIPEA, DMF, RT, 4 h.

4-bromo-6-chloropyridazin-3-amine (1)



To a stirred solution of 6-chloropyridazin-3-amine (20.0 g, 155.02 mmol) in MeOH (100 mL) was added $NaHCO_3$ (19.53 g, 232.00 mmol) at RT and stirred for 15 min and then bromine (8.74 mL, 170.52 mmol) was added drop wise to the reaction mixture over period of 1 h at 0 °C and stirred for 16 h at RT. After completion of the reaction (monitored by TLC), reaction mixture was quenched with water (100 mL). Dark brown colored solid was precipitated which was filtered and washed with water (50 mL) and dried under vacuum to afford crude **1**. The solid was washed with 20% EtOAc in hexane and diethyl ether to afford pure **1** (15.0 g, 46%). ^{1}H NMR (400 MHz, DMSO-d_6) δ 7.99 (s, 1H), 6.97 (bs, 2H). LC-MS: *m/z* 208.0 $[M+H]^+$.

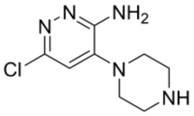
tert-butyl 4-(3-amino-6-chloropyridazin-4-yl)piperazine-1-carboxylate (2).



To a stirred solution of 4-bromo-6-chloropyridazin-3-amine **1** (20.0 g, 96.66 mmol) in DMF (400 mL) was added tert-butyl piperazine-1-carboxylate (53.92 g, 289.9 mmol) at RT and stirred for 16 h at 90 ºC under nitrogen atmosphere. Then the reaction mixture was quenched with cold water

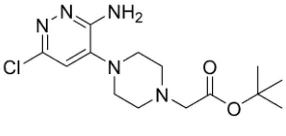
(200 mL) and the brown solid obtained was washed with diethyl ether, filtered and dried under vacuum. The same procedure was repeated four times to afford pure **2** (15.0 g, 49.57%). ^{1}H NMR (400 MHz, DMSO-*d$_6$*) *δ* 6.91 (s, 1H), 6.21 (s, 2H), 3.49 (t, *J* = 4.4 Hz, 4H), 2.92 (t, *J* = 4.8 Hz, 4H), 1.41 (m, 9H). LC-MS: *m/z* 314.2 $[M+H]^+$.

6-chloro-4-(piperazin-1-yl)pyridazin-3-amine hydrochloride (3).



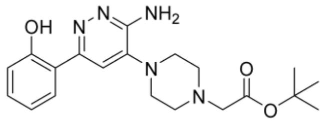
To a stirred solution of tert-butyl 4-(3-amino-6-chloropyridazin-4-yl)piperazine-1-carboxylate **2** (15.0 g, 47.92 mmol) in DCM (100 mL) was added 4 M HCl in 1,4-dioxane (75 mL) at 0 °C under nitrogen atmosphere and stirred for 16 h at RT. The reaction mixture was concentrated under reduced pressure to get a crude brown solid compound. The brown solid was washed with diethyl ether (2x 50 mL), filtered and dried under vacuum to afford **3** (11.98 g, 100%). ^{1}H NMR (400 MHz, DMSO-*d$_6$*) δ 9.44 (bs, 2H), 7.75 (bs, 2H), 7.32 (s, 1H), 3.30-3.25 (m, 8H). LC-MS: *m/z* 214.1 $[M+H]^+$.

tert-butyl 2-(4-(3-amino-6-chloropyridazin-4-yl)piperazin-1-yl)acetate (4).



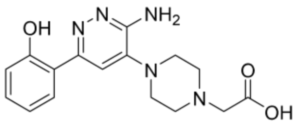
To a stirred solution of 6-chloro-4-(piperazin-1-yl)pyridazin-3-amine hydrochloride **3** (12.0 g, 48.18 mmol) in DMF (100 mL) in a sealed tube were added DIPEA (25.70 mL, 144.54 mmol) and tert-butyl 2-bromoacetate (10.53 mL, 72.27 mmol) at RT and stirred for 16 h at 60 ºC. The reaction mixture was quenched with water (100 mL) and extracted with EtOAc (2 X 150 mL). The combined organic layer was washed with water (100 mL), brine (100 mL), dried over anhydrous sodium sulphate and concentrated under reduced pressure to give the crude product which was purified by combi flash using 60% ethyl acetate in hexane as eluent to afford **4** (10.0 g, 63.7%). ^{1}H NMR (400 MHz, DMSO-*d*$_6$) *δ* 6.89 (s, 1H), 6.07 (s, 2H), 3.165 (s, 2H), 3.05-2.95 (m, 4H) 2.70-2.65 (m, 4H), 1.45 (s, 9H). LC-MS: *m/z* 328.2 [M+H]$^+$.

tert-butyl 2-(4-(3-amino-6-(2-hydroxyphenyl)pyridazin-4-yl)piperazin-1-yl)acetate (5).



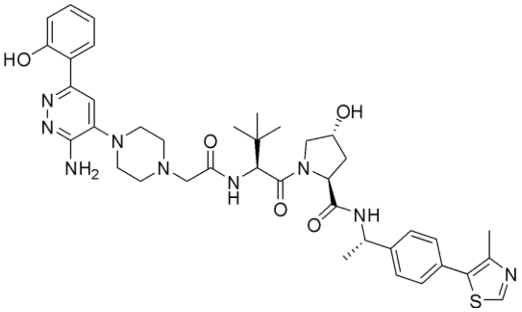
To a stirred solution of tert-butyl 2-(4-(3-amino-6-chloropyridazin-4-yl)piperazin-1-yl)acetate **4** (1.8g, 5.47 mmol) and (2-hydroxyphenyl)boronic acid (1.7g, 10.94 mmol) in 1,4-dioxane (20 mL) was added K_2CO_3 (2M) solution (4.52 g, 32.8 mmol) and degassed with nitrogen for 5 min. followed by Pd(dppf)Cl_2•DCM (0.44 g, 0.54 mmol) was added and the reaction mixture was heated for 6h at 120 ºC in a sealed tube. Once the reaction was completed (monitored by TLC), the reaction mixture was diluted with EtOAc. The combined organic layer was washed with water, brine, dried over anhydrous sodium sulphate and concentrated under vacuum to give the residue which was purified by combi flash column chromatography using 50-60% ethyl acetate in hexane as eluent to afford **5** (0.9 g, 63.7%).^{1}H NMR (400 MHz, DMSO-d_6) δ 14.15 (s, 1H), 7.94 (dd, J_1 = 1.6, J_2 = 8.4 Hz,1H), 7.26-7.21 (m, 2H), 6.90-6.87 (m, 2H), 6.22 (s, 2H), 3.18-3.16 (m, 2H), 3.11-3.09 (m, 4H), 2.67-2.66 (m, 4H), 1.43 (s, 9H). LC-MS: *m/z* 386.1 [M+H]$^+$.

2-(4-(3-amino-6-(2-hydroxyphenyl)pyridazin-4-yl)piperazin-1-yl)acetic acid (6).



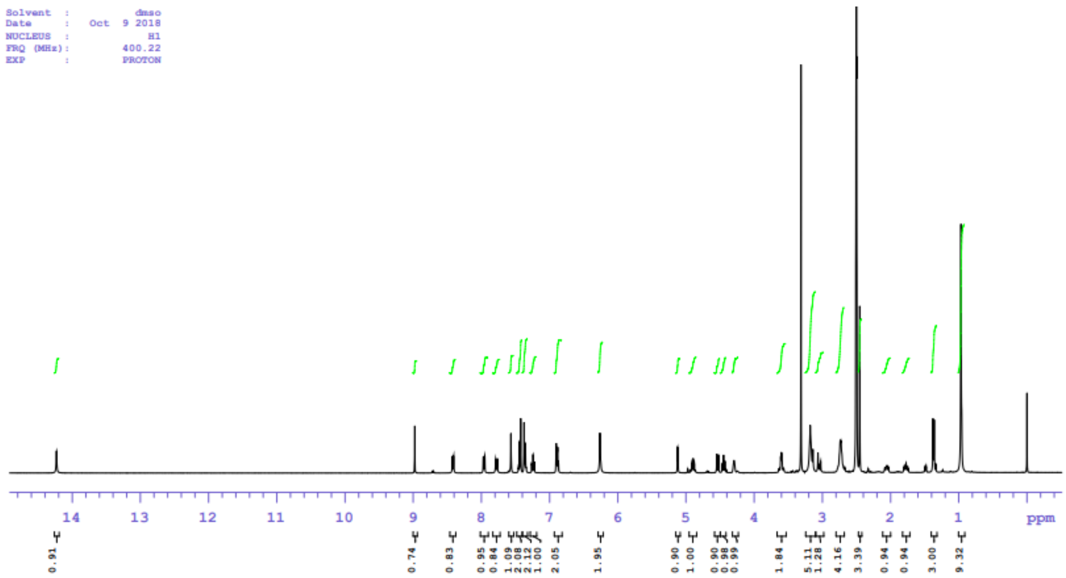
To a stirred solution of tert-butyl 2-(4-(3-amino-6-(2-hydroxyphenyl)pyridazin-4-yl)piperazin-1-yl)acetate **5** (5 g, 0.258 mmol.) in DCM (50 vol.) was added 4 M HCl in 1, 4-Dioxane (75 mL) at 0 °C under nitrogen atmosphere and stirred for 16 h at RT. After completion of the reaction (monitored by TLC), reaction mixture was concentrated under reduced pressure. The residue was washed with diethyl ether, filtered and dried under vacuum to afford compound **6** (4.27 g, 95%). ^{1}H NMR (400 MHz, DMSO-d_6) δ 7.72 (s, 1H), 7.64 (d, J_1 = 1.6, J = 7.6 Hz, 1H), 7.43 (t, J = 7.6 Hz, 1H), 7.04 (m, 2H), 4.25 (s, 2H), 3.72 (bs, 8H). LC-MS: *m/z* 330.1 [M+H]$^+$.

(2S,4R)-1-((S)-2-(2-(4-(3-amino-6-(2-hydroxyphenyl)pyridazin-4-yl)piperazin-1-yl)acetamido)-3,3-dimethylbutanoyl)-4-hydroxy-N-((S)-1-(4-(4-methylthiazol-5-yl)phenyl)ethyl)pyrrolidine-2-carboxamide (AU-15330)

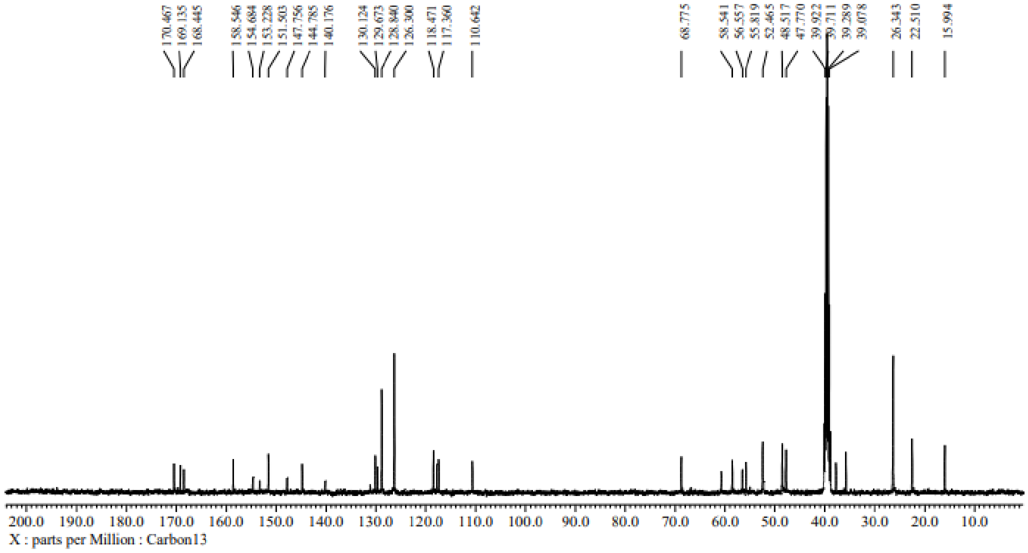


To a stirred solution of 2-(4-(3-amino-6-(2-hydroxyphenyl)pyridazin-4-yl)piperazin-1-yl)acetic acid **6** (0.3 g, 0.91 mmol) and (2S,4R)-1-((S)-2-amino-3,3-dimethylbutanoyl) -4-hydroxy-N-((S)-1-(4-(4-methylthiazol-5-yl)phenyl)ethyl)pyrrolidine-2-carboxamide (0.65 g, 1.36 mmol) in DMF (3 mL) was added HATU (0.51 g, 1.336 mmol) at 0 ºC followed by the drop wise addition of DIPEA (0.65 mL, 3.64 mmol). Stirring was continued at RT for 16 h. After completion of the reaction (monitored by TLC) the reaction mixture was poured into crushed ice. Solid formed was filtered off to get the crude product which was purified by combi flash using 4 % MeOH in DCM as eluent to afford **AU-15330** (0.25 g, 37%) as a pale yellow solid. HPLC purity 99.4% (RT- 6.54 min); ^{1}H NMR (400 MHz, DMSO-d_6) δ 14.22 (s, 1H), 8.97 (s, 1H), 8.41 (d, J = 7.8 Hz, 1H), 7.95 (d, J = 7.4 Hz, 1H), 7.77 (d, J = 9.8 Hz, 1H), 7.56 (s, 1H), 7.43 (d, J = 8.3 Hz, 2H), 7.36 (d, J = 8.3 Hz, 2H), 7.25 (t, J = 6.8 Hz, 1H), 6.88 (dd, J_1 = 5.8 Hz, J_2 = 4.4 Hz, 2H), 6.25 (s, 2H), 5.11 (d, J = 3.5 Hz, 1H), 4.89 (d, J = 7.3 Hz, 1H), 4.53 (d, J = 9.8 Hz, 1H), 4.44 (t, J = 7.8 Hz, 1H), 4.29 (bs, 1H), 3.60 (d, J = 5.9 Hz, 2H), 3.17 (bs, 4H), 3.13 (s, 1H), 3.0 (d, J = 16.0 Hz, 1H), 2.69 (dd, J_1 = 4.4 Hz, J_2 = 2.0 Hz, 4H), 2.45 (s, 3H), 2.08 (d, J = 13.2 Hz, 1H), 1.81-1.74 (m, 1H), 1.36 (d, J = 6.9 Hz, 3H), 0.96 (s, 9H). ^{13}C NMR (100 MHz, DMSO-d_6) δ 170.4, 169.1, 168.4, 158.5, 154.6, 153.2, 151.5, 147.7, 144.7, 140.1, 131.1, 130.1, 129.6, 128.8, 126.3, 118.4, 117.7, 117.3, 110.6, 68.7, 60.7, 58.5, 56.5, 55.8, 52.4, 48.5, 47.7, 40.1, 39.9, 39.0, 38.8, 37.7, 35.8, 26.3, 22.5, 15.9. HRMS (ESI) for $C_{39}H_{49}N_9O_5S$ $[M+H]^+$ calculated 756.3611, found 756.3635.

Supporting information: S1: ^{1}*H NMR of AU-15330:*



S2: ^{13}*C NMR of AU-15330\:*



Computational modeling of AU-15330 - SMARCA2-Bromodomain binding

The binding model of AU-15330 in complex with SMARCA2-BD and VHL was generated using Aurigene's proprietary computing algorithm, ALMOND (ALgorithm for MOdeling Neosubstrate Degraders). The algorithm is developed using the ICM-Pro integrated modeling platform (http://www.molsoft.com/icm_pro.html) and trained to predict models of ternary complexes of bi-functional molecules with very short or no linkers. The process employs protein-protein docking simulation, exhaustive conformational sampling, small molecule-protein docking, and site-directed scoring of predicted ternary complex models. The computed score estimates the force of induced interactions in the predicted target – E3 ligase complex and used as a basis for prioritization of degrader binding models. The images were prepared using PyMOL(https://www.schrodinger.com/products/pymol).

Cell lines, antibodies, and compounds

Most cell lines were originally obtained from ATCC, DSMZ, ECACC, or internal stock. C4-2B cells were generously provided by Evan Keller, Ph.D. at the University of Michigan, CWR-R1 cells, and a series of enzalutamide-resistant prostate cancer cell lines (LNCaP_Parental, LNCaP_EnzR, CWR-R1_Parental, CWR-R1_EnzR, VCaP_Parental and VCaP_EnzR) were generously provided by Donald Vander Griend, Ph.D. at the University of Illinois at Chicago. All the cells were genotyped to confirm their identity at the University of Michigan Sequencing Core and tested routinely for Mycoplasma contamination. Additionally, all the cell lines were genotyped every two months to confirm their identity. LNCaP, 22RV-1, CWR-R1, PC-3, and DU145 were grown in Gibco RPMI-1640 + 10% FBS (ThermoFisher, Waltham, MA). VCaP was grown in Gibco DMEM + 10% FBS (ThermoFisher, Waltham, MA). Sources of all antibodies are described in **Supplementary Table S2**. AU-15330 were synthesized by Aurigene, dBRD9 and VZ 185 were purchased from Tocris Bioscience (Bristol, UK) and MDV3100 was purchased from Selleck Chemicals (Huston, TX).

Cell viability assay

Cells were plated onto 96-well plates in their respective culture medium and incubated at 37C in an atmosphere of 5% CO2. After overnight incubation, a serial dilution of compounds was prepared and added to the plate. The cells were further incubated for 5 days, and the CellTiter-Glo assay (Promega, Madison, WI) was then performed according to the manufacturer's instruction to determine cell proliferation. The luminescence signal from each well was acquired using the Infinite M1000 Pro plate reader (Tecan, Zürich, Switzerland), and the data was analyzed using the GraphPad Prism software (GraphPad Software Inc, La Jolla, CA).

Incucyte proliferation assays/Caspase-3/7 Green Apoptosis Assay

A total of 4,000 cells per well were seeded in clear 96-well plates. After overnight incubation, compounds were added to the cells at logarithmic dose series. One day and 8 days after seeding, cellular ATP content was measured using CellTiterGlo (Promega). Measurements after 8 days were divided by the measurement after 1 day (that is, the T0 plate) to derive fold proliferation.

For online analysis of cell growth, 4,000 cells per well were seeded in clear 96-well plates (costar no. 3513). IncuCyte Caspase-3/7 Green Apoptosis Assay Reagent (1:1,000, Essen BioScience no. 4440) was added and cells were incubated at 37 °C and 5% CO2 overnight. On the next day, compounds were added at the desired concentration using the HP digital dispenser D300 and plates were read in an Incucyte ZOOM. Every 2 h, phase object confluence (percentage area) for proliferation and green object count for apoptosis were measured. Values for apoptosis were normalized for the total number of cells.

Western Blot and Immunoprecipitations

Cell lysates were prepared in RIPA buffers (ThermoFisher Scientific, Waltham, MA) supplemented with cOmpleteTM protease inhibitor cocktail tablets (Sigma-Aldrich, St. Louis, MO), and total protein was measured by Pierce™ BCA Protein Assay Kit (ThermoFisher Scientific, Waltham, MA). An equal amount of protein was resolved in NuPAGE™ 3 to 8%, Tris-Acetate Protein Gel (ThermoFisher Scientific, Waltham, MA) or NuPAGE™ 4 to 12%, Bis-Tris Protein Gel (ThermoFisher Scientific, Waltham, MA), and blotted with primary antibodies. Following incubation with HRP-conjugated secondary antibodies membranes were imaged on an Odyssey CLx Imager (LiCOR Biosciences. Lincoln, NE).

Immunoprecipitations were performed in LNCaP and VCaP cells treated as described. 600 µg of nuclear extracts isolated using the NE-PER™ Nuclear and Cytoplasmic Extraction Reagents (ThermoFisher Scientific, Waltham, MA) were immunoprecipitated with BAF155, AR, FOXA1, ERG antibodies according to manufacturer protocol. Eluted proteins were subjected to Western blot or mass spectrometry analysis.

RNA isolation and quantitative real-time PCR

Total RNA was isolated from cells using the Direct-zol kit (Zymo, Irvine, CA), and cDNA was synthesized from 1,000 ng total RNA using Maxima First Strand cDNA Synthesis Kit for RT-qPCR (Thermo Fisher Scientific, Waltham, MA). Quantitative real-time PCR was performed in triplicate using standard SYBR green reagents and protocols on QuantStudio 5 Real-Time PCR system (Applied Biosystems). The target mRNA expression was quantified using the ΔΔCt method and normalized to ACTB expression. All primers were designed using Primer 3 (http://frodo.wi.mit.edu/primer3/) and synthesized by Integrated DNA Technologies (Coralville, IA). Primer sequences are listed in **Supplementary Table S2**.

CRISPR Knock-out and inducible shRNA knockdown

Guide RNAs (sgRNAs) targeting the exons of human SMARCA2/BRM or SMARCA4/BRM were designed using the Benchling (https://www.benchling.com/). Non-targeting sgRNA, SMARCA2/BRM or SMARCA4/BRM targeting sgRNAs were cloned into lentiCRISPR v2 plasmid according to published literature [51], lentiCRISPR v2 plasmid was a gift from Feng Zhang (Addgene plasmid #52961). LNCaP cells were transiently transfected with lentiCRISPR v2 encoding non-targeting or pool of three independent SMARCA2/BRM or SMARCA4/BRM-targeting sgRNAs. Twenty-four hours after transfection, cells were selected with 1 μg/mL puromycin for three days. Western blot was performed to examine the knock-out efficiency. The sgRNA sequences are listed in **Supplementary Table S2**.

ATAC-seq and analysis

ATAC-seq was performed as previously described[27]. In brief, 50,000 cancer cells were washed in cold PBS and resuspended in cytoplasmic lysis buffer (CER-I from the NE-PER kit, Invitrogen, cat. no. 78833). This single-cell suspension was incubated on ice for 5-8 min (depending on the cell line) with gentle mixing by pipetting every 2 min. The lysate was centrifuged at 1,300g for 5 min at 4 °C. Nuclei were resuspended in 50μl of 1× TD buffer, then incubated with 2-2.5 μl Tn5 enzyme for 30 min at 37 °C (Nextera DNA Library Preparation Kit; cat. no. FC-121-1031). Samples were immediately purified by Qiagen minElute column and PCR-amplified with the NEBNext High-Fidelity 2X PCR Master Mix (NEB; cat. no. M0541L) following the original protocol[27]. qPCR was used to determine the optimal PCR cycles to prevent over-amplification. The amplified library was further purified by Qiagen minElute column and SPRI beads (Beckman Coulter; cat. no. A63881). ATAC-seq libraries were sequenced on the Illumina HiSeq 2500 (125-nucleotide read length).

Paired-end .fastq files were trimmed and uniquely aligned to the GRCh38/hg38 human genome assembly using Novoalign (Novocraft) (with the parameters -r None -k -q 13 -k -t 60 -o sam –a CTGTCTCTTATACACATCT), and converted to .bam files using SAMtools (version 1.3.1). Reads mapped to mitochondrial or duplicated reads were removed by SAMtools and PICARD MarkDuplicates (version 2.9.0), respectively. Filtered .bam files from replicates were merged for downstream analysis. MACS2 (2.1.1.20160309) was used to call ATAC-seq peaks. The coverage tracks were generated using the program bam2wig (http://search.cpan.org/dist/Bio-ToolBox/) with the following parameters:–pe–rpm–span–bw. Bigwig files were then visualized using the IGV (Broad Institute) open-source genome browser and the final figures were assembled using Adobe Illustrator.

De novo and known motif enrichment analysis

All de novo and known motif enrichment analyses were performed using the HOMER (v.4.10) suite of algorithms42. Peaks were called by the findPeaks function (-style factor -o auto) at 0.1% false discovery rate; de novo motif discovery and enrichment analysis of known motifs were performed with findMotifsGenome.pl (--size given --mask). The top 10 motifs from the results are shown and motifs were generally ascribed to the protein family instead of specific family members (unless known).

RNA-seq and analysis

RNA-seq libraries were prepared using 200–1,000 ng of total RNA. PolyA+ RNA isolation, cDNA synthesis, end-repair, A-base addition, and ligation of the Illumina indexed adapters were performed according to the TruSeq RNA protocol (Illumina). Libraries were size selected for 250–300 bp cDNA fragments on a 3% Nusieve 3:1 (Lonza) gel, recovered using QIAEX II reagents (QIAGEN), and PCR amplified using Phusion DNA polymerase (New England Biolabs). Library quality was measured on an Agilent 2100 Bioanalyzer for product size and concentration. Paired-end libraries were sequenced with the Illumina HiSeq 2500, (2×100 nucleotide read length) with sequence coverage to 15-20M paired reads.

Libraries passing quality control were trimmed of sequencing adaptors and aligned to the human reference genome, GRCh38. Samples were demultiplexed into paired-end reads using Illumina's bcl2fastq conversion software v2.20. The reference genome was indexed using bowtie2-build and reads were aligned onto the GRCh38/hg38 human reference genome using TopHat2 [52] with strand-specificity and allowing only for the best match for each read. The aligned file was used to calculate strand-specific read count for each gene using HTSeq-count[53]. EdgeR [54] was used to compute differential gene expression using raw read-counts as input. Heatmaps were generated using the ComplexHeatmap[55] package in R. For gene enrichment analysis (GSEA), we first defined ERG and FOXA1 gene signatures from VCaP or LNCaP cells treated with control siRNA or siRNA targeting ERG[30] of FOXA1 (generated in this study) containing 250 significantly down-regulated genes. For AR and MYC, the Hallmark gene signatures were used. These gene signatures were used to perform a fast pre-ranked GSEA using fgsea bioconductor package[56] in R. We used the function fgsea to estimate the net enrichment score and p-value of each pathway and the plotEnrichment function was used to plot enrichment for the pathways of interest.

Chromatin immunoprecipitation with massively parallel DNA sequencing (ChIP-seq) assay and data analysis

ChIP experiments were carried out using the HighCell# ChIP-Protein G kit (Diagenode) as per the manufacturer's protocol. Chromatin from five million cells was used per ChIP reaction with 10 μg of the target protein antibody. In brief, cells were trypsinized and washed twice with 1× PBS, followed by cross-linking for 8 min in 1% formaldehyde solution. Crosslinking was terminated by the addition of 1/10 volume 1.25 M glycine for 5 min at room temperature followed by cell lysis and sonication (Bioruptor, Diagenode), resulting in an average chromatin fragment size of 200 bp. Fragmented chromatin was then used for immunoprecipitation using various antibodies, with overnight incubation at 4 °C. ChIP DNA was de-crosslinked and purified using the iPure Kit V2 (Diagenode) using the standard protocol. Purified DNA was then prepared for sequencing as per the manufacturer's instructions (Illumina). ChIP samples (1–10 ng) were converted to blunt-ended fragments using T4 DNA polymerase, E. coli DNA polymerase I large fragment (Klenow polymerase), and T4 polynucleotide kinase (New England BioLabs (NEB)). A single A base was added to fragment ends by Klenow fragment (3′ to 5′ exo minus; NEB) followed by ligation of Illumina adaptors (Quick ligase, NEB). The adaptor-ligated DNA fragments were enriched by PCR using the Illumina Barcode primers and Phusion DNA polymerase (NEB). PCR products were size-selected using 3% NuSieve agarose gels (Lonza) followed by gel extraction using QIAEX II reagents (Qiagen). Libraries were quantified and quality checked using the Bioanalyzer 2100 (Agilent) and sequenced on the Illumina HiSeq 2500 Sequencer (125-nucleotide read length). Paired-end, 125 bp reads were trimmed and aligned to the human reference genome (GRC h38/hg38) with the Burrows-Wheeler Aligner (BWA 0.7.17-r1198)[57]. The SAM file obtained after alignment was converted into BAM format using SAMTools (Version: 1.9). MACS2 *callpeak* was used for performing peaking calling with the following option: 'macs2 callpeak–call-summits–verbose 3 -g hs -f BAM -n OUT–qvalue 0.05'. For H3K27ac data --broad option was used. Using deepTools bamCoverage, a coverage file (bigWig format) for each sample was created. The coverage is calculated as the number of reads per bin, where bins are short consecutive counting windows. While creating the coverage file, the data was normalized with respect to each library size. ChIP peak profile plots and read-density heat maps were generated using deepTools, and cistrome overlap analyses were carried out using the ChIPpeakAnno or ChIPseeker packages in R.

HiC coupled with chromatin immunoprecipitation with massively parallel DNA sequencing (HiChIPseq) assay and data analysis

HiChIP assay was performed on $5x10^6$ DMSO or AU-15330 treated VCaP cells. Frozen cells were resuspended in 1X PBS and crosslinked with 3mM DSG and 1% formaldehyde. Washed cells were digested with 0.5 uL MNase in 100 uL of Nuclease digest buffer with $MgCl_2$. Cells were lysed with 1X RIPA and clarified lysate (approximately 1400 ng) was used for ChIP. The antibody amount used per ChIP and vendor information are as follows: CTCF – 1.14 μg of Cell Signaling, Cat. 3418; H3K4me3 – 3.4 μg of Cell Signaling, Cat. 9751; H3K27ac – 0.4 μg of Cell Signaling, Cat. 8173. The Protein A/G bead pulldown, proximity ligation, and libraries were prepared as described in the Dovetail protocol (Dovetail™ HiChIP MNase Kit). Libraries were sequenced on an Illumina HiSeq 4000. Raw fastq files were aligned using BWA mem with the -5SP options with an index containing only the main chromosome from the human genome release hg38 (available from the UCSC genome). The aligned paired reads were annotated with pairtools parse (https://github.com/open2c/pairtools) with the following options --min-mapq 40 --walks-policy 5unique --max-inter-align-gap 30 and the --chroms-path file corresponding to the size of the chromosome used for the alignment index. The paired reads were further processed to remove duplicated reads, sorted with unaligned reads removed with the pairtools sort and the pairtools dedup tools with the basic option to produce an alignment file in the bam format as well as the location of the valid pair. The valid pairs were finally converted to the .cool and .mcool format using the cooler cload and cooler zoomify tools [58] and to the .hic format using the juicer tool [59].

For the generation of the Aggregate Peak Analyses (APA) plots, we used the HiCExplorer tools and the hicAggregateContacts command with --*range* 50000:100000 --*numberOfBins* 30. Plots for all chromosomes were individually computed and summated to generate the global APA plots. The ComplexHeatmaps package [55] in R was used for the generation of the final heatmap. For HiChIP contact heatmap, .hic files were uploaded to the WashU Epigenome Browser (https://epigenomegateway.wustl.edu/) and screenshots from gene loci-of-interest were downloaded using the default viewing conditions.

Super-enhancer analysis

Super enhancer regions were identified with findPeaks function from HOMER[60] using options "-style super -o auto". In addition, the option "-superSlope -1000" was added to include all potential peaks, which were used to generate the super-enhancer plot (super-enhancer score versus ranked peaks). The slope value of greater than or equal to 1 was used to indetify super-enhancer clusters. The input files to findPeaks were tag directories generated from alignment files in SAM format with makeTagDirectory function from HOMER.

Preclinical safety evaluation of AU-15330 in male CD-1 mice

<u>Vehicle and Formulation</u>

Dose (0, 10 & 30 mg/kg/day)	:	10% w/v HPβCD + 5% w/v Dextrose + Purified water q.s.
Dose (60 mg/kg, (2+/5-) for 2 cycles or Once every three days)	:	10% w/v HPβCD + 5% w/v Dextrose + Purified water q.s. (pH adjusted to ~4.5 to 4.8 with 1N HCl)

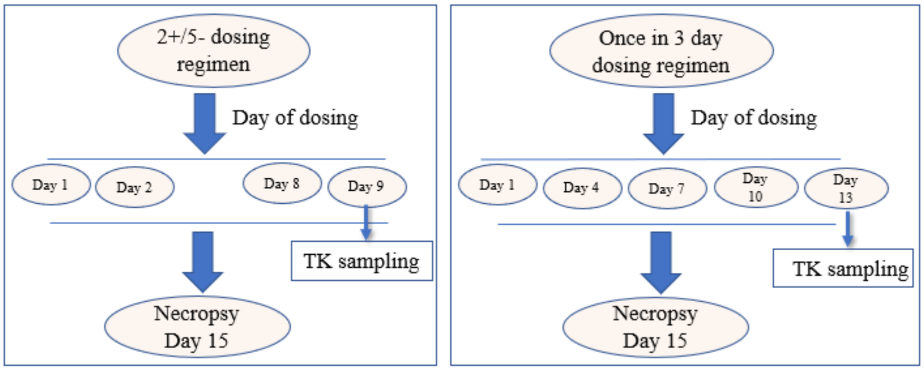
Intravenous formulation preparation details

The test item AU-15330 was formulated in 10% w/v HPβCD + 5% w/v Dextrose + Purified water q.s for 10 and 30 mg/kg doses. AU-15330 was formulated in 10% w/v HPβCD + 5% w/v Dextrose + Purified water q.s with pH adjusted to ~4.5 to 4.8 with 1N HCl for 60 mg/kg infrequent intravenous administration. Formulation was stored at room temperature (~25°C) in a screw capped glass bottle until use and should be dosed within 2h of its preparation.

Test system: Male Hsd: ICR (CD-1®) mouse of 7-9 week age (6 animals/group, main study and 3 animals/group, toxicokinetic satellite group) were selected as test species.

Treatment: AU-15330 was administered once daily up to 7 days at doses of 0, 10, and 30 mg/kg/day by slow (10 sec to 15 sec) intravenous injection through the tail vein.

In intermittent dosing study, the frequency of dose administration, duration of treatment, and dosing procedure was followed as shown below fig (2 on/5 off and once every three days dosing for 2 cycles at 0, 60 mg/kg by slow (10 sec to 15 sec) intravenous injection through the tail vein)



The dose-volume was 5 ml/kg for all groups.

Mortality/Moribundity

Animals were observed for mortality/moribundity at least once daily during the acclimatization period and twice daily during the dosing period (once in the morning hours and once in the afternoon hours).

Clinical Signs of Toxicity

A routine clinical examination was performed at least once daily during the acclimatization period and twice daily during the dosing period ((once in the morning hours (~ within 30 mins of dosing) and once in the afternoon hours).

Body Weight

Individual animal body weights were recorded daily using a weighing balance connected to a printer throughout the pretest treatment periods.

Food Consumption

Pre-weighed amounts of 50 gm food were offered to each cage daily and the quantity of food left in each cage was weighed approximately after 24 hours each day throughout the experimental period. Food consumption (gm/mice/day) was calculated. The formula is mentioned below

$$\text{Food Consumption (gm/mice/day)} = \frac{\text{Total food offered (g)} - \text{Food left over (g)}}{\text{No. of surviving animals}}$$

Clinical pathology (blood sampling procedure)

At scheduled necropsy, mice were anesthetized by isoflurane. Blood samples were collected from the inferior vena cava and animals were exsanguinated after the opening of the abdominal vessels. Food was withdrawn 4-6 hours before blood collection.

Group-wise animals were anesthetized in sequence with Isoflurane ~3-5 mins before blood collection, and blood samples were collected from the inferior vena cava after the opening of the abdominal cavity. Blood samples from each animal were collected into two different tubes, one for hematology and another for clinical chemistry. Additionally, blood smears were prepared for further evaluation from all groups.

Hematology analysis

For evaluation of hematology parameters, approximately 0.4 mL of blood was collected into tubes containing Ethylenediaminetetraacetic acid dipotassium salt dihydrate solution as an anticoagulant (2% K2 EDTA, pH 7.4). The following parameters were measured using ADVIA 2120 hematology analyzer (Siemens):

List of Hematology Parameters Analyzed

Parameter	**Unit**
Hemoglobin (HGB)	g/dL
Red blood cell counts (RBC)	x 10^6 cells/µL
Hematocrit (HCT)	%
Mean corpuscular hemoglobin (MCH)	pg
Mean corpuscular hemoglobin concentration (MCHC)	g/dL
Mean corpuscular volume (MCV)	fL
Platelet count (PLT)	x 10^3 cells/µL
Total white blood cell count (WBC)	x 10^3 cells/µL
Mean Platelet Volume (MPV)	fL
Differential white blood cell count Neutrophil (N) Lymphocytes (L) Basophil (B) Eosinophil (E) Monocyte (M)	% and x 10^3 cells/µL
Reticulocytes count	% and x 10^9 cells/L

Clinical chemistry analysis

For clinical biochemistry estimations, approximately 0.6 mL of blood was collected in plain tubes and allowed to clot at room temperature for ~1 hour. The serum was separated by centrifugation at 5000 rpm for 10 minutes at about 5 °C. The following parameters were measured:

List of Clinical Chemistry Parameters Analyzed

Parameter in Serum	**Unit**	**Abbreviation**	**Instrument**
Glucose	mg/dL	GLUC	(a)
Urea	mg/dL	UREA	(a)
Creatinine	mg/dL	CREA	(a)
Creatine Kinase	U/L	CK	(a)
Total bilirubin	mg/dL	T BIL	(a)
Cholesterol	mg/dL	CHOL	(a)
Triglycerides	mg/dL	TRIGS	(a)
Total protein	g/dL	TP	(a)
Albumin (A)	g/dL	ALB	(a)
Globulin (G)	g/dL	GLOB	Total protein-Albumin
Albumin/globulin ratio	NA	A/G Ratio	Albumin/Globulin
Alanine aminotransferase	U/L	ALT	(a)
Aspartate aminotransferase	U/L	AST	(a)
Alkaline Phosphatase	U/L	ALP	(a)
Gamma-glutamyl transferase	U/L	GGT	(a)
Lactate dehydrogenase	U/L	LDH	(a)
Aldolase	U/L	ALS	(a)

Instruments

a) RX monaco, Randox Autoanalyzer; NA- Not Applicable

Bioanalysis and toxicokinetics

Blood Sampling Time Points and Bleeding Schedule

For daily dosing study (TK sampling time point was followed as mentioned below)

On day 1, TK study animals from each test item treated dose group were bled at 0, 0.25, 1, 3, 7 and 24 hours post-dose from the saphenous vein for drug levels estimation with n=3 animals per time point. On day 7, All animals from each test item treated dose group along with control animals were bled at 0, 0.25, 1, 3, 7 and 24 hours post-dose from the saphenous vein for drug levels estimation with n=3 animals per time point. Additionally, 0.08 & 0.5 hr time point were part of intravenous sampling.

For intermittent dosing study (TK sampling time point was followed as mentioned below)

All animals from each test item treated dose group along with control animals were bled at each time points on Day 9 (2 on/5 off) and Day 13 (for once every three days dosing) at 0, 0.08, 0.5 1, 3, 7, and 24 hours post-dose from a saphenous vein for drug levels estimation.

Approximately 0.60-0.80 mL of blood samples were collected in 2% K2EDTA as an anticoagulant) and was centrifuged at 6000 rpm for 5 minutes at 5 °C temperature to separate the plasma. Plasma samples for drug levels estimation were stored at -80 °C until drug level estimation.

Bioanalysis and toxicokinetic analysis

The analysis was carried out using LC/MS-MS bioanalytical method (qualified fit for the purpose method). A brief description of the bioanalytical method is provided in the Bioanalysis and Toxicokinetic Report. The toxicokinetic parameters such as C_{max}/C_0, $AUC_{(0-t)}$, and T_{max} were determined by Non-Compartmental Analysis using Phoenix version 8.0 (Pharsight Corporation, USA).

Necropsy examination

A detailed necropsy examination was performed ~4-6 hour fasting for all the treated. The animals were weighed just before necropsy and this weight was used for relative organ weight calculations. Approximately 4 hours after the last dosing the blood was collected under isoflurane anesthesia for hematology in 2% K2 EDTA solution and in plain tubes to separate serum for biochemistry. After blood collection, the animals were necropsied, and a detailed necropsy examination was carried out by a Veterinary Pathologist. The necropsy included an examination of the external surface, external orifices, abdominal, thoracic, and cranial cavities, organs, and tissues. On completion of the gross pathology examination, the tissues were collected, weighed, and preserved from all animals.

Histopathological examination

In the daily dosing study, the following tissues were evaluated microscopically: liver, kidneys, heart, gastrocnemius muscle, stomach, testes, spleen, thymus, mesenteric lymph nodes, injection site, femur, bone marrow smear, and blood smear from all the groups.

In the intermittent dosing study, the following tissues were evaluated microscopically: liver, spleen, thymus, mesenteric lymph nodes, pancreas, femur, bone marrow smear, and blood smear from all the groups.

AU-15330 and enzalutamide formula for *in vivo* studies

AU-15330 was added in 40% of 2-Hydroxypropyl-beta-cyclodextrin (HPβCD) and sonicated until completely dissolved, solution and then further mixed with D5W (dextrose 5% in water) to reach a final concentration of 10% HPβCD. AU-15330 was freshly prepared right before administration to mice. AU-15330 was delivered to mice by intravenous (IP) injection either through the tail vein or retro-orbital injection unless otherwise indicated. Enzalutamide was added in 1% carboxymethyl cellulose (CMC) with 0.25% Tween-80 and sonicated until homogenized.

Human prostate tumor xenograft models

Six-week-old male CB17 severe combined immunodeficiency (SCID) mice were procured from the University of Michigan breeding colony. Subcutaneous tumors were established at both sides of the dorsal flank of mice. Tumors were measured at least biweekly using digital calipers following the formula (π/6) (L× W2), where L = length and W = width of the tumor. At the end of the studies, mice were sacrificed and tumors extracted and weighed. The University of Michigan Institutional Animal Care & Use Committee (IACUC) approved all in vivo studies.

For the VCaP non-castrated tumor model, 3×106 VCaP cells were injected subcutaneously into the dorsal flank on both sides of the mice in a serum-free medium with 50% Matrigel (BD Biosciences). Once tumors reached a palpable stage (~200 mm3), mice were randomized and treated with either 10, 30 mg/kg AU-15330, or vehicle by IP 5 days per week for 3 weeks.

For the VCaP castration-resistant tumor model, 3×106 VCaP cells were injected subcutaneously into the dorsal flank on both sides of the mice in a serum-free medium with 50% Matrigel (BD Biosciences). Once tumors reached a palpable stage (~200 mm3), tumor-bearing mice were castrated. Once tumors grew back to the pre-castration size, mice were randomized and treated with either 60 mg/kg AU-15330, or vehicle by IP injection 3 days per week, and with or without 10 mg/kg MDV3100 by oral gavage 5 days per week for 5 weeks.

Prostate patient-derived xenograft models

The University of Texas M.D. Anderson Cancer Center (MDACC) patient-derived xenografts (PDX) series has been previously described[61]. PDXs were derived from men with CRPC undergoing cystoprostatectomy using described protocols. MDA-PCa-146-12 was derived from a CRPC patient diagnosed with Gleason 5+4=9 prostate adenocarcinomas. MDA-PCa-146-12 was derived from a specimen obtained from the left bladder wall and demonstrated conventional adenocarcinoma (AR+). PDXs were maintained in male SCID mice by surgically implanting 2 mm3 tumors coated with 100% Matrigel to both flanks of mice. Once tumors reached ~200 mm3 in size, mice were randomized and divided into different treatment groups receiving either 60 mg/kg AU-15330 or vehicle by subcutaneous injection 3 days per week, and with or without 10 mg/kg MDV3100 by oral gavage 5 days per week for 3 weeks.

H929 xenograft generation, dosing, and measurement of antitumor activity

H929 cells (procured from ATCC) were cultured in RPMI1640 medium supplemented with 10% FBS, 1X penicillin-streptomycin, and 0.05 mM 2-mercaptoethanol. For generating xenografts, H929 cells were suspended in a 1:1 mixture of ECM gel (Sigma, E1270) and Hank’s Balanced Salt Solution (HBSS) and injected subcutaneously into the right flank of the female NOD-SCID mice (NOD/MrkBomTac-Prkdcscid) at a density of 5 x 106 cells/animal. When the mean tumor volume (TV) reached approximately 240-250 mm3, animals were selected and randomized based on TV into treatment groups of 5 animals each. Once-daily intravenous administration of vehicle and AU-15330 was started on Day 1 and continued for the duration of the study. Each treatment was administered at a dose volume of 5 mL/kg based on the most recent body weight. AU-15330 was formulated fresh every day using a vehicle containing 10% w/v 2-Hydroxypropyl-β-cyclodextrin (HPβCD) in 5% dextrose. The treatment-related changes in body weight (BW) and other clinical signs were monitored daily, and tumor measurements were recorded at least twice a week.

The TV in mm3 was calculated according to the following formula:

TV = length x width2 x 0.5

length: largest diameter of tumor (mm)

width: diameter perpendicular to length (mm)

Tumor growth inhibition (TGI %) was calculated as a measure of anti-tumor efficacy using the following formula

$$\% \text{ TGI} = 100 - ((\Delta T/\Delta C) \text{ X } 100)$$

where changes in tumor volume (Δ volumes) for each treated (T) and control (C) group was derived by subtracting the mean tumor volume on the first day of treatment (starting day) from the mean tumor volume on the specified observation day. All procedures relating to animal care, handling, and treatment were performed according to guidelines approved by the Institutional Animal Ethics Committee (IAEC) of Aurigene Discovery Technologies Ltd. All treatments were well tolerated with no clinical signs observed in all studies presented.

H&E and Immunohistochemistry (IHC), and Alcian blue staining

Immunohistochemistry (IHC):

IHC was performed on formalin-fixed paraffin-embedded 4um sections of mouse or xenograft tissues. Slides with tissue sections were incubated at 58C overnight and the next day were deparaffinized in xylene, followed by serial hydration steps in ethanol (100%, 70%) and water for 5 mins each. Endogenous tissue peroxidase activity was blocked by placing slides in 3% H_20_2-methanol solution for 1 hour at room temperature. Antigen retrieval was performed by microwaving slides in a solution of Citrate Buffer (pH=6) for 15 minutes, followed by blocking in 2.5% normal horse serum (Vector Laboratories, Cat#S-2012-50) for 2 hours. The slides were then incubated in the following primary antibodies overnight at 4 degrees: BRG1(Abcam Cat#108318, 1:100), AR(Millipore Cat#06-680, 1:2000), BRM1(Millipore sigma Cat#HPA029981, 1:100), FOXA1 (Thermo Fisher Scientific Cat#PA5-27157, 1:1000), ERG (Cell Signaling Technology Cat#97249S, 1:500). ImmPRESS-HRP conjugated anti-mouse/rabbit cocktail from Vector Laboratories (Cat#MP-7500-50) was used as secondary antibodies (room temperature, 1 hour) Visualization of staining was done as per the manufacturer's protocol (Vector Laboratories, Cat#SK-4100). Following DAB staining, slides were dehydrated in ethanol, xylene (5 minutes each) and mounted using EcoMount (Thermo Fisher, Cat#EM897L)

Alcian Blue staining:

Alcian blue staining was performed as per the manufacturer's protocol (Alcian Blue Stain Kit (pH 2.5) Cat#ab150662). Following an overnight incubation of tissue sections at 58 degrees celsius, slides were deparaffinized in Xylene followed by hydration in ethanol (100,70%) and water for 5 minutes each. Slides were then incubated in Acetic acid solution for 3 minutes followed by a 30 min incubation at room temperature in Alcian Blue stain (pH 2.5). Excess Alcian blue was removed by rinsing slides in Acetic acid solution for 1 minute, and three water washes for 2 minutes each. Nuclear Fast Red solution was used as a counterstain for 5 minutes. Slides were subsequently washed in running tap water, dehydrated in ethanol, xylene, and mounted using EcoMount (Thermo Fisher, Cat#EM897L)

Histopathological analysis of the various organs harvested for drug toxicity:

For the present study, organs- liver, spleen, kidney, colon, small intestine, prostate, and testis were harvested and fixed in 10% neutral buffered formalin followed by embedding in paraffin to make tissue blocks. These blocks were sectioned at 4 microns and stained with Harris hematoxylin and alcoholic eosin-Y stain (both reagents from Leica Surgipath) and staining performed on Leica autostainer-XL (automatic) platform. The stained sections were evaluated by two different pathologists using a brightfield microscope in a blinded fashion between the control and treatment groups for general tissue morphology and coherence of architecture. A detailed comprehensive analysis of the changes noted at the cellular and subcellular level were performed using scoring systems as described below for colon and testicular assessment. An evaluation summary for specific tissues is described below:

Evaluation of liver: liver tissue sections were evaluated for normal architecture and regional analysis for all three zones was performed for inflammation, necrosis, and fibrosis.

Evaluation of spleen: splenic tissue sections were evaluated for the organization of hematogenous red and lymphoid white pulp regions including necrosis and fibrotic changes if any.

Evaluation of kidney: kidney tissue sections were examined for changes noted if any in all the four renal functional components namely; glomeruli, interstitium, tubules, and vessels.

Evaluation of colon: colonic tissue sections were examined for mucosal (epithelium and lamina propria), submucosal, and seromuscular layer changes including crypt changes, goblet cells, inflammatory infiltrate granulation tissue, and mucosal ulceration. A detailed goblet cell evaluation was also performed utilizing alcian-blue staining wherein goblet cells and epithelial cells were counted in ten colonic crypt epithelium in each experimental animal of the various sub-groups. Summation of all the goblet and epithelial cells was done and a ratio of goblet cell to epithelial cell (GC ratio) was calculated per sample.

Evaluation of small intestine: small intestine tissue sections were examined for mucosal changes such as villous blunting, villous: crypt ratio and evaluated for inflammatory changes including intraepithelial lymphocytes, extent (mucosal, submucosal, serosal) and type of inflammatory infiltrate including tissue modulatory effect due to the same if any.

Evaluation of prostate: prostate tissue sections were evaluated to note for any epithelial abnormality and stromal changes identified in all four lobes (dorsal, anterior, lateral, and ventral). Additionally, any overt inflammatory infiltrate was also examined.

Evaluation of testis: testicular tissues were examined for the architectural assessment of seminiferous tubules (orderly maturation of germinal epithelial cells devoid of maturation arrest and Sertoli cell prominence if any), Leydig cells, and interstitial reaction. For an in-depth comprehensive analysis to comment upon the spermatogenesis a semi-quantitative "testicular biopsy score count (Johnsen score)" in 100 orderly cross-sections of seminiferous tubules in each animal of all the sub-groups at 20X magnification was performed. Each of the 100 seminiferous tubules assessed was given a score (score range:0-10). In the end, the average score was calculated (Total sum of score/100).

TMT Mass Spectrometry

VCaP cells were seeded at 5 × 106 cells on a 100 mm plate 24 h before treatment. Cells were treated in triplicate by the addition of test compounds. After 4 h, the cells were harvested and processed by using EasyPep™ Mini MS Sample Prep Kit (Thermo Fisher, #A40006). Samples were quantified using a micro BCA protein assay kit (Thermo Fisher Scientific) and cell lysis samples were proteolyzed and labeled with TMT 10-plex Isobaric Label Reagent (Thermo Fisher Scientific, #90110) essentially following the manufacturer's protocol. Briefly, upon reduction and alkylation of cysteines, the proteins were precipitated by adding 6 volumes of ice-cold acetone followed by overnight incubation at 20C. The precipitate was spun down, and the pellet was allowed to air dry. The pellet was resuspended in 0.1M TEAB and digested overnight with trypsin (1:50; enzyme:protein) at 37C with constant mixing using a thermomixer. The TMT 10-plex reagents were dissolved in 41 ml of anhydrous acetonitrile and labeling was performed by transferring the entire digest to the TMT reagent vial and incubating it at room temperature for 1 hour. The reaction was quenched by adding 8 ml of 5% hydroxylamine and a further 15 min incubation. Labeled samples were mixed together, and dried using a vacufuge. An offline fractionation of the combined sample (200 mg) into 10 fractions was performed using high pH reversed-phase peptide fractionation kit according to the manufacturer's protocol (Pierce; Cat #84868). Fractions were dried and reconstituted in 12 ml of 0.1% formic acid/2% acetonitrile in preparation for LC-MS/MS analysis.

In order to obtain superior accuracy in quantitation, we employed multinotch-MS3 [62] which minimizes the reporter ion ratio distortion resulting from fragmentation of co-isolated peptides during MS analysis. Orbitrap Fusion (Thermo Fisher Scientific) and RSLC Ultimate 3000 nano-UPLC (Dionex) was used to acquire the data. The sample (2 ml) was resolved on a PepMap RSLC C18 column (75 mm i.d. x 50 cm; Thermo Scientific) at the flow-rate of 300 nl/min using 0.1% formic acid/acetonitrile gradient system (2-22% acetonitrile in 150 min; 22-32% acetonitrile in 40 min; 20 min wash at 90% followed by 50 min re-equilibration) and direct spray into the mass spectrometer using EasySpray source (Thermo Fisher Scientific). The mass spectrometer was set to collect one MS1 scan (Orbitrap; 60K resolution; AGC target 2x105; max IT 100 ms) followed by data-dependent, ''Top Speed'' (3 seconds) MS2 scans (collision-induced dissociation; ion trap; NCD 35; AGC 5x103; max IT 100 ms). For multinotch-MS3, top 10 precursors from each MS2 were fragmented by HCD followed by Orbitrap analysis (NCE 55; 60K resolution; AGC 5x104; max IT 120 ms, 100-500 m/z scan range). Proteome Discoverer (v2.1; Thermo Fisher) was used for data analysis. MS2 spectra were searched against SwissProt human protein database (release 2015-11-11; 42084 sequences) using the following search parameters: MS1 and MS2 tolerances were set to 10 ppm and 0.6 Da, respectively; carbamidomethylation of cysteines (57.02146 Da) and TMT labeling of lysine and N-termini of peptides (229.16293 Da) were considered static modifications; oxidation of methionine (15.9949 Da) and deamidation of asparagine and glutamine (0.98401 Da) were considered variable. Identified proteins and peptides were filtered to retain only those that passed %1% FDR threshold. Quantitation was performed using high-quality MS3 spectra (Average signal-to-noise ratio of 20 and <30% isolation interference).

Meta-analyses of protein interactomes: Interactome proteomics data of AR and ERG was downloaded from published literature (ERG from [30] and AR from [31]), the FOXA1 nuclear co-immunoprecipitation/mass spectrometry experiment was performed in this study as described above. The protein interactomes of AR, ERG, and FOXA1 were ranked based on FDR at the top 10% and the intersection was taken from these three independent studies.

Assessment of drug synergism

To determine the presence of synergy between two drug treatments, cells were treated with increasing concentrations of either drug for 120 h, followed by the determination of viable cells using the CellTiter-Glo Luminescent Cell Viability Assay (Promega). The experiment was carried out in four biological replicates. The data were expressed as percentage inhibition relative to baseline, and the presence of synergy was determined by the Bliss method using the synergy finder R package [39].

Chapter acknowledgements:

Contents of this chapter have been summarized in research manuscript that is currently under review at a reputable scientific journal. The author list and their specific contributions are listed below:

Author list:

Lanbo Xiao*, Abhijit Parolia*, Yuanyuan Qiao*, Pushpinder Bawa, Rahul Mannan, Sandra E. Carson, Sanjana Eyunni, Yu Chang, Xiaoju Wang, Yuping Zhang, Josh N. Vo, Steven Kregel, Stephanie A. Simko, Andrew D. Delekta, Ingrid J. Apel, Lisa McMurry, Mustapha Jaber, Fengyun Su, Rui Wang, Sylvia Zelenka-Wang, Sanjita Sasmal, Leena K. Satyam, Subhendu Mukerjee, Chandrasekhar Abbineni, Kiran Aithal, Mital S. Bahkta, Jay Ghurye, Xuhong Cao, Nora M. Navone, Alexey I. Nesvizhskii, Rohit Mehra, Ulka Vaishampayan, Marco Blanchette, Yuzhuo Wang, Susanta Samajdar, Murali Ramachandra, Arul M. Chinnaiyan.

* These authors equally contributed to this work.

Author contributions:

L.X., A.P., Y.Q., and A.M.C. conceived and designed the studies; L.X. and A.P. performed all of the in vitro and functional genomics experiments with assistance from S.E.C., S.E., X.W., S.K., I.J.A, and M.J.; Y.Q. performed all of the animal efficacy studies with assistance from S.A.S. and A.D.D.; A.P. and P.B. carried out all of the bioinformatics analyses with assistance from Y.Z. and J.N.V.; R. Mannan and R. Mehra carried out all of the histopathological evaluations of drug toxicity as well as quantified all of the histology-based data; S.E. and S.Z. carried out all of the immunohistochemistry with L.M. helping with tissue processing and cross-sectioning. M.S.B., J.G., and M.B. helped with the HiChIP-seq experiment and data analyses. Y.C. helped with modeling drug-protein interaction. F.S. and R.W. generated next-generation sequencing libraries, and X.C. performed the sequencing. S. Sasmal. L.K.S., S.M., C.A., S. Samajdar, K.A., and M.R. are from Aurigene involved in the discovery, synthesis and initial profiling of AU-15330 compound. N.M.N., U.V., and Y.W. provided various key preclinical and clinical resources. A.I.N. guided all of the proteomics analyses. L.X., A.P., and A.M.C. wrote the manuscript and organized the final figures.

References:

1. Kornberg RD. Chromatin structure: a repeating unit of histones and DNA. Science. 1974;184: 868–871.
2. Bednar J, Horowitz RA, Grigoryev SA, Carruthers LM, Hansen JC, Koster AJ, et al. Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. Proc Natl Acad Sci U S A. 1998;95: 14173–14178.
3. Adelman K, Lis JT. How does Pol II overcome the nucleosome barrier? Molecular cell. 2002. pp. 451–452.
4. Li B, Carey M, Workman JL. The role of chromatin during transcription. Cell. 2007;128: 707–719.
5. Maston GA, Evans SK, Green MR. Transcriptional regulatory elements in the human genome. Annu Rev Genomics Hum Genet. 2006;7: 29–59.
6. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489: 57–74.
7. Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, et al. An atlas of active enhancers across human cell types and tissues. Nature. 2014;507: 455–461.
8. Blackwood EM, Kadonaga JT. Going the distance: a current view of enhancer action. Science. 1998;281: 60–63.
9. Visel A, Rubin EM, Pennacchio LA. Genomic views of distant-acting enhancers. Nature. 2009;461: 199–205.
10. Sur I, Taipale J. The role of enhancers in cancer. Nat Rev Cancer. 2016;16: 483–493.
11. Bradner JE, Hnisz D, Young RA. Transcriptional Addiction in Cancer. Cell. 2017;168: 629–643.
12. Chen H, Li C, Peng X, Zhou Z, Weinstein JN, Cancer Genome Atlas Research Network, et al. A Pan-Cancer Analysis of Enhancer Expression in Nearly 9000 Patient Samples. Cell. 2018;173: 386–399.e12.
13. Corces MR, Granja JM, Shams S, Louie BH, Seoane JA, Zhou W, et al. The chromatin accessibility landscape of primary human cancers. Science. 2018;362. doi:10.1126/science.aav1898
14. Dobersch S, Rubio K, Barreto G. Pioneer Factors and Architectural Proteins Mediating Embryonic Expression Signatures in Cancer. Trends Mol Med. 2019;25: 287–302.
15. Northcott PA, Lee C, Zichner T, Stütz AM, Erkek S, Kawauchi D, et al. Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. Nature. 2014;511: 428–434.
16. Parolia A, Cieslik M, Chu S-C, Xiao L, Ouchi T, Zhang Y, et al. Distinct structural classes of activating FOXA1 alterations in advanced prostate cancer. Nature. 2019;571: 413–418.
17. Flavahan WA, Drier Y, Liau BB, Gillespie SM, Venteicher AS, Stemmer-Rachamimov AO, et al. Insulator dysfunction and oncogene activation in IDH mutant gliomas. Nature. 2016;529: 110–114.
18. Kadoch C, Hargreaves DC, Hodges C, Elias L, Ho L, Ranish J, et al. Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. Nat Genet. 2013;45: 592–601.
19. Shain AH, Pollack JR. The spectrum of SWI/SNF mutations, ubiquitous in human cancers. PLoS One. 2013;8: e55119.
20. Kassabov SR, Zhang B, Persinger J, Bartholomew B. SWI/SNF unwraps, slides, and rewraps

the nucleosome. Mol Cell. 2003;11: 391–403.

21. Helming KC, Wang X, Roberts CWM. Vulnerabilities of mutant SWI/SNF complexes in cancer. Cancer Cell. 2014;26: 309–317.
22. Kadoch C, Crabtree GR. Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics. Sci Adv. 2015;1: e1500447.
23. Centore RC, Sandoval GJ, Soares LMM, Kadoch C, Chan HM. Mammalian SWI/SNF Chromatin Remodeling Complexes: Emerging Mechanisms and Therapeutic Strategies. Trends Genet. 2020;36: 936–950.
24. Mashtalir N, D'Avino AR, Michel BC, Luo J, Pan J, Otto JE, et al. Modular Organization and Assembly of SWI/SNF Family Chromatin Remodeling Complexes. Cell. 2018;175: 1272–1288.e20.
25. Robinson D, Van Allen EM, Wu Y-M, Schultz N, Lonigro RJ, Mosquera J-M, et al. Integrative clinical genomics of advanced prostate cancer. Cell. 2015;161: 1215–1228.
26. Bolli N, Biancon G, Moarii M, Gimondi S, Li Y, de Philippis C, et al. Analysis of the genomic landscape of multiple myeloma highlights novel prognostic markers and disease subgroups. Leukemia. 2018;32: 2604–2616.
27. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods. 2013;10: 1213–1218.
28. Zhou B, Hu J, Xu F, Chen Z, Bai L, Fernandez-Salas E, et al. Discovery of a Small-Molecule Degrader of Bromodomain and Extra-Terminal (BET) Proteins with Picomolar Cellular Potencies and Capable of Achieving Tumor Regression. J Med Chem. 2018;61: 462–481.
29. Wang Z, Civelek M, Miller CL, Sheffield NC, Guertin MJ, Zang C. BART: a transcription factor prediction tool with query gene sets or epigenomic profiles. Bioinformatics. 2018 [cited 8 May 2018]. doi:10.1093/bioinformatics/bty194
30. Sandoval GJ, Pulice JL, Pakula H, Schenone M, Takeda DY, Pop M, et al. Binding of TMPRSS2-ERG to BAF Chromatin Remodeling Complexes Mediates Prostate Oncogenesis. Mol Cell. 2018;71: 554–566.e7.
31. Stelloo S, Nevedomskaya E, Kim Y, Hoekman L, Bleijerveld OB, Mirza T, et al. Endogenous androgen receptor proteomic profiling reveals genomic subcomplex involved in prostate tumorigenesis. Oncogene. 2018;37: 313–322.
32. Remillard D, Buckley DL, Paulk J, Brien GL, Sonnett M, Seo H-S, et al. Degradation of the BAF Complex Factor BRD9 by Heterobifunctional Ligands. Angew Chem Int Ed Engl. 2017;56: 5738–5743.
33. Zoppi V, Hughes SJ, Maniaci C, Testa A, Gmaschitz T, Wieshofer C, et al. Iterative Design and Optimization of Initially Inactive Proteolysis Targeting Chimeras (PROTACs) Identify VZ185 as a Potent, Fast, and Selective von Hippel-Lindau (VHL) Based Dual Degrader Probe of BRD9 and BRD7. J Med Chem. 2019;62: 699–726.
34. Viswanathan SR, Ha G, Hoff AM, Wala JA, Carrot-Zhang J, Whelan CW, et al. Structural Alterations Driving Castration-Resistant Prostate Cancer Revealed by Linked-Read Genome Sequencing. Cell. 2018;174: 433–447.e19.
35. Takeda DY, Spisák S, Seo J-H, Bell C, O'Connor E, Korthauer K, et al. A Somatically Acquired Enhancer of the Androgen Receptor Is a Noncoding Driver in Advanced Prostate Cancer. Cell. 2018;174: 422–432.e13.
36. Pott S, Lieb JD. What are super-enhancers? Nat Genet. 2015;47: 8–12.
37. Asangani IA, Wilder-Romans K, Dommeti VL, Krishnamurthy PM, Apel IJ, Escara-Wilke J,

et al. BET Bromodomain Inhibitors Enhance Efficacy and Disrupt Resistance to AR Antagonists in the Treatment of Prostate Cancer. Mol Cancer Res. 2016;14: 324–331.

38. Faivre EJ, McDaniel KF, Albert DH, Mantena SR, Plotnik JP, Wilcox D, et al. Selective inhibition of the BD2 bromodomain of BET proteins in prostate cancer. Nature. 2020;578: 306–310.
39. Ianevski A, He L, Aittokallio T, Tang J. SynergyFinder: a web application for analyzing drug combination dose-response matrix data. Bioinformatics. 2017;33: 2413–2415.
40. Kregel S, Chen JL, Tom W, Krishnan V, Kach J, Brechka H, et al. Acquired resistance to the second-generation androgen receptor antagonist enzalutamide in castration-resistant prostate cancer. Oncotarget. 2016;7: 26259–26274.
41. Gonzalez-Perez A, Jene-Sanz A, Lopez-Bigas N. The mutational landscape of chromatin regulatory factors across 4,623 tumor samples. Genome Biol. 2013;14: r106.
42. Hodges C, Kirkland JG, Crabtree GR. The Many Roles of BAF (mSWI/SNF) and PBAF Complexes in Cancer. Cold Spring Harb Perspect Med. 2016;6. doi:10.1101/cshperspect.a026930
43. Schick S, Rendeiro AF, Runggatscher K, Ringler A, Boidol B, Hinkel M, et al. Systematic characterization of BAF mutations provides insights into intracomplex synthetic lethalities in human cancers. Nat Genet. 2019;51: 1399–1410.
44. Xue Y, Meehan B, Fu Z, Wang XQD, Fiset PO, Rieker R, et al. SMARCA4 loss is synthetic lethal with CDK4/6 inhibition in non-small cell lung cancer. Nat Commun. 2019;10: 557.
45. Mills IG. Maintaining and reprogramming genomic androgen receptor activity in prostate cancer. Nat Rev Cancer. 2014;14: 187–198.
46. Watson PA, Arora VK, Sawyers CL. Emerging mechanisms of resistance to androgen receptor inhibitors in prostate cancer. Nat Rev Cancer. 2015;15: 701–711.
47. Cyrta J, Augspach A, de Filippo MR, Prandi D, Thienger P, Benelli M, et al. Role of Specialized Composition of SWI/SNF Complexes in Prostate Cancer Lineage Plasticity. 2020. p. 2020.03.06.949131. doi:10.1101/2020.03.06.949131
48. Marshall TW, Link KA, Petre-Draviam CE, Knudsen KE. Differential requirement of SWI/SNF for androgen receptor activity. J Biol Chem. 2003;278: 30605–30613.
49. Link KA, Burd CJ, Williams E, Marshall T, Rosson G, Henry E, et al. BAF57 governs androgen receptor action and androgen-dependent proliferation through SWI/SNF. Mol Cell Biol. 2005;25: 2200–2215.
50. Link KA, Balasubramaniam S, Sharma A, Comstock CES, Godoy-Tundidor S, Powers N, et al. Targeting the BAF57 SWI/SNF subunit in prostate cancer: a novel platform to control androgen receptor activity. Cancer Res. 2008;68: 4551–4558.
51. Sanjana NE, Shalem O, Zhang F. Improved vectors and genome-wide libraries for CRISPR screening. Nat Methods. 2014;11: 783–784.
52. Kim D, Salzberg SL. TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. Genome Biol. 2011;12: R72.
53. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31: 166–169.
54. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26: 139–140.
55. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics. 2016;32: 2847–2849.
56. Korotkevich G, Sukhov V, Sergushichev A. Fast gene set enrichment analysis. Cold Spring

Harbor Laboratory. 2019. p. 060012. doi:10.1101/060012

57. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25: 1754–1760.
58. Abdennur N, Mirny LA. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. Bioinformatics. 2020;36: 311–316.
59. Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, et al. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. cels. 2016;3: 95–98.
60. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010;38: 576–589.
61. Palanisamy N, Yang J, Shepherd PDA, Li-Ning-Tapia EM, Labanca E, Manyam GC, et al. The MD Anderson Prostate Cancer Patient-derived Xenograft Series (MDA PCa PDX) Captures the Molecular Landscape of Prostate Cancer and Facilitates Marker-driven Therapy Development. Clin Cancer Res. 2020;26: 4933–4946.
62. McAlister GC, Nusinow DP, Jedrychowski MP, Wühr M, Huttlin EL, Erickson BK, et al. MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes. Anal Chem. 2014;86: 7150–7158.

Chapter 4

Novel Overarching Concepts, Conclusions, and Future Directions

AR is a nodal "*lineage co-oncogene*" in prostate adenocarcinoma

Hormone receptor-driven cancers best exemplify the concept of lineage-addiction. In the current clinical workflow of prostate cancer treatment, following surgical resection of the tumor, all targeted therapies inhibit the androgen/AR signaling axis. However, now that we have gone on to sequence thousands of treatment-naive primary prostate cancer genomes, I find it extremely intriguing that these tumors harbor no recurrent alterations in the AR gene locus or its coding sequence [1]. The same holds true for *ESR1* mutations in primary breast adenocarcinomas. These findings imply that hormone-receptors like AR and estrogen receptor-alpha (ER) are not conventional oncogenes in primary prostate or breast cancers, respectively, and highlight the need to understand their cancer-essential functions within a new conceptual framework.

A pertinent question then becomes: what explains the profound therapeutic benefit derived from AR inhibition in prostate cancer patients? One likely explanation could be that driver oncogenes make requisite interactions with the AR-signaling complex and repurpose it to activate tumorigenic and hyper-proliferative gene programs. This lays the foundation of the herein proposed "lineage-addiction" theory in cancer cells (**Figure 1-1**). The lineage-addiction model implies that cancers that originate from the differentiated cell lineages (e.g., the luminal prostate/breast epithelia) retain and aberrantly rewire the lineage-specific core-transcriptional machinery to activate survival and/or hyper-proliferative gene programs. To better define such functional roles in tumorigenesis, we propose *AR*, and other similar lineage-survival genes (like *ESR1*) to be termed as "*lineage co-oncogenes.*" In other tumors, these genetically unaltered genes would likely comprise essential epigenetic or transcriptional determinants of cell lineage identity, which get hijacked by oncogenes to drive the malignant phenotype. It is also important to note that most likely hyper-activation of lineage co-oncogenes by themselves cannot drive transformation, but their activity would be absolutely essential for oncogene-driven transformation.

In breast and prostate glands, ER and AR, respectively, drive terminal differentiation of the luminal epithelial cells (the putative cell-of-origin) [2,3]. Genetic knockout studies in mice have revealed an essential role for AR signaling in normal development and glandular morphogenesis of the prostate gland [3,4]. A recent study using single-cell RNA-seq presented compelling evidence for the essentiality of AR activity in the regeneration of luminal prostate epithelial cells after acute depletion of androgen [5]. Thus, AR is considered as a master instructor of the prostatic luminal epithelial identity. Notably, cancers that originate from these differentiated cells seem to continue to rely on the activity of lineage-specifying genes for survival and proliferation [2,3,6,7]. This concept is similarly embodied by the acute dependence of melanoma cells on lineage-defining microphthalmia-associated transcription factor (MITF)[8]. Thus, lineage-specifying genes could comprise a novel class of cancer dependencies that can be therapeutically targeted.

It is equally important to note the limitations of the cell lineage model of human carcinogenesis described above. Most likely the lineage-addiction model pertains only to cancers that originate from a lineage-committed or differentiated normal cell. Thus, one can envision most adenocarcinomas originating from epithelial cells (or other specialized cell types) to retain addiction to their inherent lineage programs, but sarcomas originating from undifferentiated cells (e.g. Ewing sarcomas) may not harbor such lineage-dependencies.

Coercive evolution of the prostate cancer genome under the constraint of therapies

AR alterations are only detected in CRPC tumors and are a direct consequence of treating the primary disease with ADTs. However, it is important to note that even in the presence of activating AR mutations or copy-amplifications, AR signaling is never hyper-activated at any stage of prostate cancer progression. The activating AR alterations only restore AR activity closer to baseline when the levels of androgen sharply fall following castration. In fact, hyper-stimulation of AR activity has anti-proliferative effects in prostate cancer cells [9], which has even prompted an evaluation of high dose testosterone administration as a potential therapy in advanced patients [10]. Thus, prostate cancer cells tolerate a very narrow range of AR activity that likely allows for its oncogenic appropriation while keeping its pro-differentiation duties in check.

Concordantly, a recent genomic study of *de novo* metastatic CRPCs found no AR alterations in patients who never received ADT [11], strongly suggesting that these alterations are functionally inconsequential in a naturally evolving disease. Thus, alterations in the *AR* gene in over 70-80% of metastatic CRPC cases only emerge when the cancer cells are challenged with androgen deprivation and serve as a testimony to the acute addiction of tumor cells to the AR-instructed prostatic lineage. It is yet to be determined if mutations in other AR cofactors that are particularly enriched in CRPC (e.g. *NCOR2* inactivating or FOXA1 class2 activating mutations) are associated with ADT or not. Despite having such high recurrence in CRPC, vestiges of FOXA1 alterations are rarely detected in neuroendocrine prostate cancer genomes [12–14], which is similar to rare detection of AR amplifications in this disease subtype [12,15]. Given our mechanistic findings, it is likely that class1 mutants of FOXA1—by safeguarding and sustaining AR activity—“lock” prostate adenocarcinoma in an AR-driven state. This is supported with our data showing class1 FOXA1 mutants to rescue AR activity under castrate-androgen levels as well as treatment with AR antagonists [13]. Here, we propose the above mentioned genomic events (and their kind) in advanced prostate cancer to be collectively referred to as “lineage reinforcing” alterations that causally drive the acquisition of resistance to lineage-targeted therapies. Interestingly, the majority of these adaptive mechanisms are acquired through genetic alterations in cancer cells.

Persistent inhibition of the androgen/AR axis in prostate cancer cells eventually results in a complete suspension of the AR-dependent lineage identity and a transfer to a neuroendocrine lineage instead. It is fascinating that this “lineage switching” or “lineage escaping” event is accompanied by the acquisition of an entirely new set of oncogenic drivers (namely, SOX2, MYCN, etc.) and abandonment of previous oncogenic drivers (e.g. FOXA1, ERG, etc.). This strongly suggests that it is the lineage identity of the cell that dominantly dictates which genes are competent for driving the malignant phenotype. In other words, oncogenes have to restrictively function within the outer confines of the cellular lineage. Thus, molecularly understanding these lineage–oncogene associations will likely hold the clues to why somatic alterations in human cancer show lineage-specificity. These research curiosities will guide my future pursuits in the field of molecular cancer biology.

The view of cancer genomes as natural CRISPR screens

Cancer is a devastating disease that has abruptly cut short numerous innocent lives. Thus, one of the primary motivations for me to work in the field of cancer biology is to make a meaningful impact in the clinical management of still incurable cancers. But, there is also a pure scientific motivation to my fascination with this research domain: From a conceptual standpoint, the accumulation of driver alterations—i.e. alterations that undergo positive selection—in cancer genomes can be viewed from a lens of a natural CRISPR screen. In other words, the study of recurrent alterations in human cancers, their clonality and clinical presentation, their genomic position and type, the relative conservation of genomic regions they alter or disrupt, their recurrence at certain linearly or three-dimensionally clustered amino acid residues, and/or recurrence within functional protein domains can guide the development of very sophisticated functional and mechanistic hypotheses. Molecular experiments along these lines will not only elucidate the aberrant function of driver proteins in diseased states but will also uncover novel facets of their physiologic functions in normal cells.

Following from above, a higher DNA affinity for the C-terminal-truncated, class2 FOXA1 mutants ascribes a DNA-binding inhibitory function to the C-terminal regulatory domain of the wild-type FOXA1 protein. Notably, the C-terminal regulatory domain has also been implicated in mediating the pioneering duties of the FOXA1 protein [16,17]. Thus, it'll be very meaningful to profile the pioneering ability of the cistomically-dominant class2-mutants of FOXA1 and compare it to that of its wild-type counterpart. Similarly, the marked bias of structural variations within the highly syntenic *FOXA1* locus to encompass the *MIPOL1* gene is highly suggestive of putative enhancers of the *FOXA1* gene being coded in that *cis*-neighborhood. Thus, being amply equipped with granular insights from cancer genomes and modern multi-omic tools, we truly might be stepping into another golden era for molecular biology, where pathologic alterations in chromatin structure or protein composition can be used to uncover their fundamental physiologic functions.

Characterizing the oncogenic effects of FOXA1 mutants using transgenic mouse models

Despite the high-recurrence of FOXA1 alterations in prostate cancer, their driver roles have not been experimentally studied in mouse models. To date, no laboratory has developed genetically-engineered mouse models (GEMMs) of prostate cancer that are driven by transgenic overexpression of either wild-type FOXA1 or its mutant onco-variants [18]. I believe the development of these physiological models is imperative in molecularly unraveling class-specific disease pathobiology and uncovering therapeutic vulnerabilities that are associated with sub-classes of FOXA1-altered prostate cancers.

Since FOXA1 is required for normal glandular morphogenesis and development of the mouse prostate gland [19], we plan to employ an exogenous knock-in strategy. In the *ROSA26* gene locus of the C57BL/6J mice, we will knock in the LoxP-STOP-LoxP-Transgene cassette encoding the human wild-type or mutant *FOXA1* transgene. This will cover all three alterations classes with knock-in of the V5-tagged, 1) R261G (Class1-missense), 2) R265-71del (Class1-inframe indel), 3) P358fs (class2-frameshift), and 4) wild-type (mimics class3 duplications) FOXA1 variants. These transgenic mice will then be crossed with the *Probasin (Pb)-Cre* mice to achieve specific activation of the *FOXA1* oncogene in AR-positive luminal prostate epithelial cells. The *FOXA1* mutation and *Pb-Cre* positive experimental mice, along with litter-matched negative control animals (i.e. negative for either *FOXA1* mutation or *Pb-Cre*), will be aged for time-lapsed histopathological assessment of the prostate gland.

Learning from previous models, we will first euthanize mice (n=2 for each line) at 12wk of age to confirm prostate-specific expression of the transgene via immunohistochemical detection of the V5-tag. Then these mice will be aged to 52 weeks for primary assessment of the prostatic histology. With the help of an expert pathologist, we will carry out a multi-field inspection in at least 10 animals covering all four prostatic lobes to detect any morphological defects and score for cellular atypia, dysplasia, or neoplasia following the established criteria. Additionally, in class2 FOXA1-driven GEMMs, we would perform quantitative evaluation of metastatic dissemination of tumor cells in the blood, bone marrow aspirates, and selective common secondary sites using V5-tag fixed-cell flow cytometry. Subsequent time-points for follow-up histologic and phenotypic inspections will be determined based on primary results at 52 weeks. In case the monogenic models show no neoplastic transformation or present with an indolent, slow-growing tumor, we will cross these mice with either *TP53* or *PTEN* flox animals based on class-specific co-alteration profiles. Notably, a recent genomic study of primary prostate cancer from Chinese men revealed activating *FOXA1* alterations (mostly class 1 in type) to recur in over 40% of the cases [20]. Our transgenic mice will be the first models of FOXA1-driven prostate cancer in the field and will help elucidate the pathobiology of a major oncogenic driver of prostate carcinogenesis and/or progression.

Potential synergism between lineage and oncogene targeted therapies

The most promising facet of lineage-targeted therapies is their potential synergism with existing therapies that target driver oncogenes. For instance, in preclinical systems of advanced prostate cancers, we saw a remarkable synergism between enzalutamide (an AR antagonist) and our SWI/SNF degrader compound. While AR antagonists block the DHT-binding pocket in the AR protein, SWI/SNF degradation impedes the binding of AR to the chromatin at *cis*-regulatory elements. Thus, the combination of these drugs results in a complementary, multi-pronged attack on the AR transcriptional complex, thus producing a stronger and more durable remission of the disease. Furthermore, SWI/SNF inhibition would attenuate other the ancillary cancer-promoting pathways that are dependent on enhancer activity, such as MYC-signaling.

Notably, we found SWI/SNF inhibition to be therapeutically effective in enzalutamide-resistant as well as AR-independent models of prostate cancer—albeit 5-10 fold less than the AR-positive disease for the latter. These prostate cancer disease subtypes currently pose a major clinical challenge. A recent paper implicated a specialized SWI/SNF complex in mediating the neuroendocrine transdifferentiation of prostate cancer cells [21]. Based on this study, the SWI/SNF complexes get dynamically reconfigured during prostate cancer progression and, as per our findings, continue to play an oncogenic role in AR-dependent as well as AR-independent stages of the disease. In these disease subtypes, the SWI/SNF complexes likely work in concert with distinct transcriptional oncogenes. Moving forward, we plan to mechanistically interrogate the growth inhibitory effect of SWI/SNF inactivation in AR-independent prostate cancer. We also plan to test the utility SWI/SNF ATPase inactivation in *SS18-SSX* fusion driven synovial sarcomas and heme-malignancies that harbor well-characterized addiction to aberrant transcriptional states (e.g. dependence of acute myeloid leukemia on the MLL-fusion/HOXA9-activated gene programs [22,23]).

As with epigenetic therapies in general, it would be very important to demonstrate that there are no unintended toxicities to normal cellular functions from complete SWI/SNF inactivation. Such dose-limiting toxicities have largely hindered the clinical evaluations of the BRD4 class of therapeutics in human cancers, which also showed great promise in preclinical systems [24–26]. Encouragingly, in our preclinical efficacy studies in mice, despite Smarca2 and Smarca4 degradation in normal tissues, we see no major changes in whole-body weights or weights of essential organs upon prolonged treatment with the SWI/SNF degrader. This was accompanied with no histological signs of toxicities in essential organs, including no indications of thrombocytopenia (i.e. acute depletion of blood platelets), gastrointestinal goblet cell depletion, and normal prostatic and testicular atrophy—these being major toxicities previously reported for BRD4 inhibitors. However, this data needs to be further corroborated by extensive tolerability and toxicities studies of SWI/SNF degraders in other animal models, including higher primates.

The roots of lineage addiction likely lie in the hardwiring of the chromatin structure

For years, we have profiled the primary differences in cancer cells compared to their normal counterparts. I wonder if it would be equally informative if we focused on what aspects of the normal cellular heritage remain unaltered in the transformed state. Even upon acquisition of aggressive characteristics, cancer cells rarely lose the identity of the originating normal cell. For instance, prostate cancer cells still closely resemble their normal epithelial precursors in terms of molecular identity. This strongly suggests that the genes and pathways that define cell identity persist in tumor cells—albeit aberrantly—and are likely essential for cancer cell growth and survival.

I predict lineage addiction in cancers to primarily stem from the physical chromatin structure of the normal precursor cells. Thus, only genes that can functionally interact with the lineage-specific transcription factor machinery are competent to drive oncogenesis. In other words, accessible regions of the chromatin that are shared between the normal cell-of-origin and the cancer cell may hold clues as to what lineage-specific transcriptional machinery continue to play a relevant role in a given tumor. Intriguingly, pioneer factors primarily bind at *cis*-regulatory elements that include several kilobases long lineage-defining super-enhancers [27,28]. Thus, pioneer factors and their chromatin cofactors have emerged as primary gatekeepers of cellular identity. For example, in prostate and breast epithelia, FOXA1 is the chief pioneer factor that delimits the accessible chromatin landscape [29]. Thus, it is tempting to speculate that pioneer factors and their chromatin remodeling cofactors molecularly bridge the lineage–oncogene interactions in cancer cells. In fact, a recent multi-omic effort to profile the accessible regulome across human cancers identified many putative pioneers and lineage co-oncogenes that need to be mechanistically explored in future studies [30,31].

A comprehensive understanding of lineage–oncogene association will facilitate the identification of novel molecular dependencies in cancer cells that can be co-targeted along with driver oncogenes. These efforts can particularly benefit from the advent of single cell mutli-omics techniques that will allow for more meaningful comparisons of tumor cells to their cell-of-origin. Most molecular profiles of normal tissues have been generated using bulk approaches, which can be potentially confounded as normal tissues comprise many cell types (e.g. epithelial, stromal, endothelial, immune cells, etc.) in varied proportions. Although this cellular milieu plays well-documented roles in tumorigenesis, tumors themselves disproportionately comprise a single cell type (e.g. mutated epithelial cells in adenocarcinomas). Thus, comparison of molecular features between normal and cancer cells at a single-cell level would provide unprecedented insights into how the lineage-specific chromatin structure and gene program is altered during transformation.

In summary, targeting lineage-addiction in cancer cells is a novel concept and is likely to synergize with existing oncogene-targeted therapies. This doctoral thesis establishes FOXA1 as a principal oncogene in hormone receptor-driven malignancies and uncovers the SWI/SNF complex as a novel lineage-associated dependency in AR-addicted prostate cancers. The work presented in Chapter 1 outlines the first classification scheme of *FOXA1* alterations in human cancers, wherein each of the alteration classes is associated with divergent clinical incidence and oncogenic gain-of-function. These mechanistic insights will facilitate mechanism-informed translational studies focused on effective clinical management of these disparate disease subtypes. The work presented in Chapter 2 is the first proof of concept that impeding chromatin accessibility at non-coding *cis*-regulatory elements constitutes a novel therapeutic strategy in enhancer-binding transcription factor-addicted tumors, including AR/FOXA1-driven prostate and MYC-driven plasma cell tumors—warranting the clinical evaluation of SWI/SNF degraders in human trials.

Chapter acknowledgements:

This chapter was written by Abhijit Parolia and summarizes novel concepts and findings from the scientific work presented in this thesis.

References:

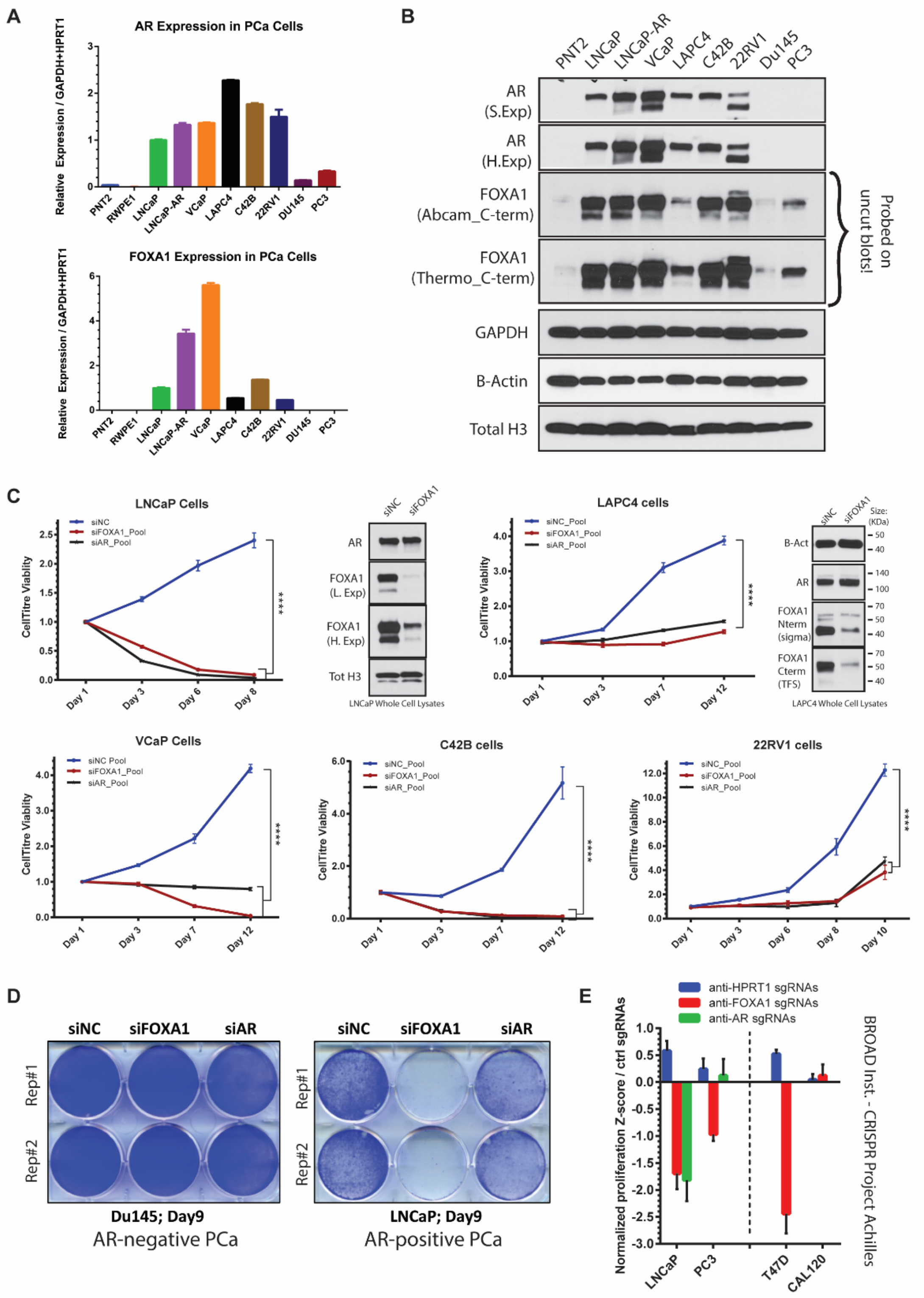
1. Cancer Genome Atlas Research Network. The Molecular Taxonomy of Primary Prostate Cancer. Cell. 2015;163: 1011–1025.
2. Mueller SO, Clark JA, Myers PH, Korach KS. Mammary gland development in adult mice requires epithelial and stromal estrogen receptor alpha. Endocrinology. 2002;143: 2357–2365.
3. Toivanen R, Shen MM. Prostate organogenesis: tissue induction, hormonal regulation and cell type specification. Development. 2017;144: 1382–1398.
4. Yadav N, Heemers HV. Androgen action in the prostate gland. Minerva Urol Nefrol. 2012;64: 35–49.
5. Karthaus WR, Hofree M, Choi D, Linton EL, Turkekul M, Bejnood A, et al. Regenerative potential of prostate luminal cells revealed by single-cell analysis. Science. 2020;368: 497–505.
6. Mills IG. Maintaining and reprogramming genomic androgen receptor activity in prostate cancer. Nat Rev Cancer. 2014;14: 187–198.
7. Higgins MJ, Baselga J. Targeted therapies for breast cancer. J Clin Invest. 2011;121: 3797–3803.
8. Levy C, Khaled M, Fisher DE. MITF: master regulator of melanocyte development and melanoma oncogene. Trends Mol Med. 2006;12: 406–414.
9. Tararova ND, Narizhneva N, Krivokrisenko V, Gudkov AV, Gurova KV. Prostate cancer cells tolerate a narrow range of androgen receptor expression and activity. Prostate. 2007;67: 1801–1815.
10. Schweizer MT, Antonarakis ES, Wang H, Ajiboye AS, Spitz A, Cao H, et al. Effect of bipolar androgen therapy for asymptomatic men with castration-resistant prostate cancer: results from a pilot clinical study. Sci Transl Med. 2015;7: 269ra2.
11. Stopsack KH, Nandakumar S, Wibmer AG, Haywood S, Weg ES, Barnett ES, et al. Oncogenic Genomic Alterations, Clinical Phenotypes, and Outcomes in Metastatic Castration-Sensitive Prostate Cancer. Clin Cancer Res. 2020;26: 3230–3238.

12. Beltran H, Prandi D, Mosquera JM, Benelli M, Puca L, Cyrta J, et al. Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. Nat Med. 2016;22: 298–305.
13. Parolia A, Cieslik M, Chu S-C, Xiao L, Ouchi T, Zhang Y, et al. Distinct structural classes of activating FOXA1 alterations in advanced prostate cancer. Nature. 2019;571: 413–418.
14. Adams EJ, Karthaus WR, Hoover E, Liu D, Gruet A, Zhang Z, et al. FOXA1 mutations alter pioneering activity, differentiation and prostate cancer phenotypes. Nature. 2019;571: 408–412.
15. Bluemn EG, Coleman IM, Lucas JM, Coleman RT, Hernandez-Lopez S, Tharakan R, et al. Androgen Receptor Pathway-Independent Prostate Cancer Is Sustained through FGF Signaling. Cancer Cell. 2017;32: 474–489.e6.
16. Zaret KS, Lerner J, Iwafuchi-Doi M. Chromatin Scanning by Dynamic Binding of Pioneer Factors. Mol Cell. 2016;62: 665–667.
17. Cirillo LA, Lin FR, Cuesta I, Friedman D, Jarnik M, Zaret KS. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. Mol Cell. 2002;9: 279–289.
18. Irshad S, Abate-Shen C. Modeling prostate cancer in mice: something old, something new, something premalignant, something metastatic. Cancer Metastasis Rev. 2013;32: 109–122.
19. Gao N, Zhang J, Rao MA, Case TC, Mirosevich J, Wang Y, et al. The role of hepatocyte nuclear factor-3 alpha (Forkhead Box A1) and androgen receptor in transcriptional regulation of prostatic genes. Mol Endocrinol. 2003;17: 1484–1507.
20. Li J, Xu C, Lee HJ, Ren S, Zi X, Zhang Z, et al. A genomic and epigenomic atlas of prostate cancer in Asian populations. Nature. 2020;580: 93–99.
21. Cyrta J, Augspach A, de Filippo MR, Prandi D, Thienger P, Benelli M, et al. Role of Specialized Composition of SWI/SNF Complexes in Prostate Cancer Lineage Plasticity. 2020. p. 2020.03.06.949131. doi:10.1101/2020.03.06.949131
22. Sun Y, Zhou B, Mao F, Xu J, Miao H, Zou Z, et al. HOXA9 Reprograms the Enhancer Landscape to Promote Leukemogenesis. Cancer Cell. 2018;34: 643–658.e5.
23. Lambert M, Alioui M, Jambon S, Depauw S, Van Seuningen I, David-Cordonnier M-H. Direct and Indirect Targeting of HOXA9 Transcription Factor in Acute Myeloid Leukemia. Cancers. 2019;11. doi:10.3390/cancers11060837
24. Faivre EJ, McDaniel KF, Albert DH, Mantena SR, Plotnik JP, Wilcox D, et al. Selective inhibition of the BD2 bromodomain of BET proteins in prostate cancer. Nature. 2020;578: 306–310.
25. Paiva S-L. Widening the window of bromodomain inhibition. Nature reviews. Drug discovery. 2020. p. 166.
26. Bolden JE, Tasdemir N, Dow LE, van Es JH, Wilkinson JE, Zhao Z, et al. Inducible in vivo silencing of Brd4 identifies potential toxicities of sustained BET protein inhibition. Cell Rep. 2014;8: 1919–1929.
27. Iwafuchi-Doi M, Donahue G, Kakumanu A, Watts JA, Mahony S, Pugh BF, et al. The Pioneer Transcription Factor FoxA Maintains an Accessible Nucleosome Configuration at Enhancers for Tissue-Specific Gene Activation. Mol Cell. 2016;62: 79–91.
28. Factor DC, Corradin O, Zentner GE, Saiakhova A, Song L, Chenoweth JG, et al. Epigenomic comparison reveals activation of "seed" enhancers during transition from naive to primed pluripotency. Cell Stem Cell. 2014;14: 854–863.

29. Lupien M, Eeckhoute J, Meyer CA, Wang Q, Zhang Y, Li W, et al. FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. Cell. 2008;132: 958–970.
30. Corces MR, Granja JM, Shams S, Louie BH, Seoane JA, Zhou W, et al. The chromatin accessibility landscape of primary human cancers. Science. 2018;362. doi:10.1126/science.aav1898
31. Klemm SL, Shipony Z, Greenleaf WJ. Chromatin accessibility and the regulatory epigenome. Nat Rev Genet. 2019;20: 207–220.

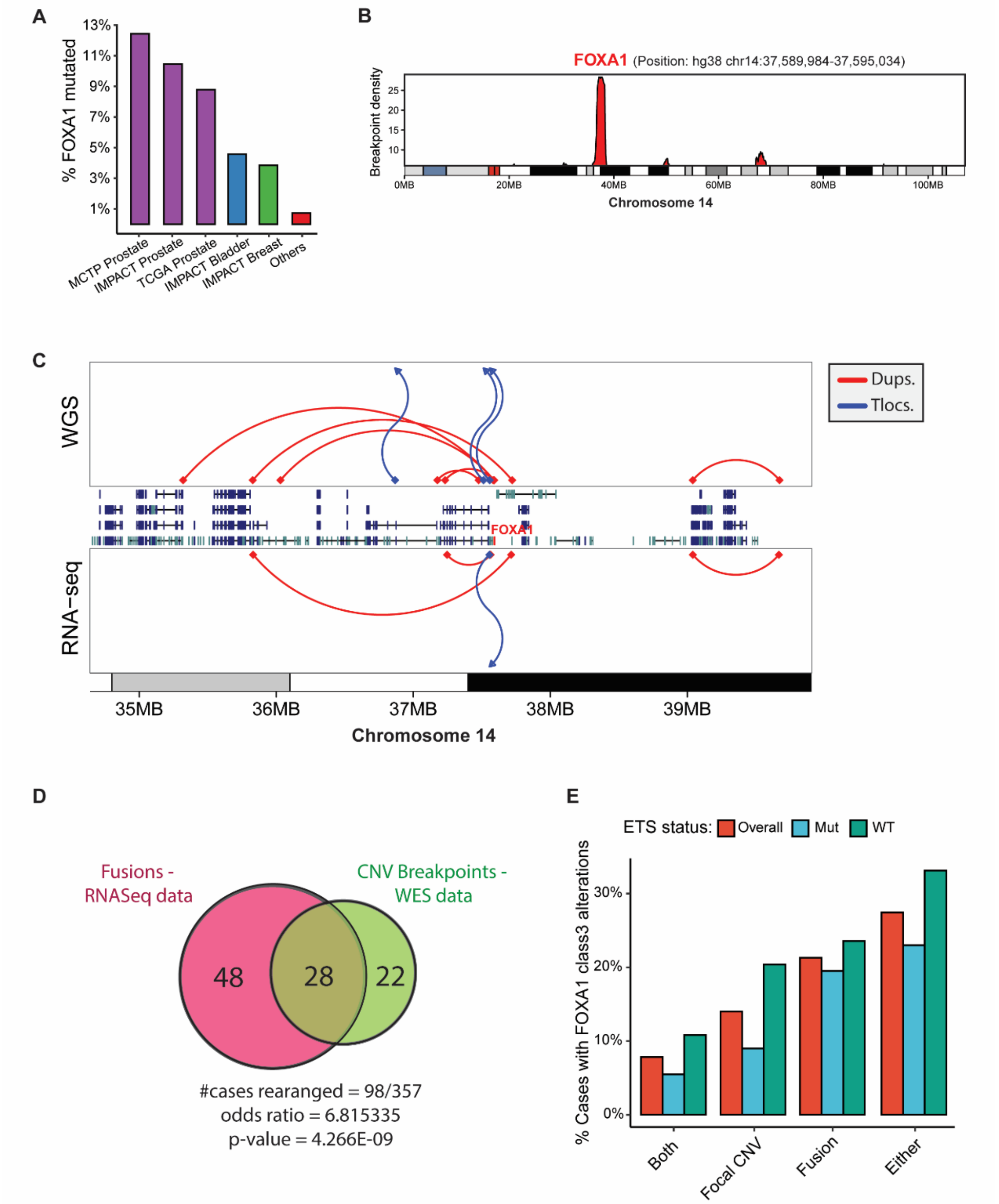
Appendix 1

Supplementary Figures for Chapter 2

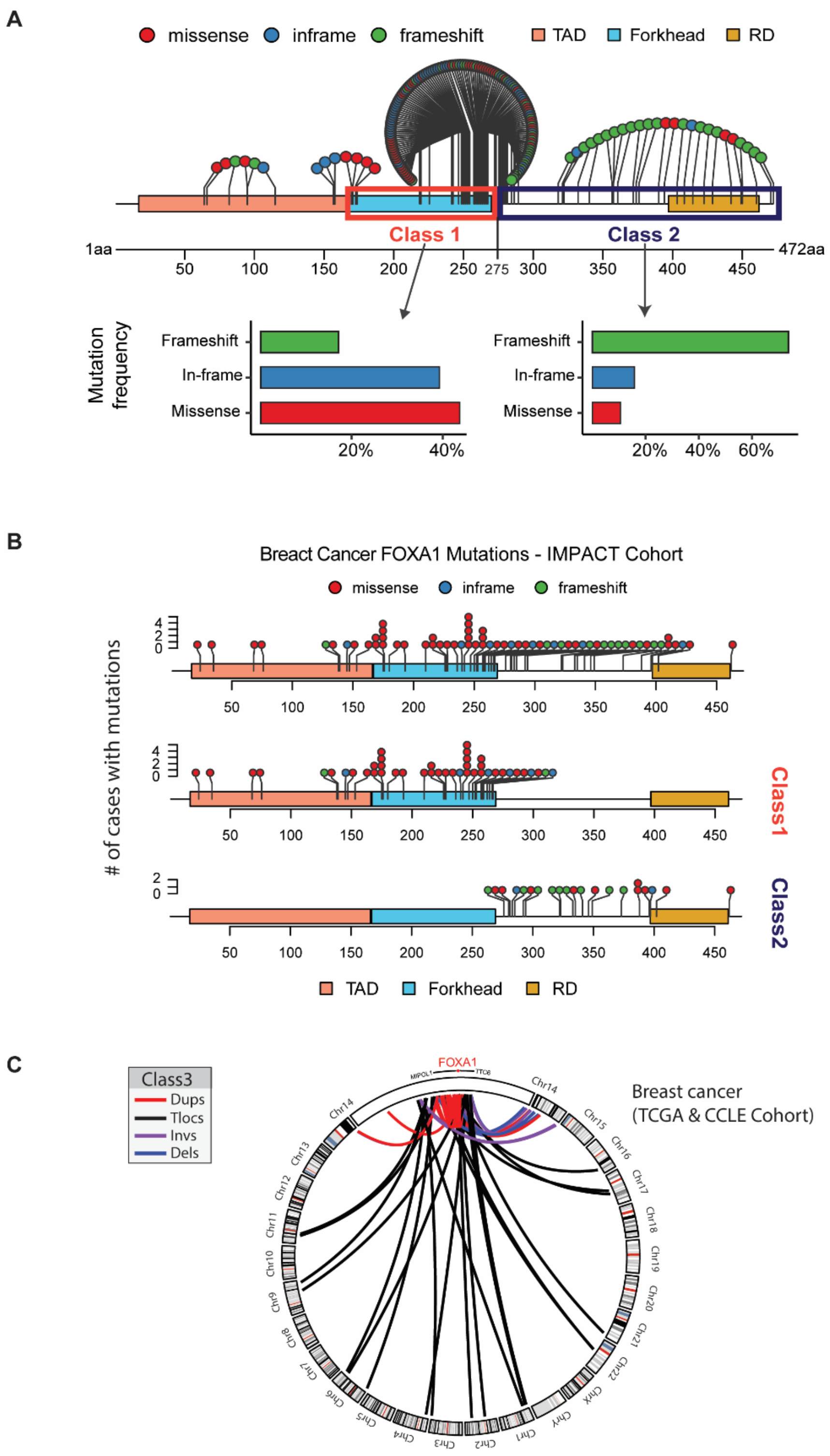


Supplementary Figure A1-1: AR and FOXA1 activity is essential for prostate cancer growth and survival. (A) mRNA (qPCR) and **(B)** protein expression of AR and FOXA1 in a panel of PCa cells. **(C)** Growth curves of AR-positive PCa cells treated with non-targeting (siNC), AR or FOXA1-targeting siRNAs (25nM at Day0 and Day1). Immunoblots confirm robust knockdown of FOXA1 protein in LNCaP and LAPC4 72h after siRNA-treatment. **(D)**

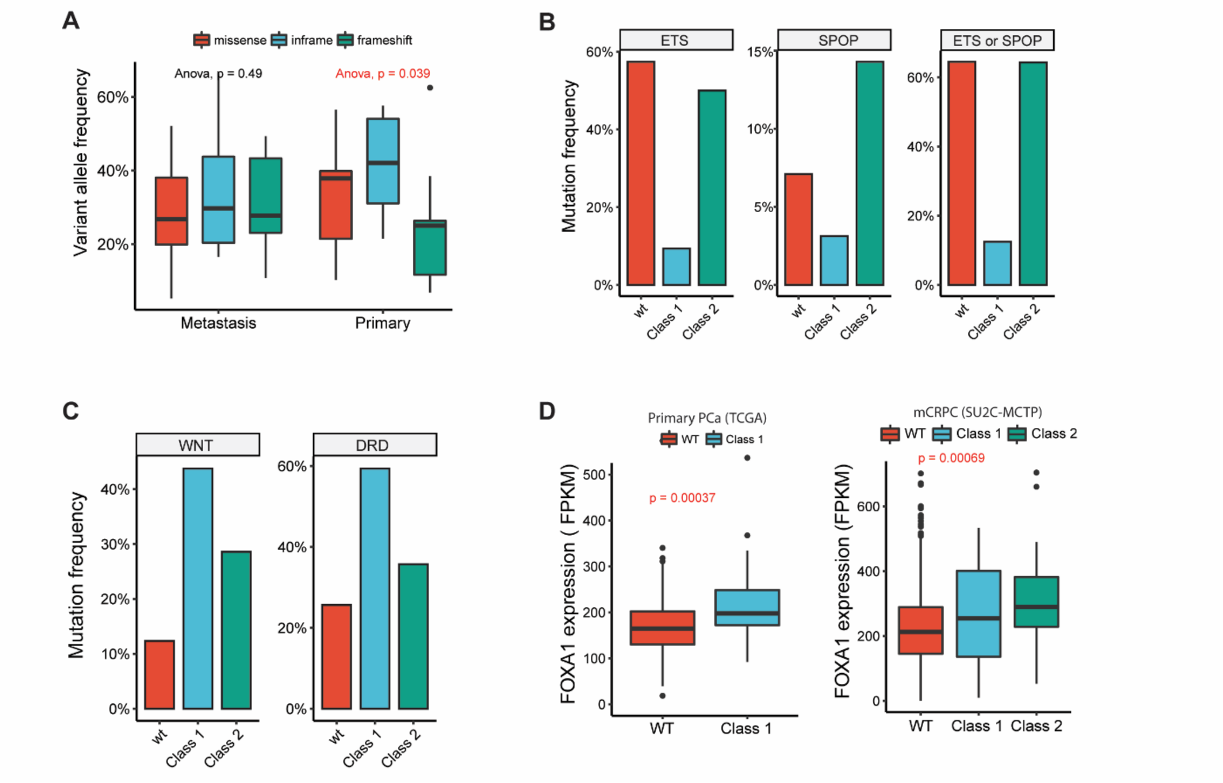
Crystal violet stain of AR-negative, DU145 PCa cells treated with siNC, AR, or FOXA1 targeting siRNAs. LNCaP is shown as a control for comparison. **(E)** Averaged proliferation Z-scores for 6 independent FOXA1-targeting sgRNAs extracted from publically available CRISPR Project Achilles data (BROAD Institute**)** in prostate and breast cancer cells. HPRT1 and AR data serve as negative and positive controls, respectively.



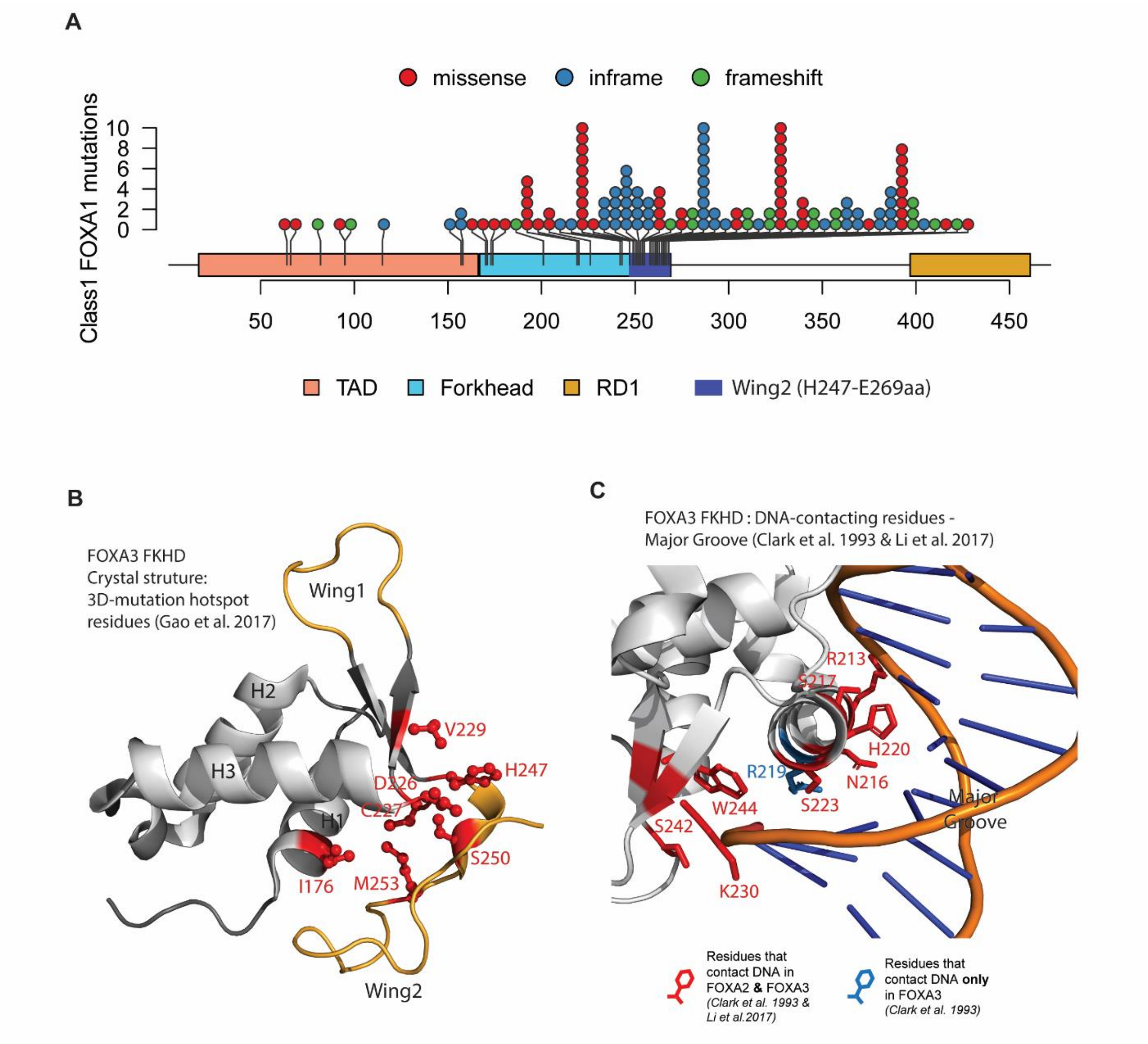
Supplementary Figure A1-2: Structural variations within the FOXA1 locus. (A) Recurrence of FOXA1 mutations across TCGA, MSK-IMPACT, and MCTP cohorts. **(B)** Density of breakpoints (RNA-seq chimeric junctions) within overlapping 1.5Mb windows along chr14 in mCRPC tumors. **(C)** Whole-genome sequencing of 7 mCRPC index cases with distinct patterns FOXA1 locus rearrangements (translocations, duplications), nominated by RNA-seq (WA46, WA37, WA57, MO_1584) or WES (MO_1778, SC_9221, MO_1637). **(D)** Concordance of RNA-seq (chimeric junctions) and WES based FOXA1 locus rearrangements calls (mCRPC cohort). **(E)** Frequency of FOXA1 locus rearrangements in mCRPC based on RNA-seq and WES.



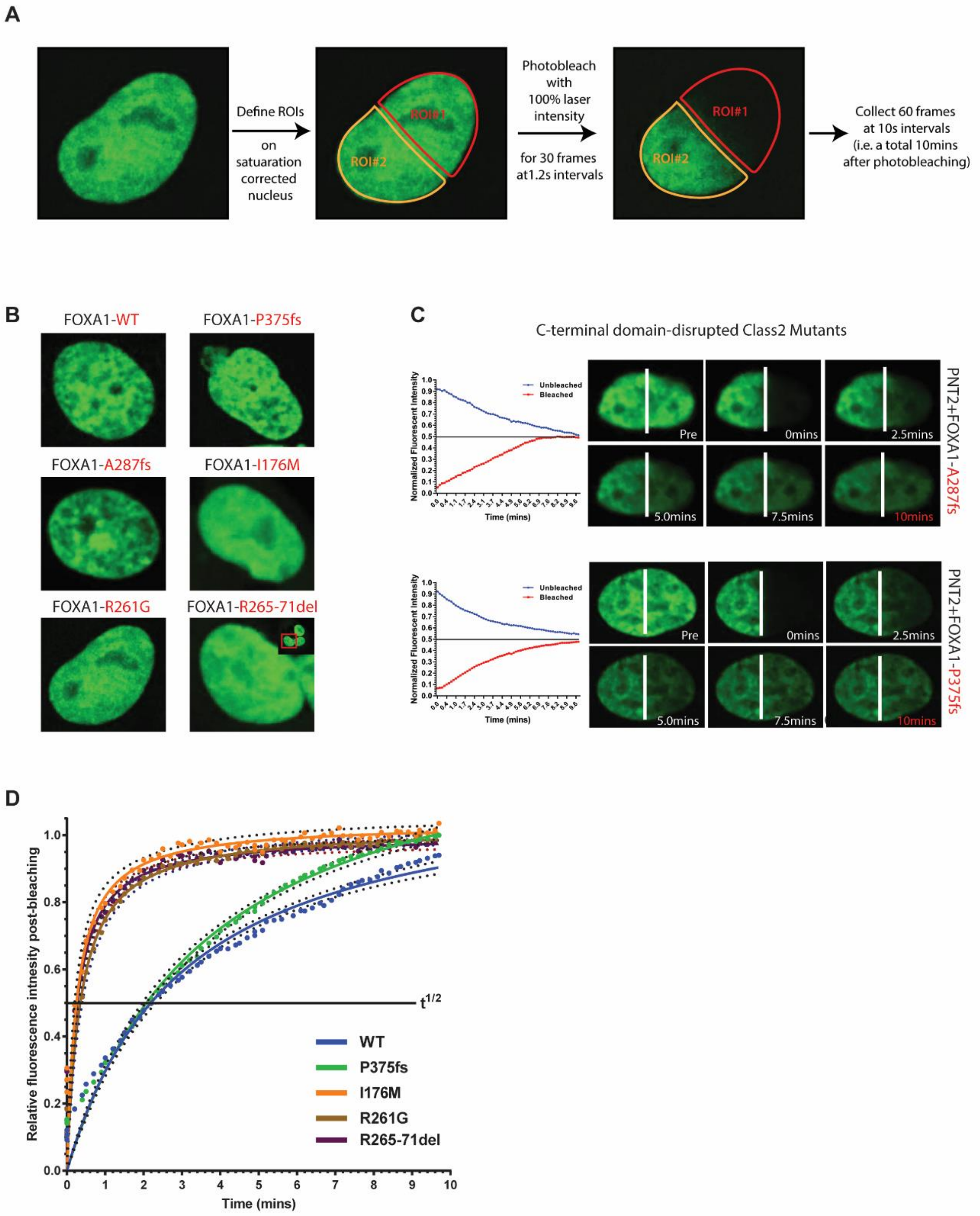
Supplementary Figure A1-3: FOXA1 alterations in breast adenocarcinomas. **(A)** Distribution and functional categorization of FOXA1 mutations (all cases) on the protein map of FOXA1. TAD, trans-activating domain; RD, regulatory domain. **(B)** Aggregate and class-specific distribution of FOXA1 mutations in advanced breast cancer (MSK-Impact cohort). **(C)** Structural classification of FOXA1 locus rearrangements in breast cancer (TCGA and CCLE cell lines).



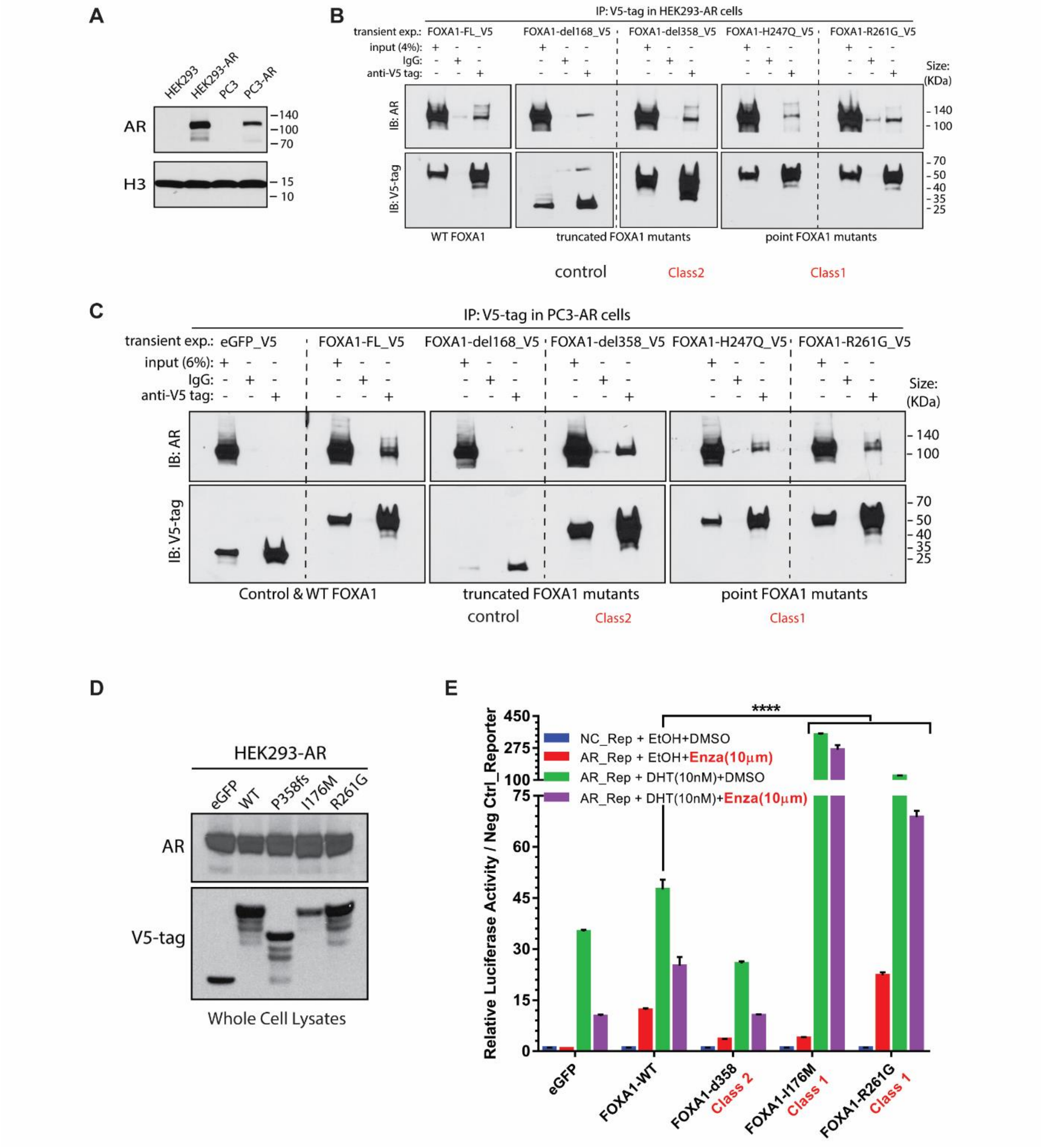
Supplementary Figure A1-4: Class-specific co-alteration profiles of FOXA1 mutations. (A) Variant allele frequency of FOXA1 mutations by tumor stage. **(B)** Mutual exclusivity of FOXA1 and ETS and/or SPOP aberrations in mCRPC. **(C)** Mutual exclusivity of FOXA1 and PI3K (PTEN, PIK3CA) aberrations. **(D)** Activation of PI3K signaling in mCRPC tumors with class1 FOXA1 mutations, based on correlation plot of gene-level fold-changes (left) and gene-set enrichment analysis (right). **(E)** Same as in d but for class2 mutants. **(F)** Co-occurrence of aberrations in the WNT-pathway and DNA repair-deficiency (homologous recombination or CDK12 deficiency) in FOXA1 mutant tumors by class. **(G)** Stage and class-specific increase in FOXA1 expression levels in primary and metastatic PCa.



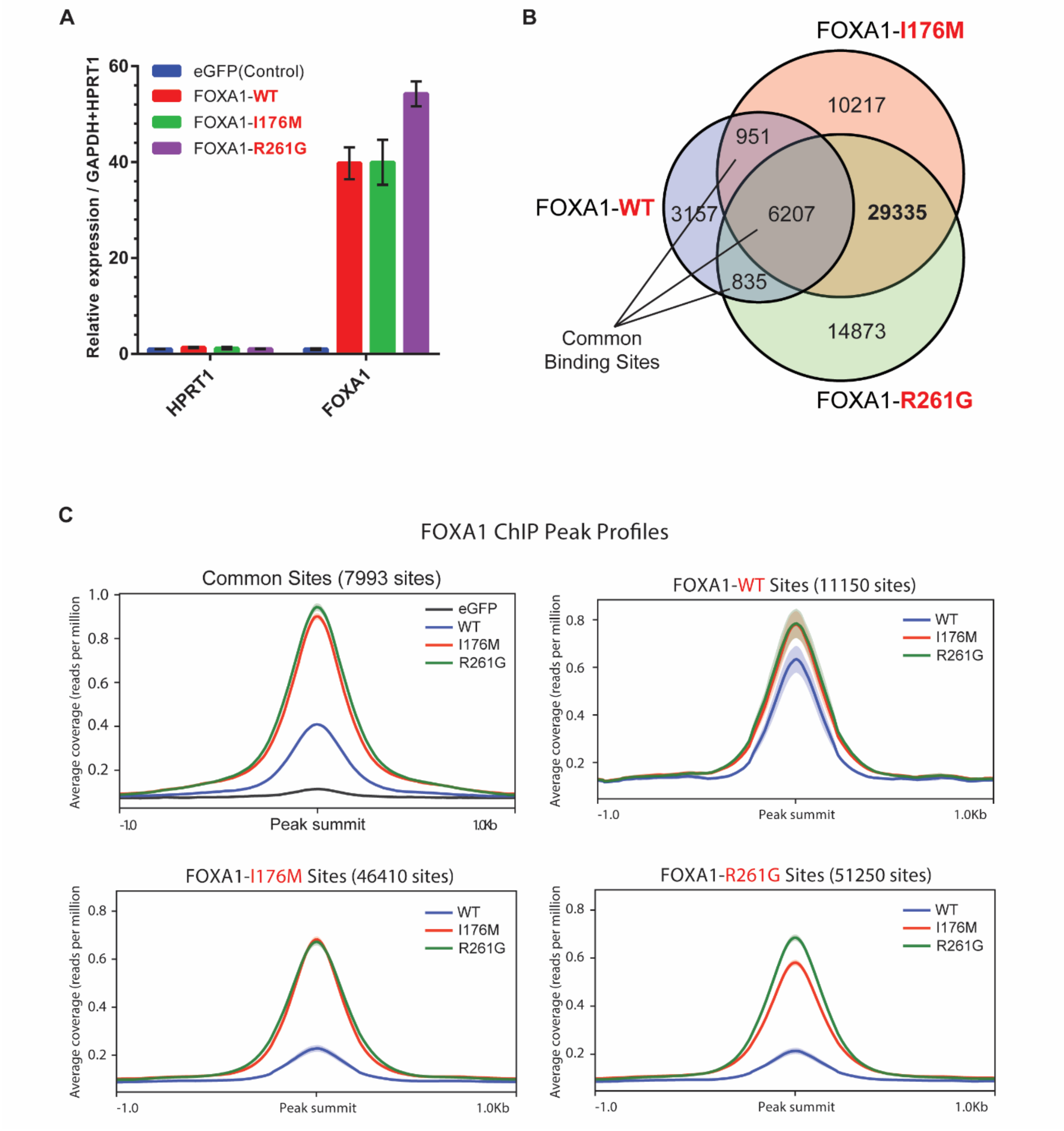
Supplementary Figure A1-5: Structural recurrence of FOXA1 class1 mutations. (A) Distribution of class1 mutations on the protein map of FOXA1. **(B)** 3D-structure of FKHD (FOXA3) with visualization of all mutated residues collectively identified as the 3D-mutational hotspot in FOXA1 across cancers. **(C)** DNA-bound 3D structure of FKHD with visualization of all residues that have been shown through crystallography to make direct base-specific contacts with the DNA in FOXA2 and FOXA3 proteins. FKHD; Forkhead domain.



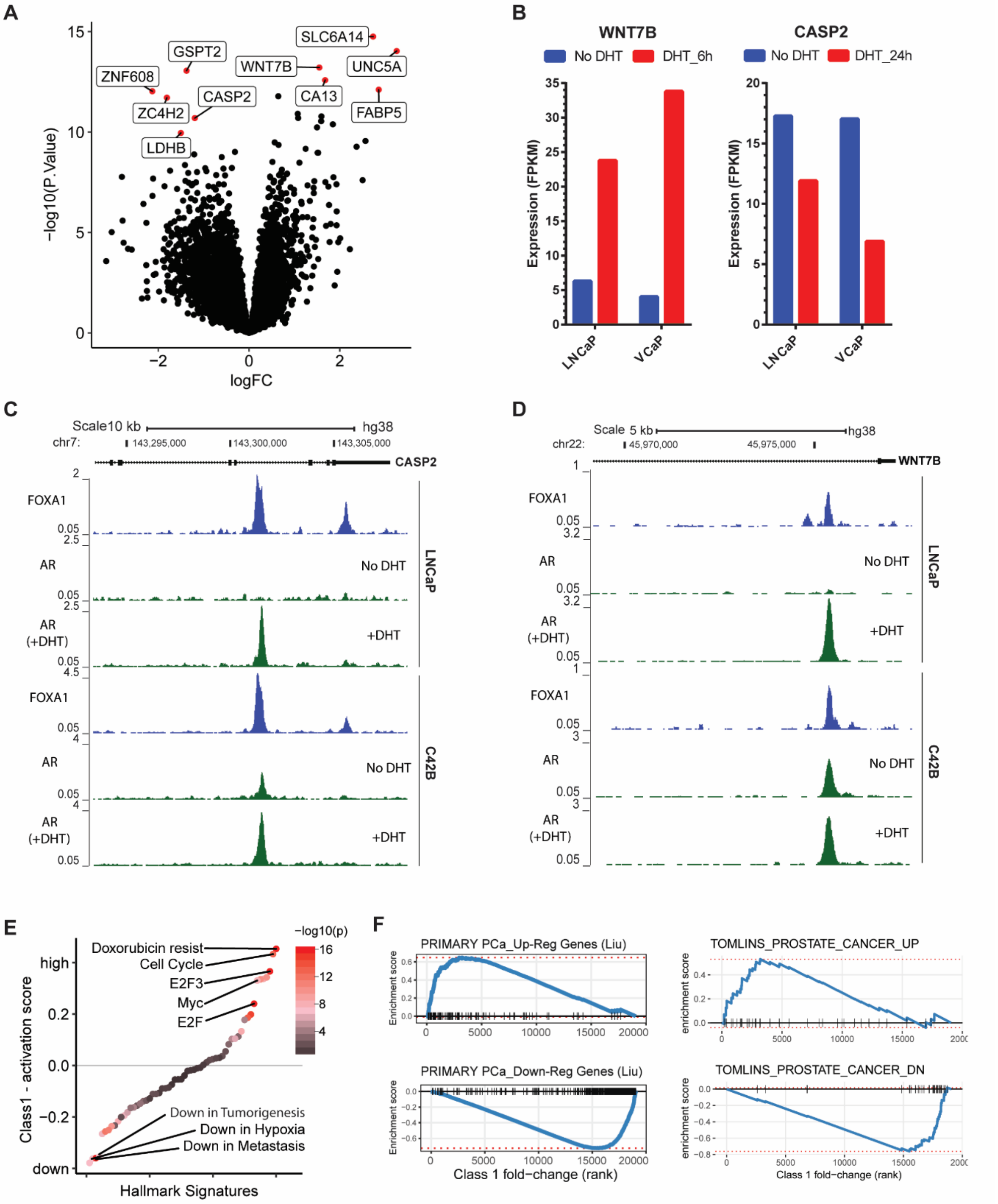
Supplementary Figure A1-6: Workflow for fluorescence recovery after photobleaching assay. (A) Workflow of the FRAP assay as carried out in this study (see Methods for more details of data collection and analyses). **(B)** Representative images of nuclei expressing different FOXA1 variants fused to GFP at the C-termini. All variants majorly localized within the nucleus and had comparable expression. **(C)** FRAP recovery kinetic plots (left) and representative time-lapse images from pre-bleaching ('Pre') to 100% recovery (red timestamps) for class2 mutants. White lines indicate the border between bleached and unbleached areas. **(D)** Representative FRAP fluorescence recovery kinetics in the bleached area for indicated FOXA1 variants. t1/2 line indicates the 50% time to recovery shown in Main Fig. 2D. Colored dots show raw data and superimposed solid curves show a hyperbolic fit with 95% confidence interval (flanking black dotted lines).



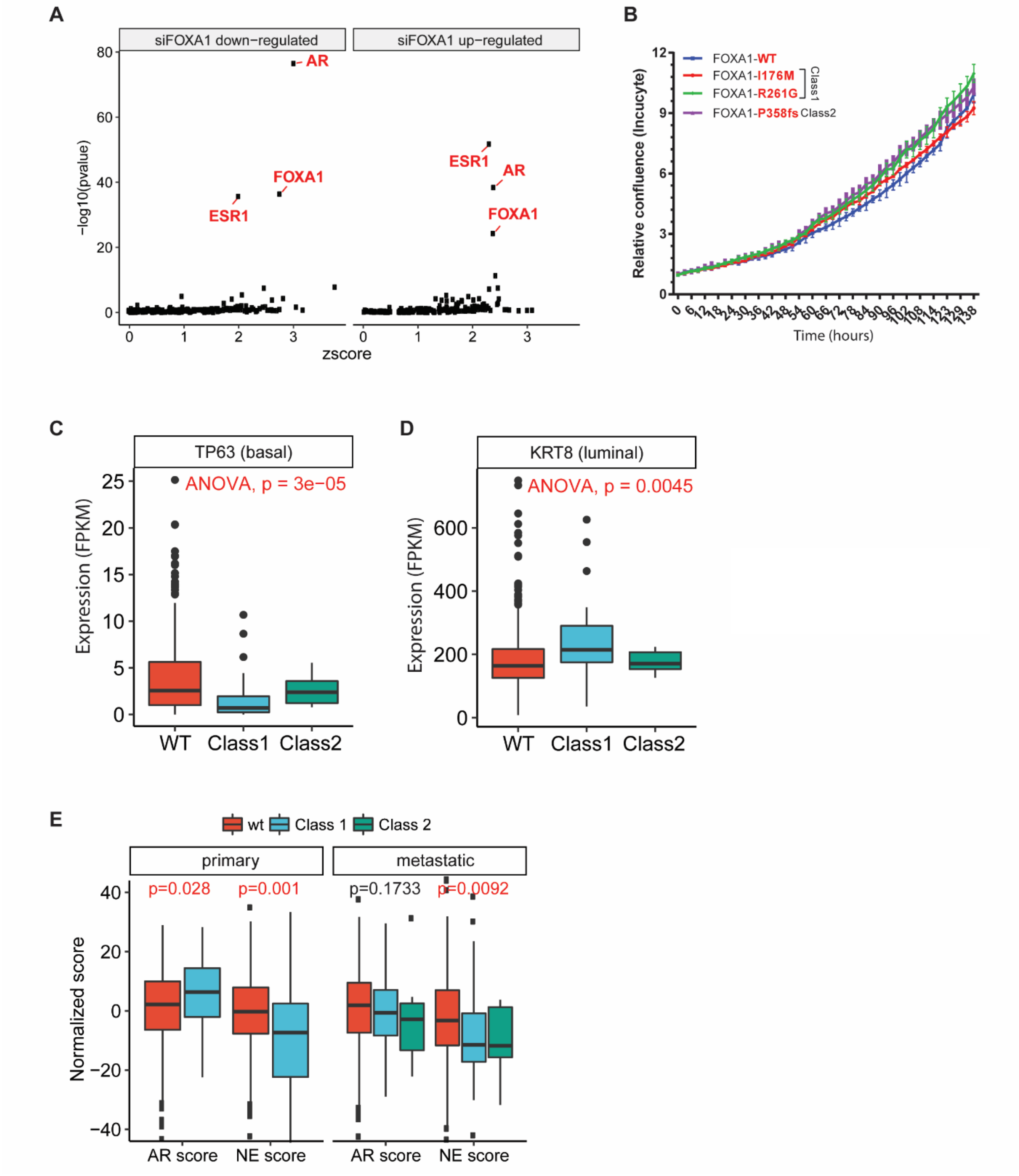
Supplementary Figure A1-7: Functional effect of FOXA1 class1 mutants on AR activity. (A) Immunoblots confirming stable overexpression of the WT AR protein in HEK293 and PC3 cells. **(B)** and **(C)** Co-immunoprecipitation assay of indicated recombinant FOXA1 variants using a V5-tag antibody in HEK293-AR and PC3-AR cells, respectively. FL is the full-length WT FOXA1. d168 and d358 are truncated FOXA1 variants with only the first 168aa (i.e. before the Forkhead domain) or 358aa of FOXA1 protein, respectively. H247Q and R261G are missense class1 mutant variants. **(D)** Immunoblots confirming comparable expression of AR and recombinant FOXA1 variants in AR reporter assay-matched HEK293 lysates. **(E)** AR dual-luciferase reporter assays with transient overexpression of indicated FOXA1 variant in HEK293-AR cells with or without DHT stimulation (6 replicates/group; ANOVA and Tukey's test). Enzalutamide (Enza**)** is a clinically used AR-inhibitor.



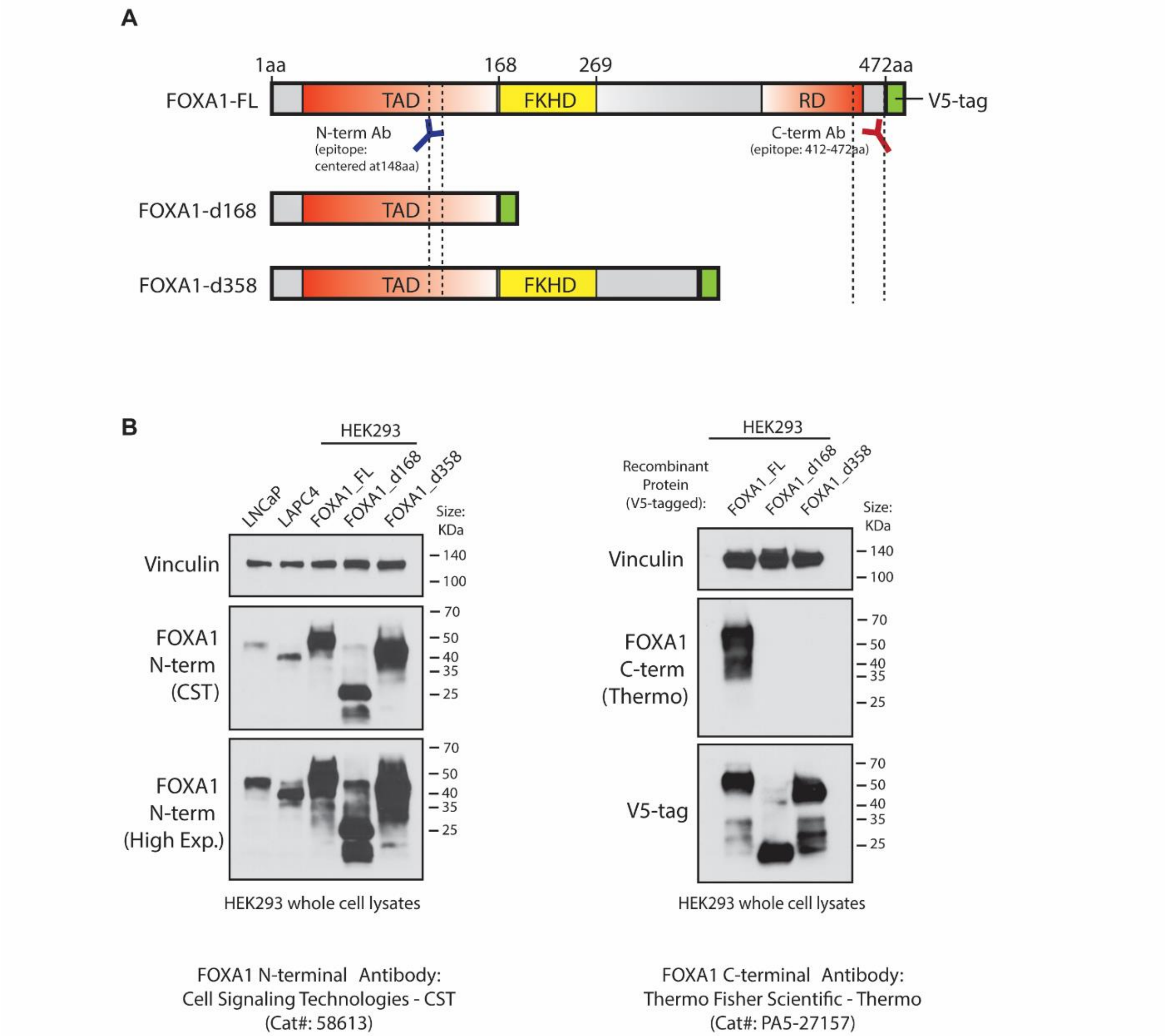
Supplementary Figure A1-8: Cistromic features of FOXA1 class1 mutants. (A) mRNA expression (qPCR) of labeled FOXA1 variants in stable, isogenic HEK293 cells. **(B)** Overlaps between FOXA1 WT and class1 mutant cistromes. Common binding sites are defined as sites shared by all or WT and either of the class1 variants. **(C)** Average peak profiles from ChIP-Seq data at FOXA1 common or variant-specific cistromes. Transparent peak extensions show standard error.



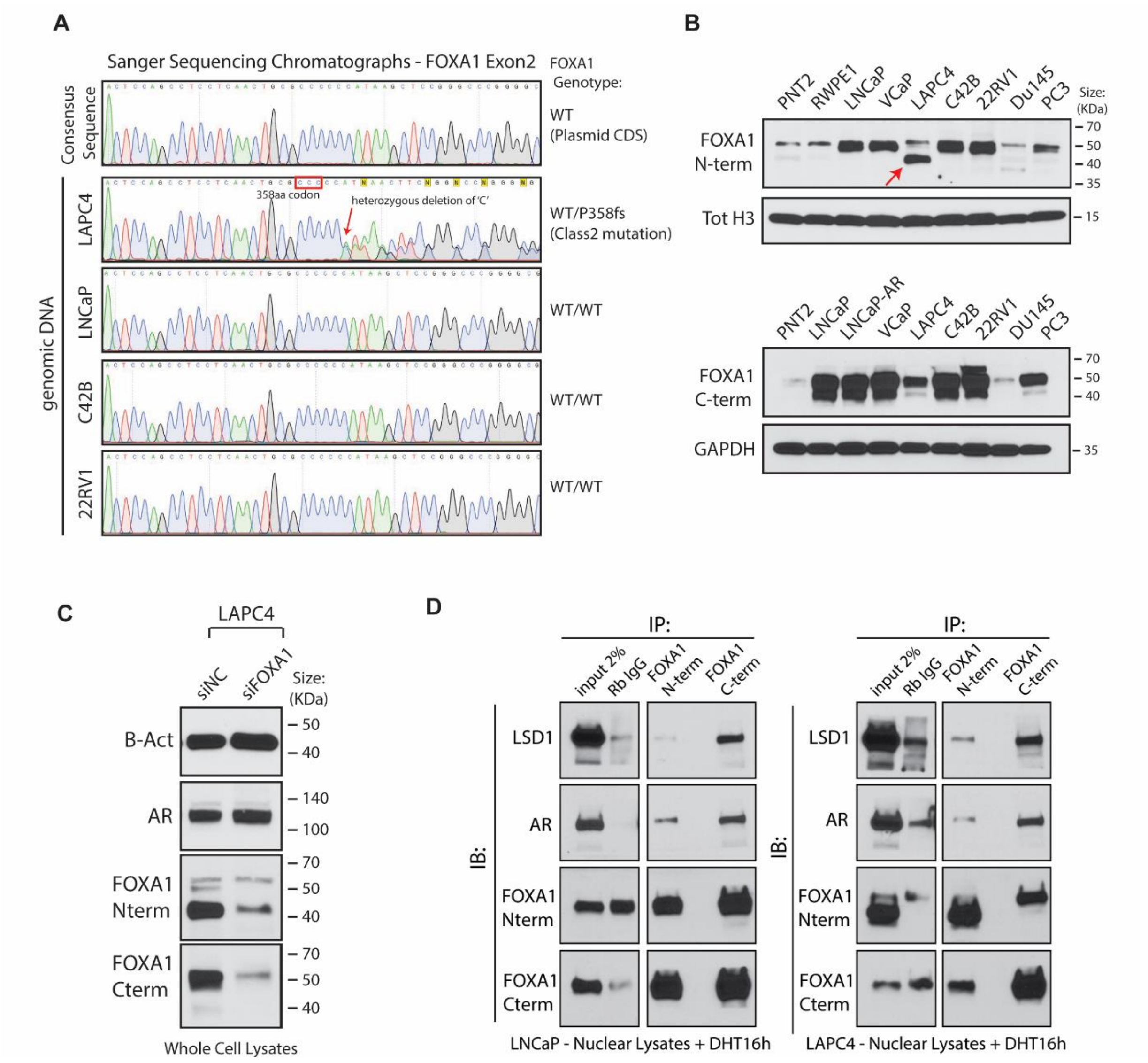
Supplementary Figure A1-9: Transcriptional features of FOXA1 class1 mutants. (A) Genes differentially expressed in class1 patient samples compared to FOXA1 WT tumors (Methods). Most significant genes are shown in red and labeled. (**B**, **C)** Gene-set enrichment analysis of the class1 gene expression signature based on a: **(B)** Differential expression of cancer hallmark signatures. **(C)** Barcode-enrichment visualization of the most-significant prostate cancer signatures showing concordance between class1 tumor and primary PCa gene signatures. **(D)** mRNA expression (RNA-Seq) of WNT7B and CASP2 in LNCaP and VCaP cells grown in 10% charcoal-stripped serum with or without treatment with DHT (10nM). **(E)** and **(F)** FOXA1 and AR ChIP-Seq normalized signal tracks at the WNT7B or CASP2 gene locus, respectively, in LNCaP and VCaP cells.



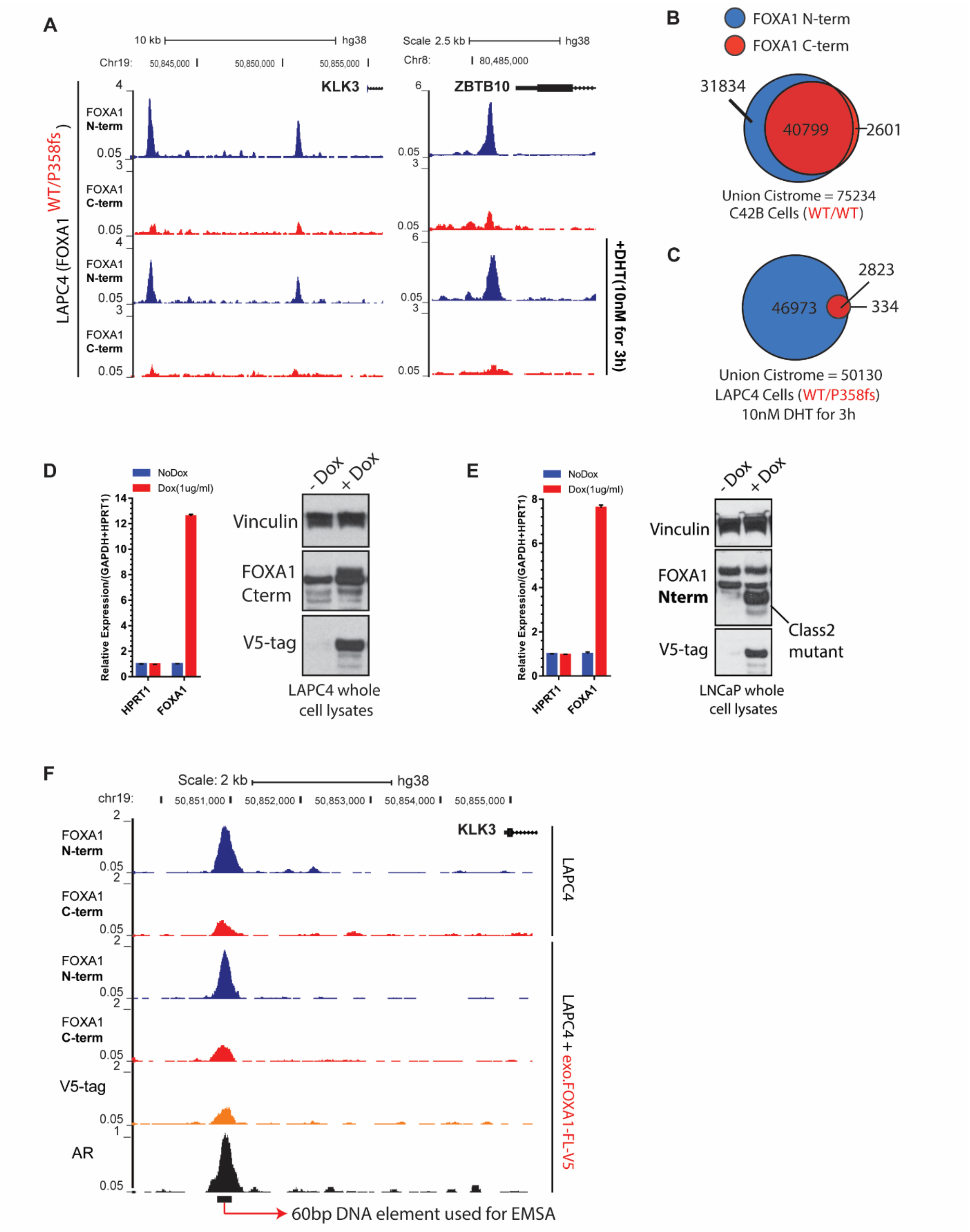
Supplementary Figure A1-10: Basal and luminal signatures in FOXA1 altered tumors. (A) BART prediction of specific TFs mediating observed transcriptional changes. The most likely (based on significance and Z-score**)** mediators of transcriptional responses in LNCaP cells treated with FOXA1-targeting siRNAs are highlighted in red. **(B)** Growth curves of 22RV1 cells with stable, dox-inducible overexpression of FOXA1 mutants or WT FOXA1 in androgen-supplemented complete medium (6 replicates/group). **(C)** and **(D)** mRNA expression (RNA-seq) of labeled basal and luminal TFs or canonical markers in FOXA1 WT, class1 or class2 mutant tumors. **(E)** Extent of AR and NE pathway activation in FOXA1 WT, class1 or class2 mutant cases from both primary and metastatic PCa. Both AR and NE scores were calculated using established gene signatures (see Methods).



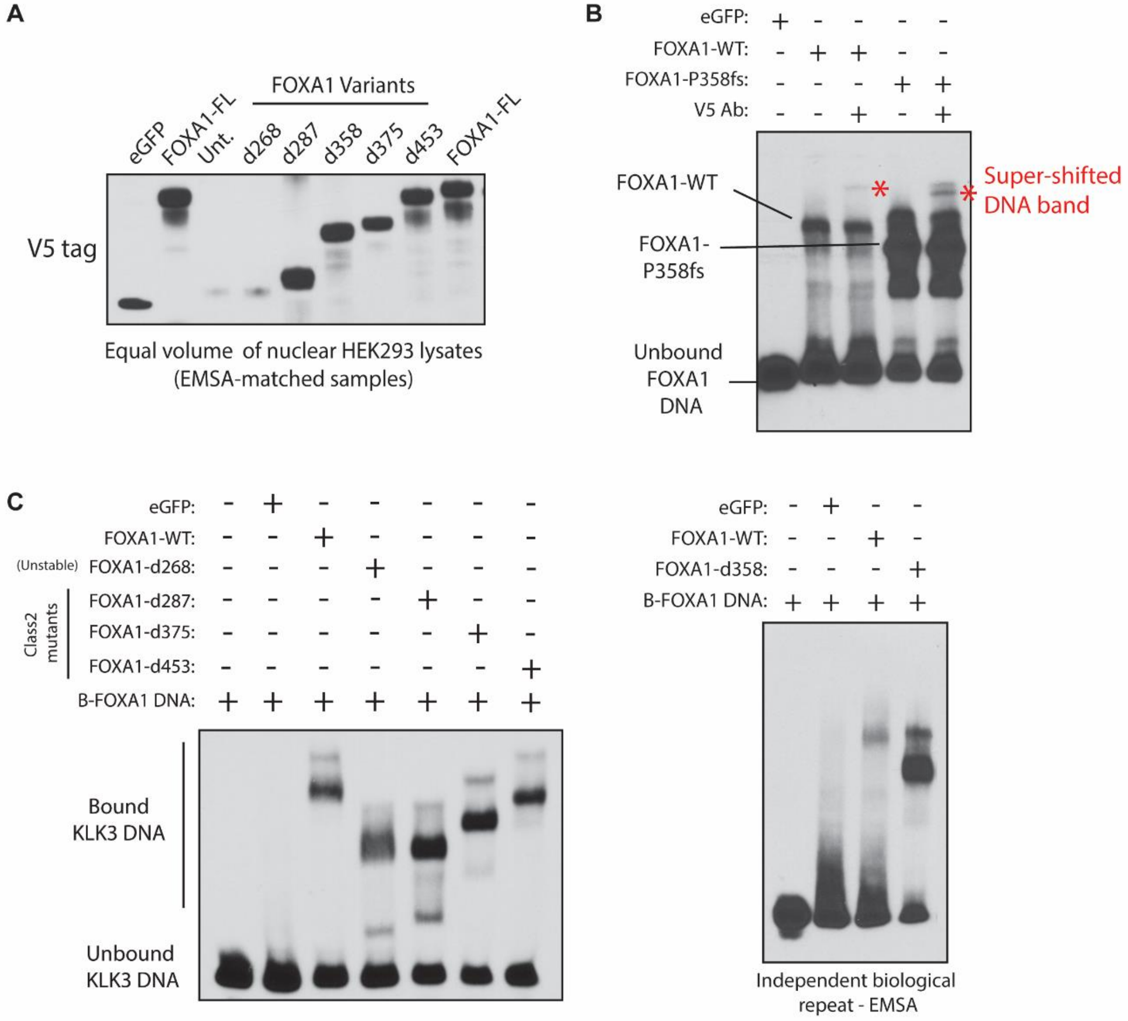
Supplementary Figure A1-11: Validation of FOXA1 antibodies. (A) FOXA1 protein maps showing the recombinant proteins used to validate the N-terminal (N-term) and C-terminal (C-term) FOXA1 antibodies. TAD, trans-activating domian; FKDH, Forkhead domain; RD, regulatory domain. **(B)** Immunoblots depicting detection of all variants by the N-term antibody (left), and detection of only the full-length WT FOXA1 protein by the C-term antibody (right). Antibody details are mentioned and included in the Methods section as well.



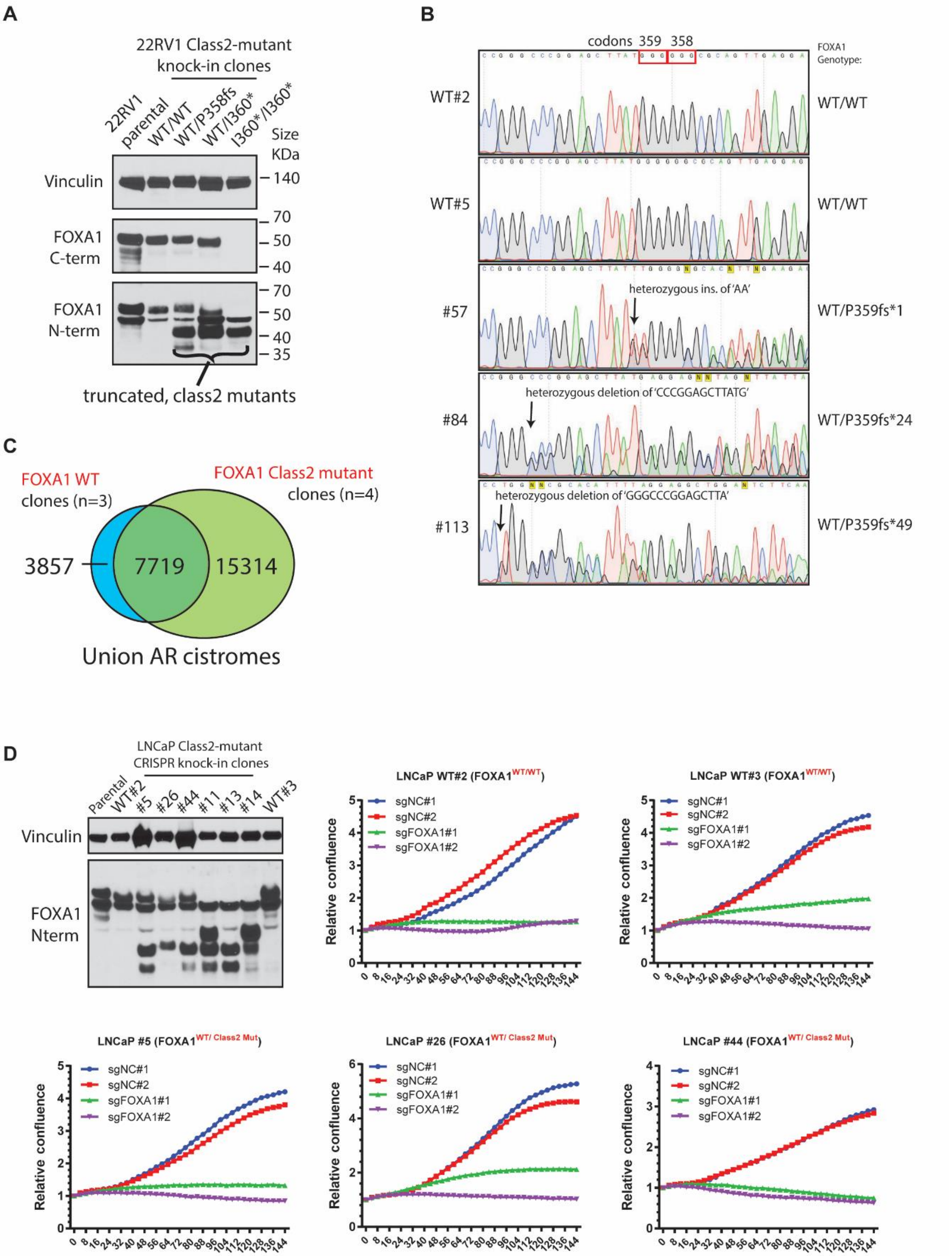
Supplementary Figure A1-12: Genomic and expression-level validation of FOXA1 class2 mutation in LAPC4 cells. (A) Sanger sequencing chromatograms showing the heterozygous class2 mutation in LAPC4 cells after the P358 codon in Exon2. All other tested PCa cells were WT for FOXA1. **(B)** Immunoblots confirming the expression of the truncated FOXA1 variant in LAPC4 at the expected ~40kDa size (top). This band is detectable only with the N-terminal (N-term) FOXA1 antibody and not the C-terminal (C-term) antibody (top and bottom). **(C)** Immunoblots confirming reduction in expression of both WT and mutant variants upon treatment of LAPC4 cells with siRNAs targeting FOXA1 vs non-targeting siRNAs (siNC). **(D)** Nuclear co-immunoprecipitation of FOXA1 from LAPC4 or LNCaP cells stimulated with DHT (10nM for 16h**)** using the N-term and C-term antibodies. Species-matched IgG was used as control. Both WT FOXA1 and P358fs mutant in LAPC4 interact with the AR signaling complex (AR, LSD1).



Supplementary Figure A1-13: Cistromic dominance of FOXA1 class2 mutants. (A) FOXA1 ChIP-seq normalized signal tracks from N-terminal (N-term) and C-terminal (C-term) antibodies in LAPC4 cells with or without DHT-stimulation (10nM for 3h**)** at KLK3 and ZBTB10 locus. **(B)** Overlap between global N-term and C-term FOXA1 cistromes in untreated C42B cells and **(C)** LAPC4 cells treated with DHT (10nM for 3h**)**. **(D)** mRNA (qPCR) and protein (immunoblots) expression of FOXA1 in LAPC4 cells with exogenous overexpression of WT FOXA1, and **(E)** in LNCaP cells with exogenous overexpression of the P358fs mutant. **(F)** FOXA1 ChIP-seq normalized signal tracks from N-term and C-term antibodies in parental LAPC4 cells and LAPC4 cells overexpressing WT FOXA1 at the KLK3 locus. The 60bp AR and FOXA1 bound KLK3 enhancer element used for EMSA is shown.

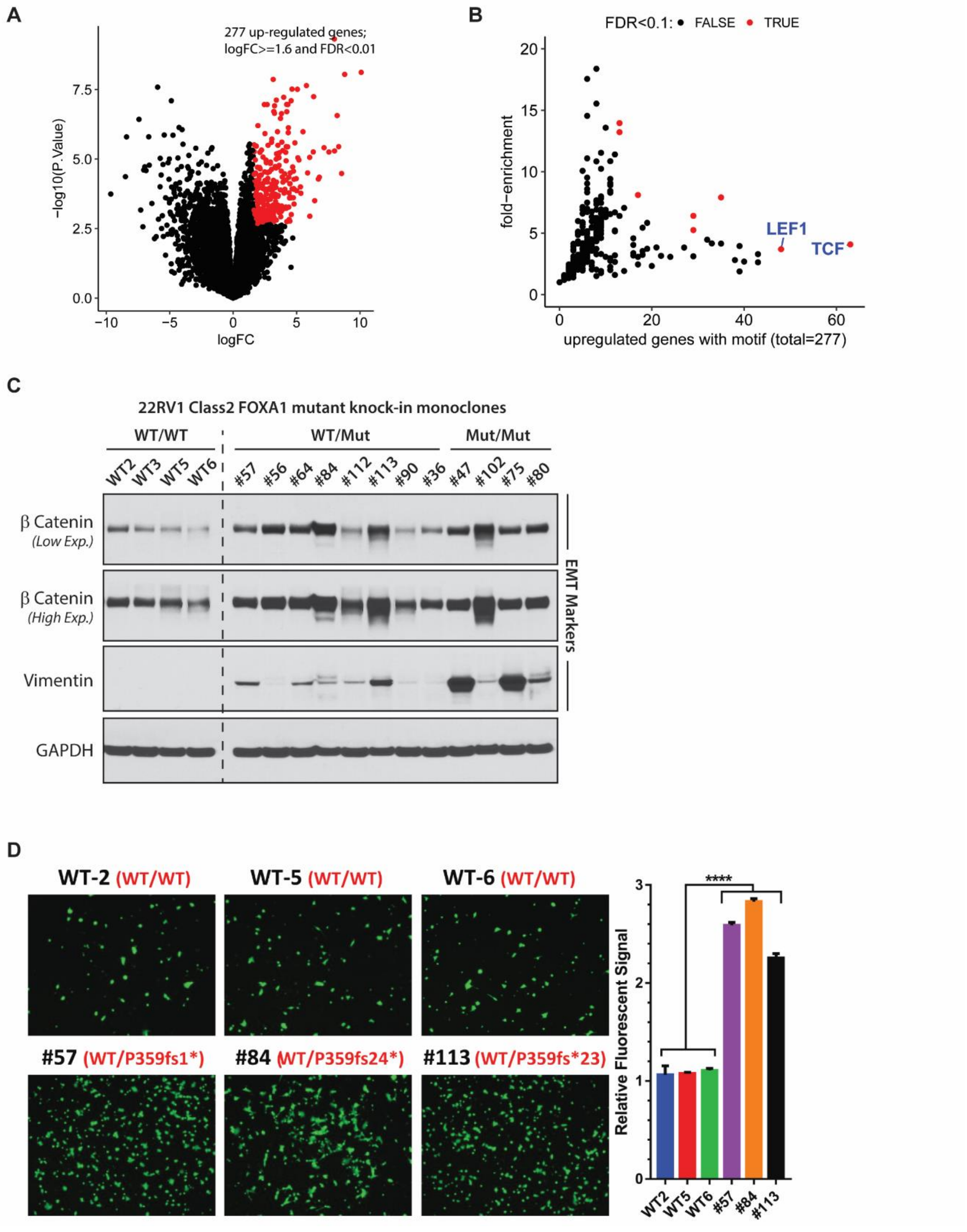


Supplementary Figure A1-14: Electromobility shift assay using distinct clinically-detected FOXA1 class2 mutants. **(A)** Immunoblot showing comparable expression of recombinant FOXA1 variants in equal volume of nuclear HEK293 lysates used to perform EMSAs. **(B)** Higher exposure of EMSA with recombinant WT or P358fs mutant and KLK-enhancer element showing the super-shifted band with the addition of the V5 antibody (red asterisks; matched to Main Fig. 3F). **(C)** EMSA with recombinant WT or different class2 mutants and KLK3 enhance element (left) and independent repeat with the P358fs mutant (right). Class2 mutants display higher affinity vs WT FOXA1**.**

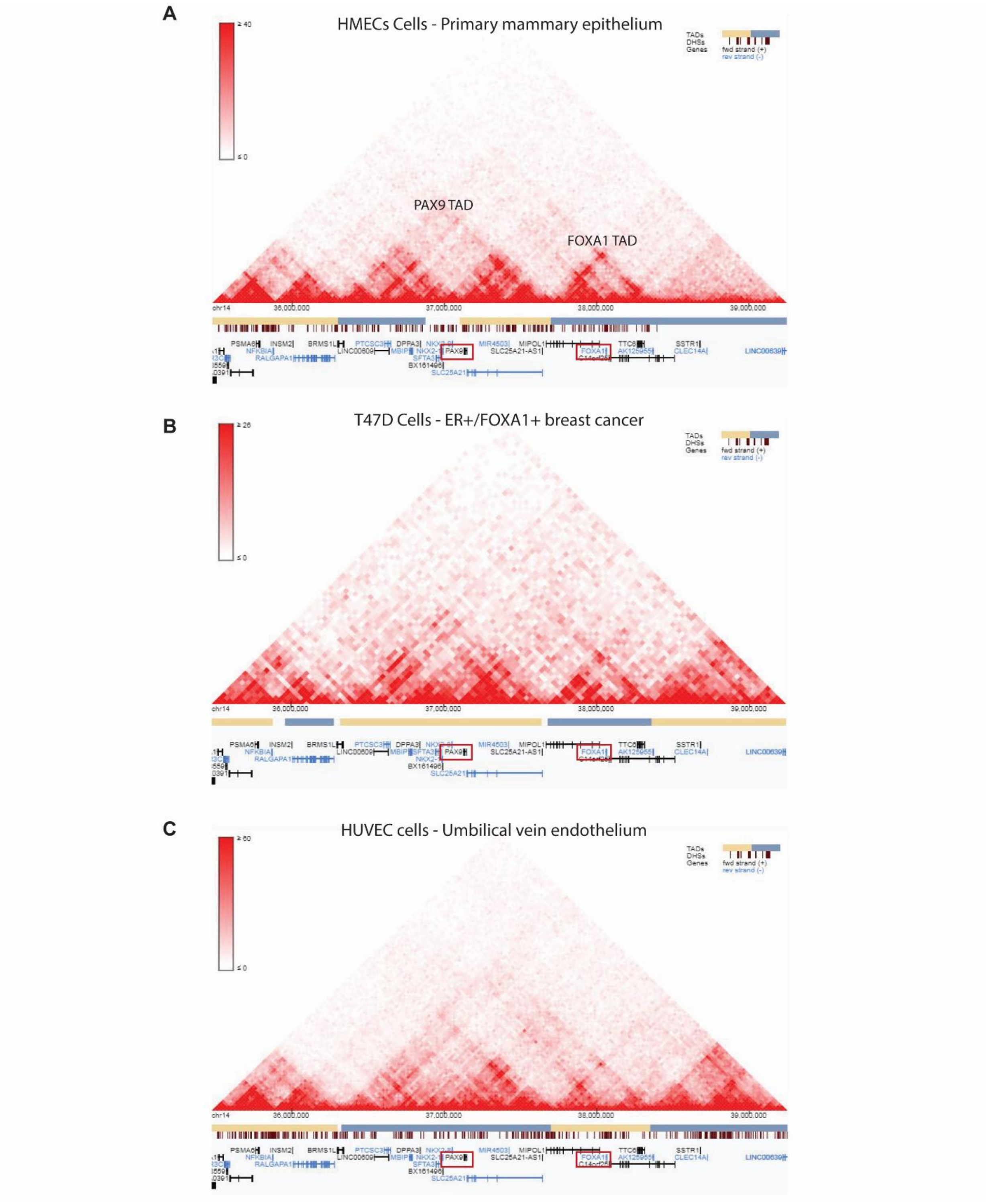


Supplementary Figure A1-15: Functional essentiality of FOXA1 class2 mutants. (A) Immunoblots showing the expression of endogenous WT or class2 mutant FOXA1 variants in parental and distinct CRISPR-engineered 22RV1 clones. **(B)** Sanger sequencing chromatograms from another set of 22RV1 CRISPR clones confirming the introduction of distinct indels in endogenous FOXA1 allele resulting in a premature stop codon. Protein mutations are identified on the right. **(C)** Overlap between union AR cistromes from WT (n=3) and class2-mutant (n=4) 22RV1 clones. **(D)**

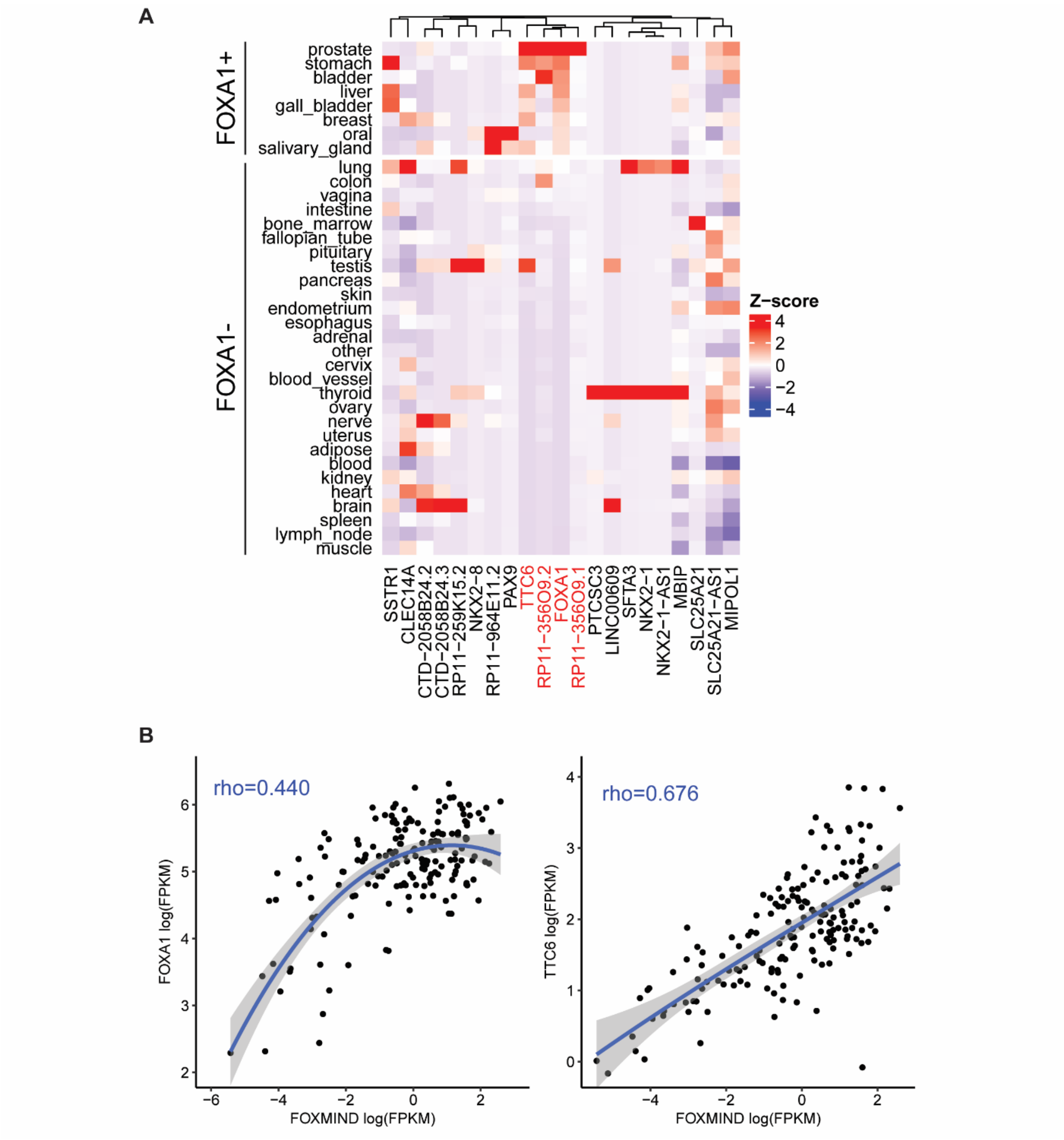
Immunoblots showing expression of FOXA1 (N-term antibody) in parental and CRISPR-engineered LNCaP clones expressing distinct class2 mutants with truncations closer to the Forkhead domain (top left). The remaining panels show growth curves of WT or mutant clones upon treatment with non-targeting or FOXA1-targeting sgRNAs and CRISPR-Cas9 protein (see Methods). Both WT and class2 clones acutely rely on FOXA1 activity for survival.



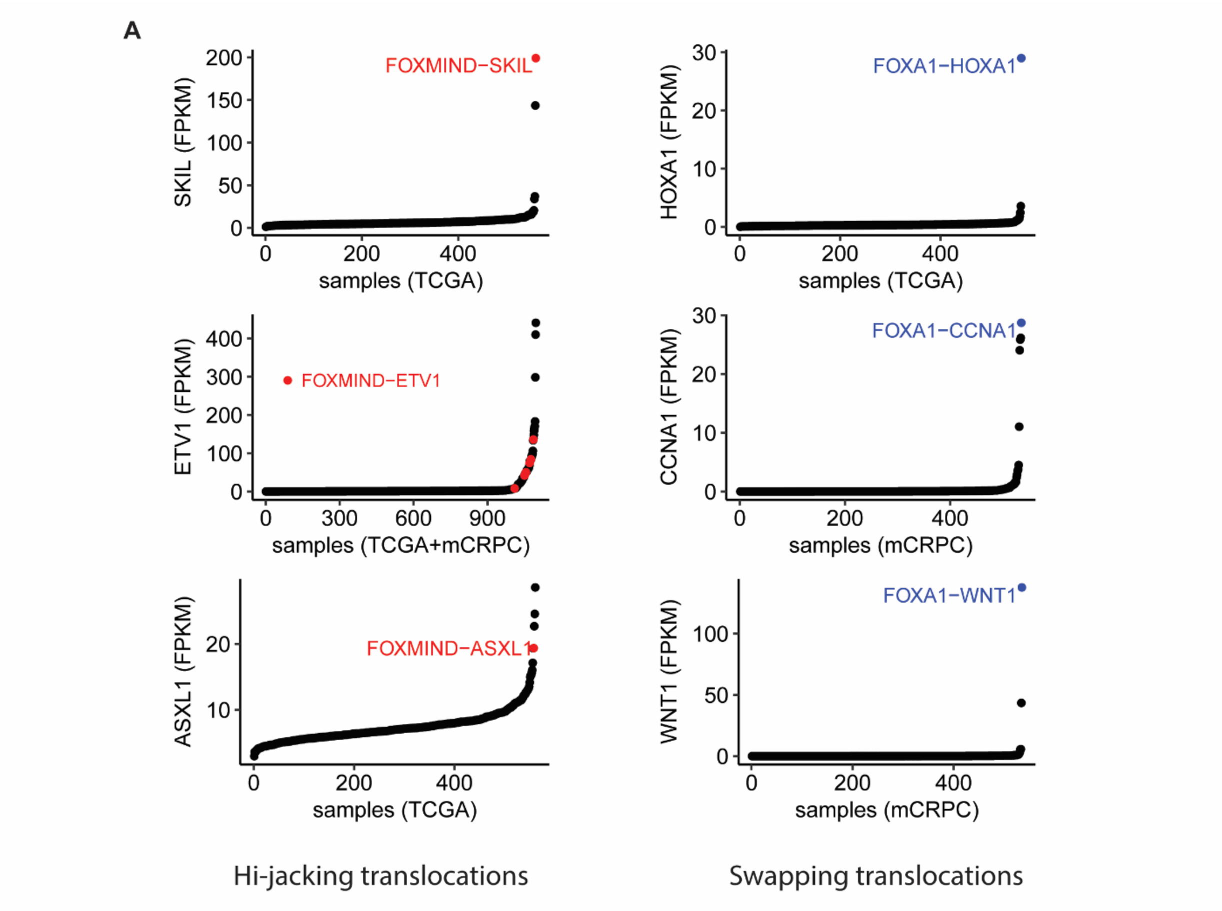
Supplementary Figure A1-16: Functional effect of FOXA1 class2 mutants on WNT-signaling. (A) Differential expression of genes in FOXA1 Class2 CRISPR mutant clones relative to FOXA1 WT clones (n=2). **(B)** Distinct TF motifs within the promoter (2kb upstream) of differentially expressed genes. TFs with the highest enrichment (fold-change, percent of up-regulated genes with the motif, and significance**)** are highlighted and labeled. **(C)** Immunoblots showing the expression of B-Catenin and Vimentin in a panel of WT and heterozygous or homozygous class2 mutant 22RV1 CRISPR clones. **(D)** Representative images of Boyden chambers showing invaded cells stained with Calcein-AM dye (left). Quantified fluorescence signals from invaded cells are shown on the right (4 replicates/group; ANOVA and Tukey's test).



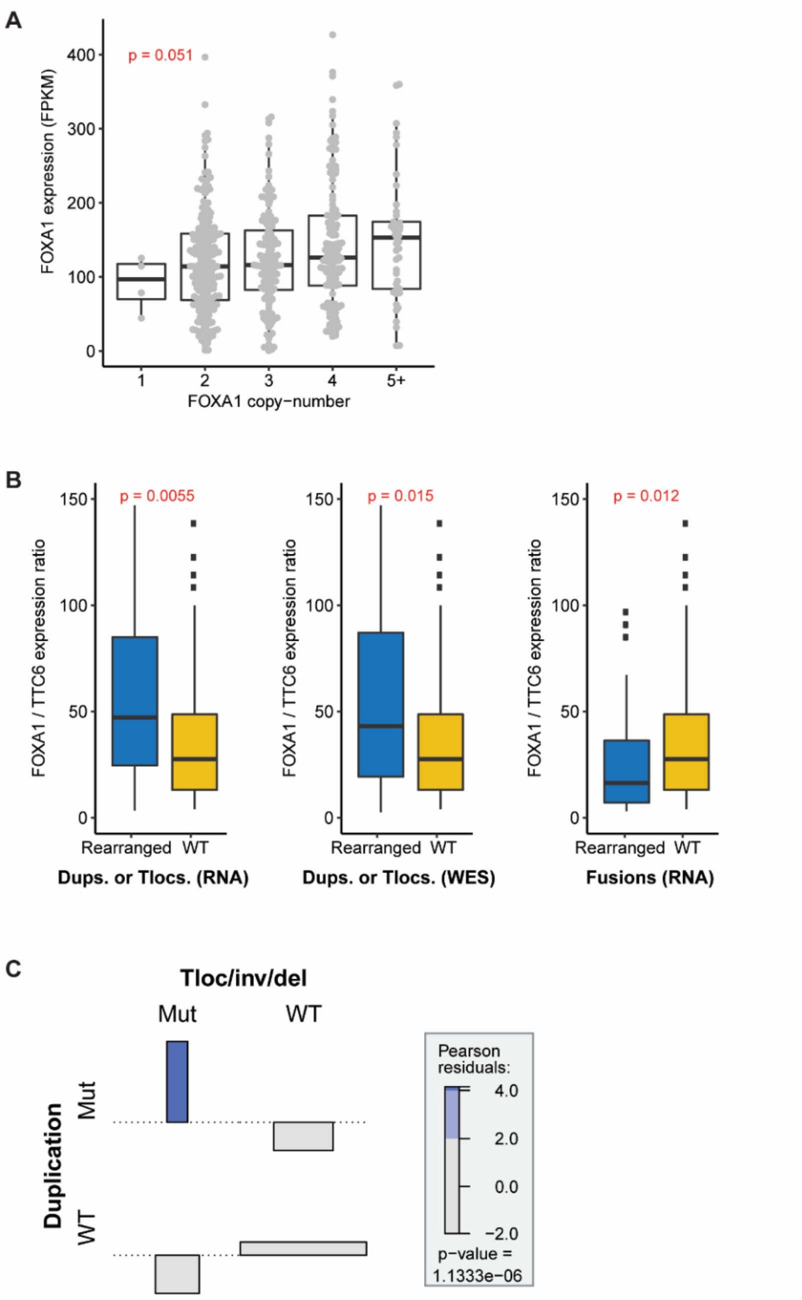
Supplementary Figure A1-17: Topologically associating domain of FOXA1. HI-C data (retrieved from http://promoter.bx.psu.edu/hi-c/view.php) depicting conserved topological domains within the PAX9/FOXA1 syntenic block in normal and FOXA1-positive cancer cell lines. **A, B,** and **C** show the TAD configuration within the FOXA1 locus in three distinct cellular lineages.



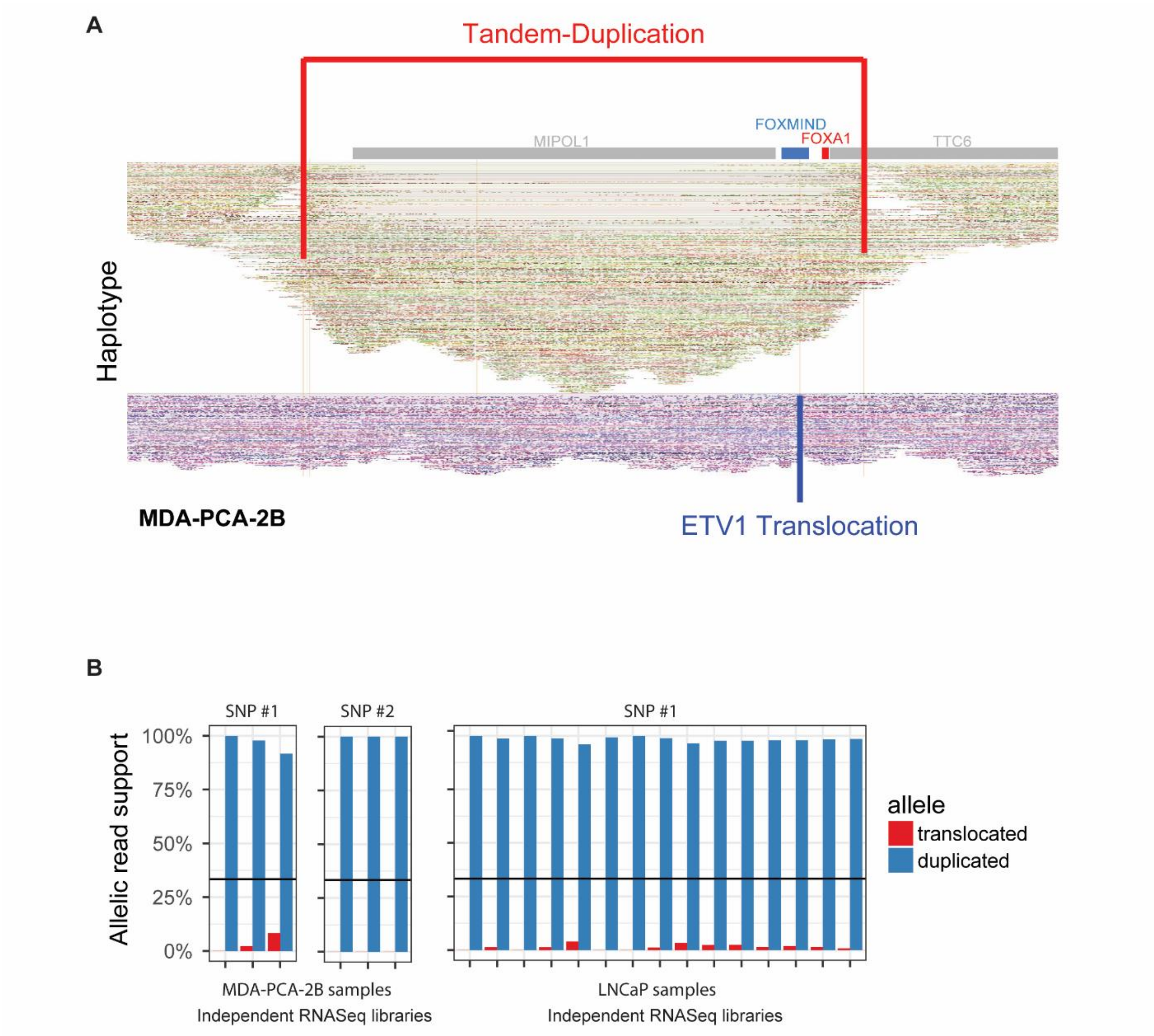
Supplementary Figure A1-18: Correlation of *FOXA1* and *FOXMIND* expression in normal and cancerous tissues. (A) Highly tissue-specific patterns of gene expression within the PAX9/FOXA1 syntenic block. Tissues were dichotomized into FOXA1 "+" and FOXA1 "-" based on FOXA1 expression levels; genes were subject to unsupervised clustering. Z-score normalization was performed for each gene across all tissues. **(B)** Correlation of FOXMIND (Methods) and FOXA1 / TTC6 expression levels across metastatic prostate cancer tissues. (Spearman rank-correlation coefficient)



Supplementary Figure A1-19: Outlier gene expression from translocations within the *FOXA1* locus. (A) Outlier expression of genes involved in translocations with the FOXA1 locus. Translocations positioning a gene between FOXMIND and FOXA1 (Hijacking**)**, are shown on the left (red**)**. Translocations positioning a gene upstream of the FOXA1 promoter (Swapping**)** are shown on the right (blue**)**.



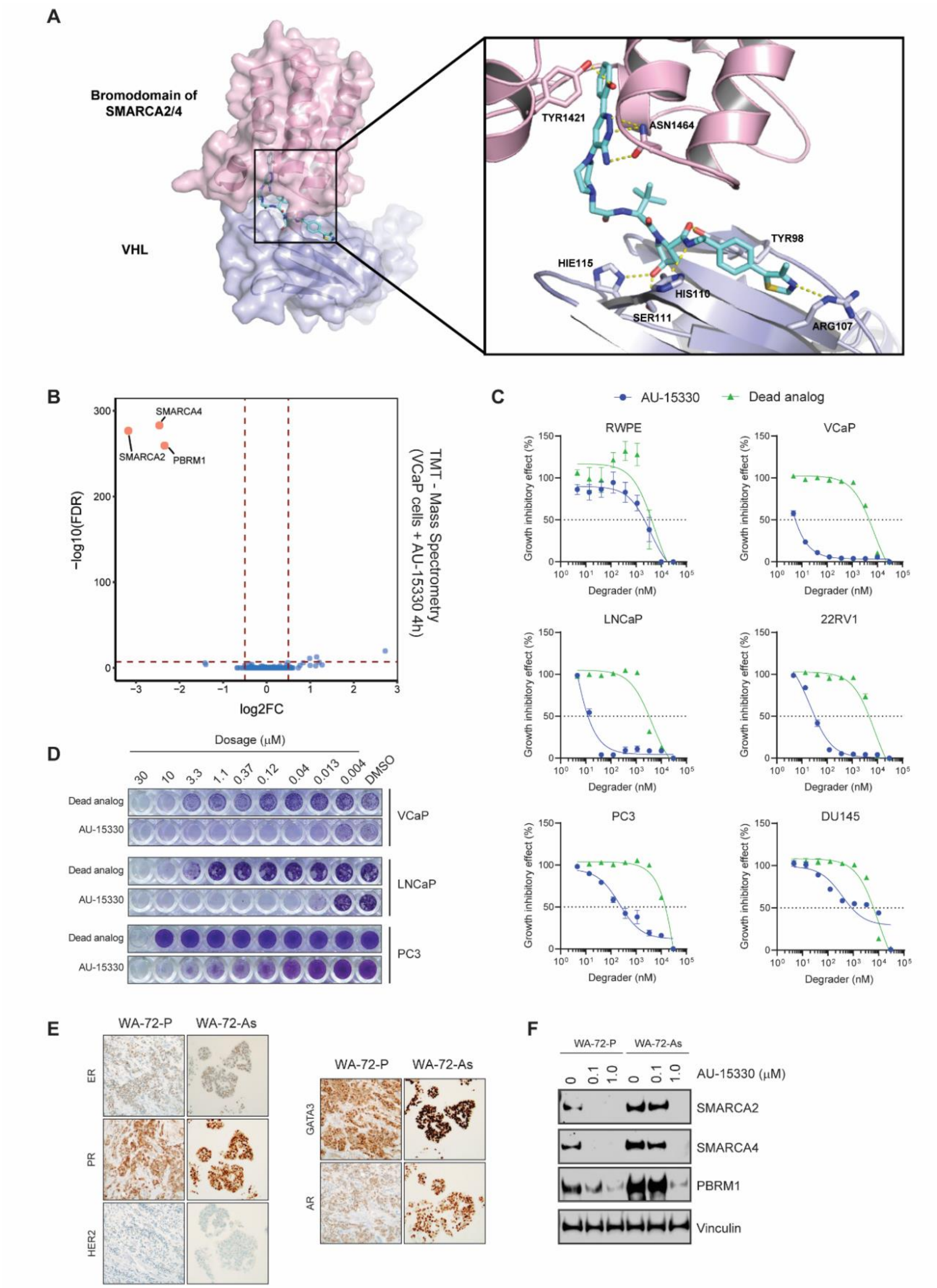
Supplementary Figure A1-20: *FOXA1* expression in rearranged prostate cancer cases. (A) Dosage sensitivity of the FOXA1 gene. Expression of FOXA1 (RNA-seq) across mCRPC tumors as a function of gene ploidy (as determined by absolute copy number at the FOXA1 locus, ANOVA). **(B)** Relative expression of FOXA1 (within the minimally amplified region) to TTC6 (outside the MAR) in rearranged (duplication or translocation) vs WT FOXA1 loci. **(C)** Association plot, visualizing the relative enrichment of cases with both translocation and duplications within the FOXA1 locus. The over-abundance of cases with both events is quantified using Pearson-residuals, the significance of this association is based on the Chi-square test without continuity correction.



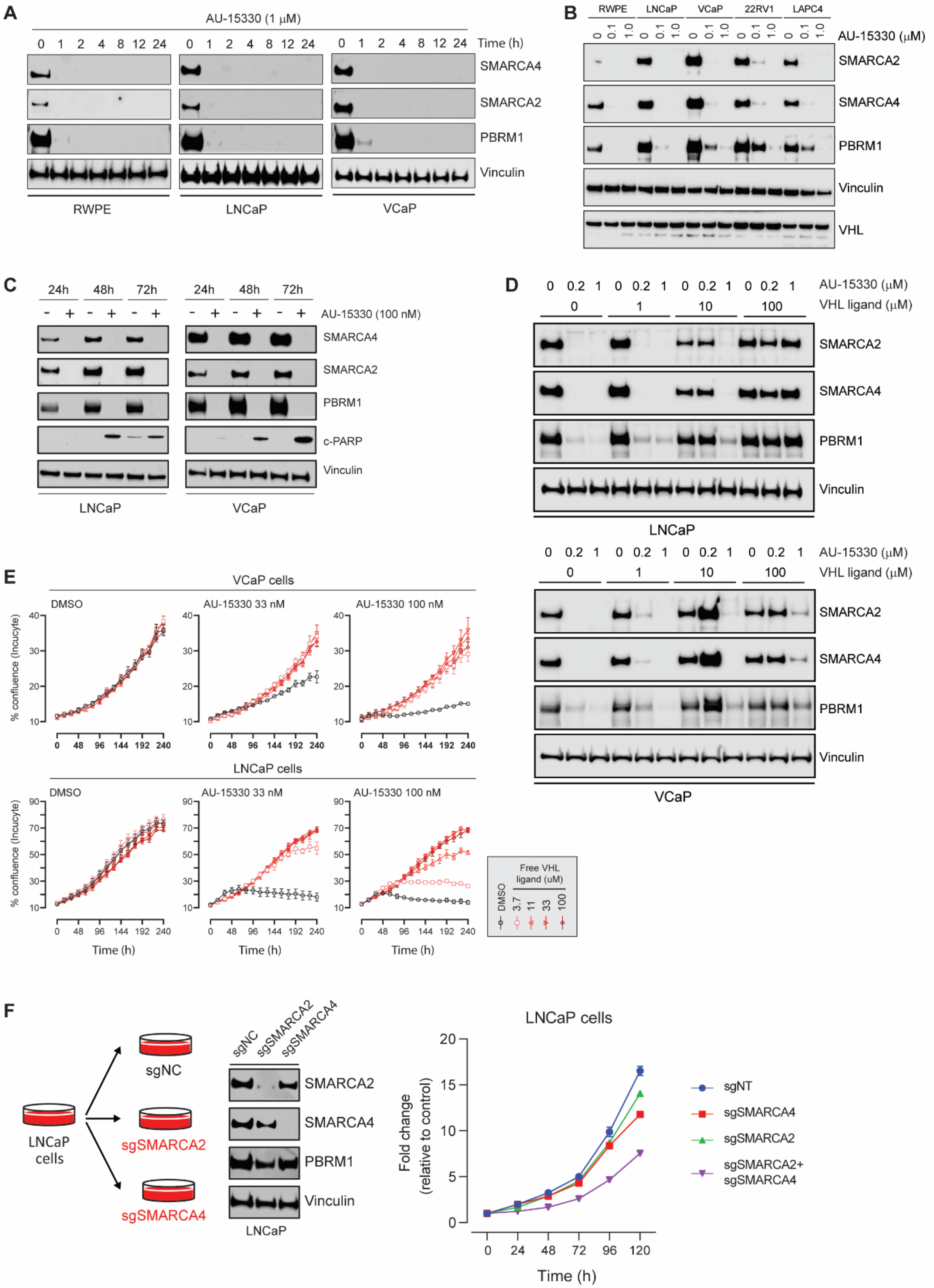
Supplementary Figure A1-21: FOXA1 duplication and monoallelic expression in prostate cancer cells. (A) FOXA1 locus visualization of linked-read (10X platform) whole genome-sequencing of the MDA-PCA-2B cell line. Alignments on the haplotype-resolved genome are shown in green and purple. Translocation and tandem-duplication calls are indicated in blue and red, respectively. **(B)** Monoallelic expression of FOXA1 cell-lines with FOXMIND-ETV1 translocations (MDA-PCA-2B and LNCaP). The phasing of FOXA1 SNPs to structural variants is based on linked-read sequencing (Methods).

Appendix 2

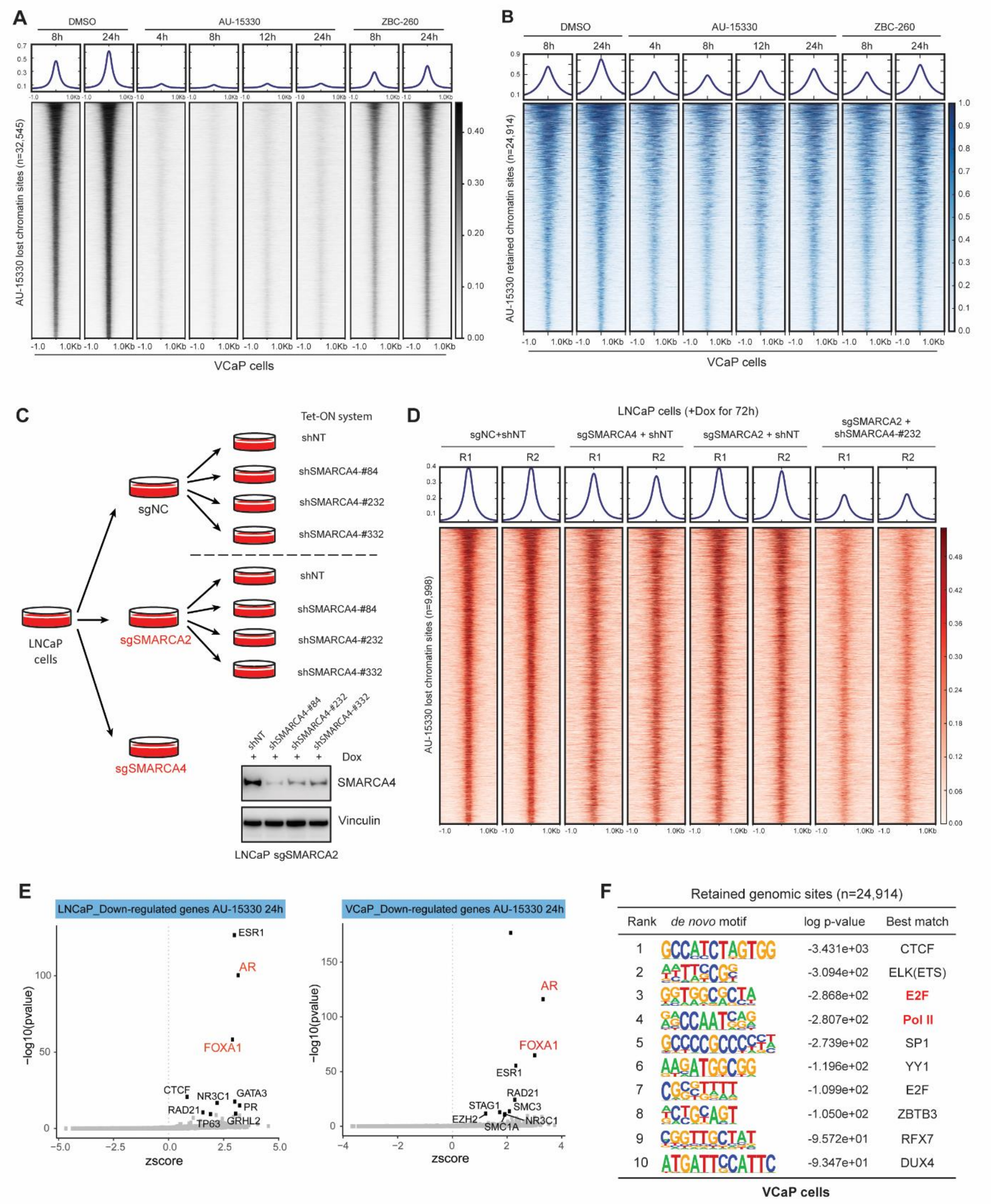
Supplementary Figures and Tables for Chapter 3



Supplementary Figure A2-1: Conformational model of AU-15330 target interaction and activity profile in diverse cell lines. (A) Docking model of AU15330 (cyan sticks) with the complex of SMARCA2 and VHL. AU15330 is suggested to fit into the pocket of SMARCA2 and VHL and capture several key interactions with the two proteins. Key hydrogen bond interactions with protein residues (pink sticks in SMARCA2, white sticks in VHL) are shown by yellow dashes. (B) Effects of AU-15330 at 1 μM for 4 hours on the proteome of VCaP cells. Data plotted Log2 of the fold change (FC) versus DMSO (dimethyl sulfoxide) control against –Log10 of the *p-value* per protein (FDR, false discovery rate) from n = 3 independent experiments. All t-tests performed were two-tailed t-tests assuming equal variances. TMT, tandem mass tag. (C) Dose-response curves of RWPE, LNCaP, VCaP, 22RV1, PC3, and DU145 cells treated with AU-15330 and dead analog. Data represent mean ± SE (n = 6) from one of three independent experiments. (D) Crystal violet staining showing the effect of AU-15330 on colony formation in VCaP, LNCaP, and PC3 cells. (E) Representative immunohistochemistry images showing the expression of ER, PR, HER2, GATA3, and AR in patient-derived breast cancer cell lines. (F) Immunoblots of noted proteins in WA-72-P or WA-72-As breast cancer cells treated with DMSO or AU-15330 at noted concentrations for 24 h. Vinculin serves as a loading control.


A
AU-15330 (1 μM)
0 1 2 4 8 12 24 Time (h)
SMARCA4
SMARCA2
PBRM1
Vinculin
RWPE
LNCaP
VCaP
B
RWPE LNCaP VCaP 22RV1 LAPC4
0 0.1 1.0 AU-15330 (μM)
SMARCA2
SMARCA4
PBRM1
Vinculin
VHL
C
24h 48h 72h
- + AU-15330 (100 nM)
SMARCA4
SMARCA2
PBRM1
c-PARP
Vinculin
LNCaP
VCaP
D
0 0.2 1 AU-15330 (μM)
0 1 10 100 VHL ligand (μM)
SMARCA2
SMARCA4
PBRM1
Vinculin
LNCaP
VCaP
E
VCaP cells
DMSO
AU-15330 33 nM
AU-15330 100 nM
% confluence (Incucyte)
0 48 96 144 192 240
LNCaP cells
Time (h)
Free VHL ligand (uM)
DMSO
3.7
11
33
100
F
LNCaP cells
sgNC
sgSMARCA2
sgSMARCA4
SMARCA2
SMARCA4
PBRM1
Vinculin
LNCaP
LNCaP cells
Fold change (relative to control)
Time (h)
0 24 48 72 96 120
sgNT
sgSMARCA4
sgSMARCA2
sgSMARCA2+ sgSMARCA4

Supplementary Figure A2-2: Verification of PROTAC design of AU-15330 and confirmation of on-target growth effects. (A) Western blot analysis showing the time-dependent effect of AU-15330 on SMARCA2, SMARCA4, and PBRM1 in RWPE, LNCaP, and VCaP cells. Vinculin serves as a loading control in all blots of this figure. (B) Immunoblots for indicated proteins in normal (RWPE) or prostate cancer cells (LNCaP, VCaP, 22RV1, and LAPC4) treated with AU-15330 at varied concentrations. (C) Immunoblots in LNCaP and VCaP cells examining time-dependent cleavage of PARP upon AU-15330 treatment. (D) Immunoblots showing the rescue effect of the free VHL ligand on AU-15330-mediated target protein degradation in LNCaP and VCaP cells. (E) Real-time measure showing the rescue effect of VHL ligand on AU-15330-mediated growth inhibition in VCaP and LNCaP cells. (F) Schematic showing the construction of polyclonal SMARCA2 and SMARCA4 knock-out LNCaP cell lines by using CRISPR/Cas9 technology. Growth curves for each cell line are plotted to the right.



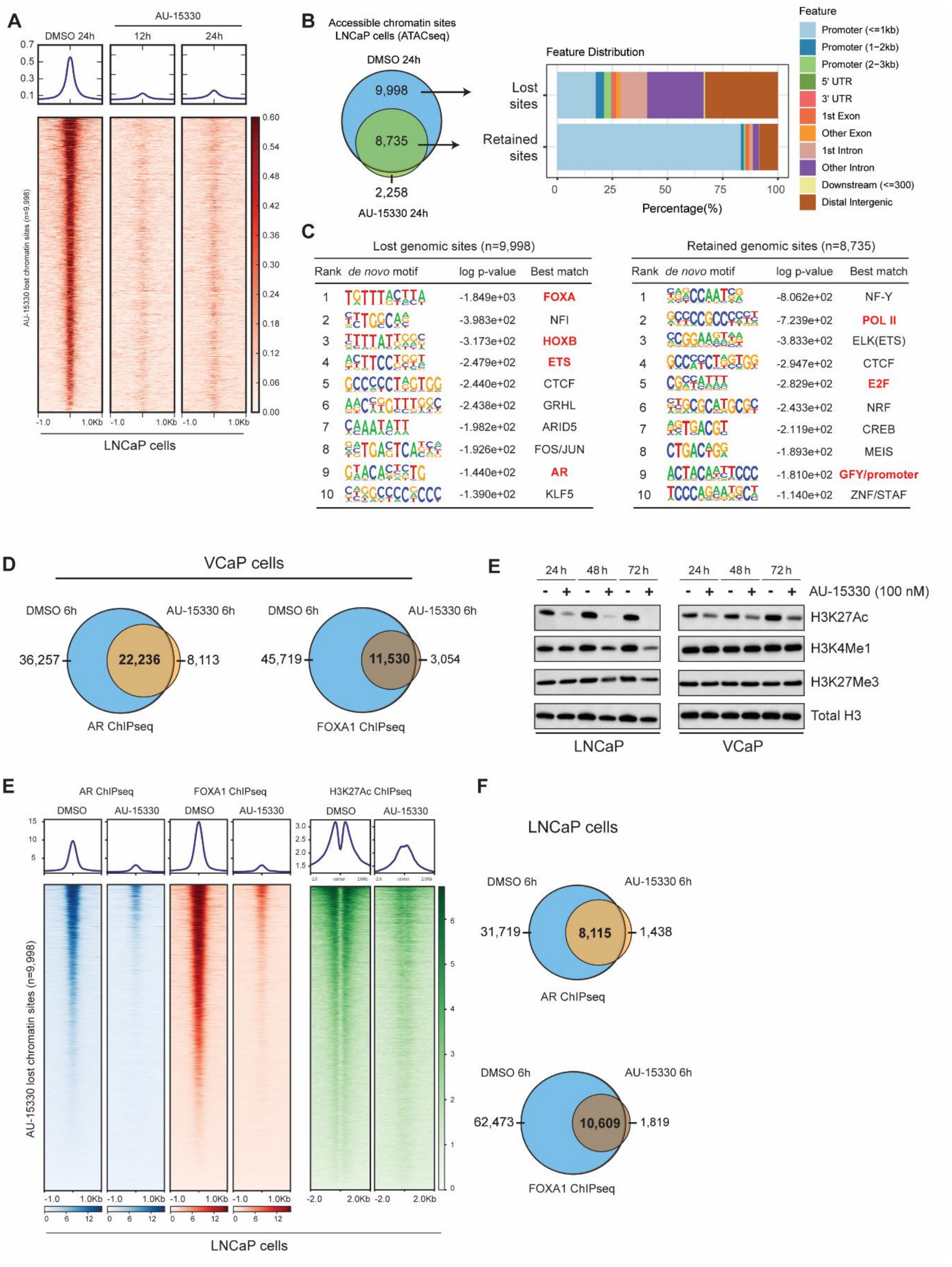
Retained genomic sites (n=24,914)

Rank	*de novo* motif	log p-value	Best match
1	GCCATCTAGTGG	-3.431e+03	CTCF
2	AATTCCGG	-3.094e+02	ELK(ETS)
3	GGTGGCGCTA	-2.868e+02	E2F
4	GACCAATCAG	-2.807e+02	Pol II
5	GCCCCGCCCCCT	-2.739e+02	SP1
6	AAGATGGCGG	-1.196e+02	YY1
7	CGCGTTTT	-1.099e+02	E2F
8	ACTGCAGT	-1.050e+02	ZBTB3
9	CGGTTGCTAT	-9.572e+01	RFX7
10	ATGATTCCATTC	-9.347e+01	DUX4

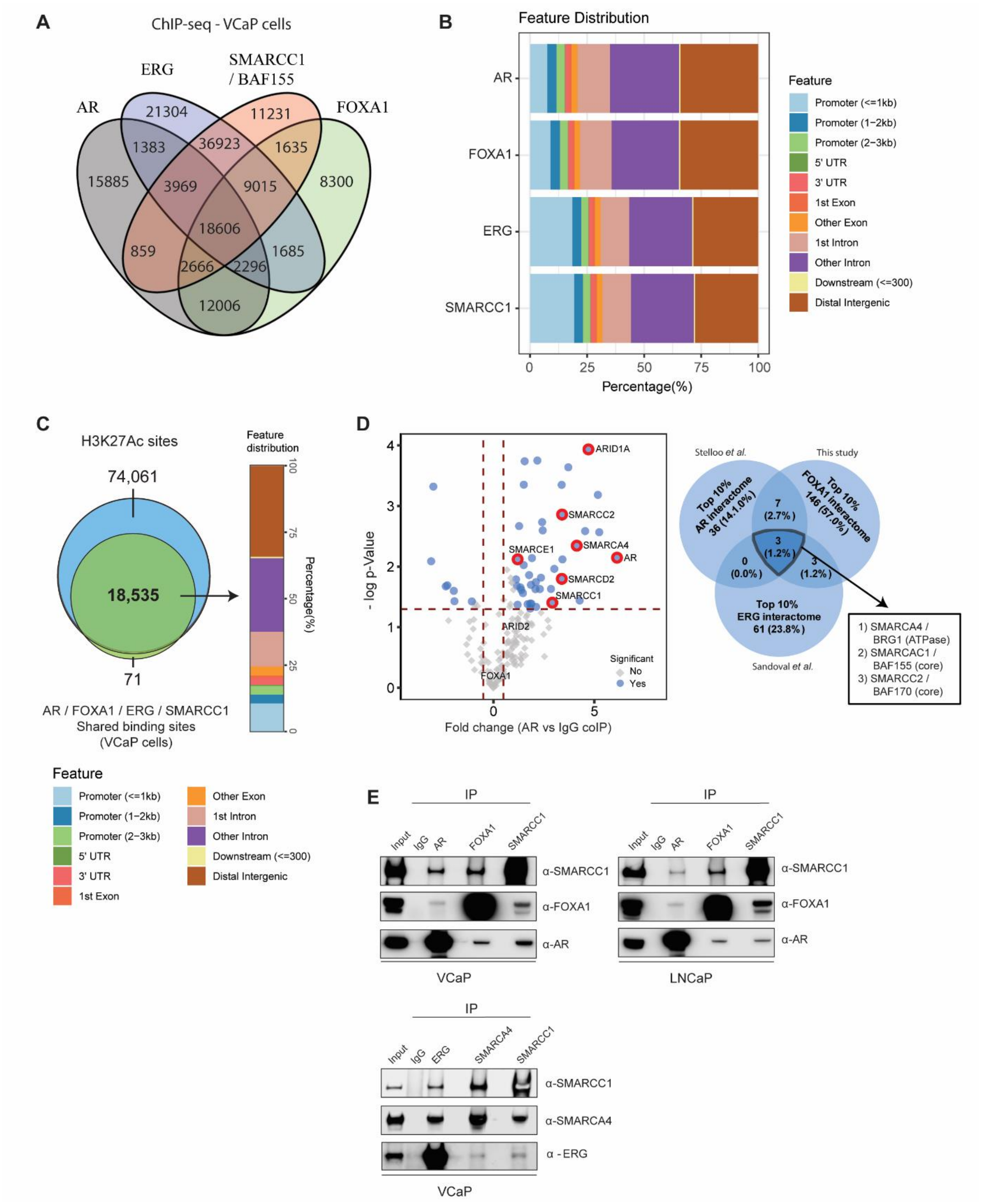
VCaP cells

Supplementary Figure A2-3: SWI/SNF ATPases, SMARCA2 and SMARCA4, mediate chromatin accessibility at numerous sites across the genome in prostate cancer cells. (A, B) ATAC-seq read-density heat maps from VCaP cells treated with DMSO (solvent control), AU-15330, or ZBC-260 (a BRD4 degrader) for indicated durations at genomic sites that are compacted (A) or remain unaltered (B) upon AU-15330 treatment. (C) Schematic outlining the CRISPR/Cas9 and shRNA-based generation of LNCaP cells with either independent or simultaneous inactivation of SWI/SNF ATPases, SMARCA2 and SMARCA4, along with immunoblots showing the decrease in target expression.

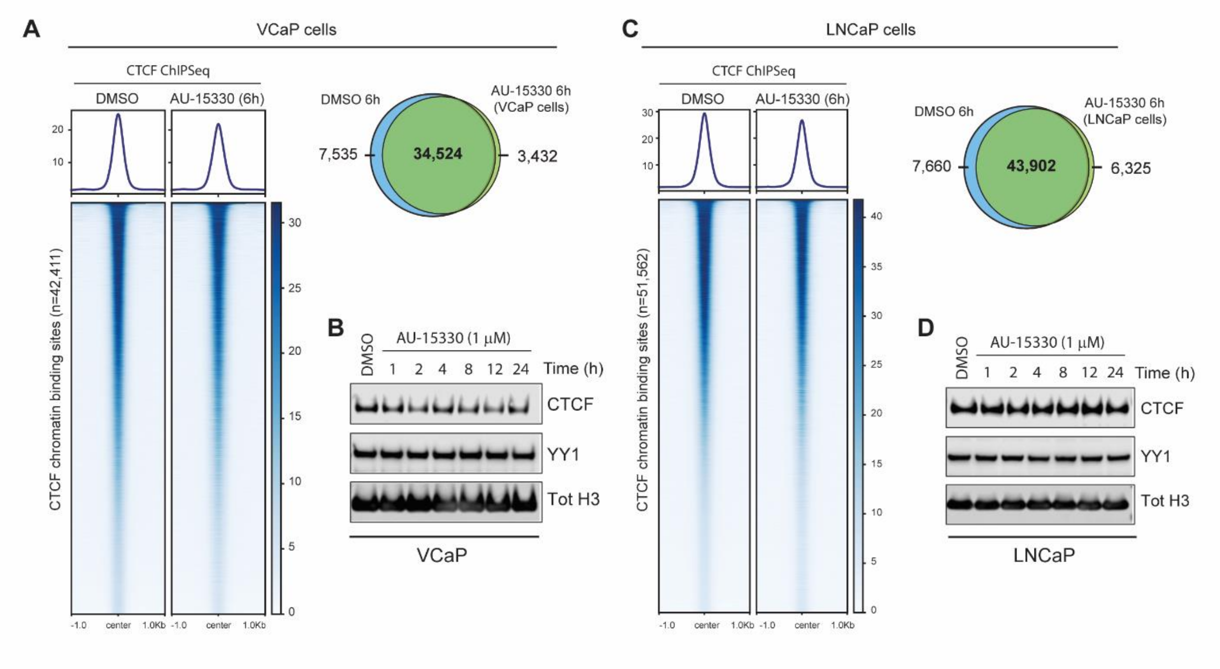
(D) ATAC-seq read-density heat maps from genetically-engineered LNCaP cells with SMARCA2 and/or SMARCA4 functional inactivation at AU-15330-compacted genomic sites. (E) Binding analysis for the regulation of transcription (BART) prediction of specific transcription factors mediating the observed transcriptional changes upon AU-15330 treatment in LNCaP or VCaP cells. The top 10 significant and strong (*z*-score) mediators of transcriptional responses are labeled (BART, Wilcoxon rank-sum test). (F) *De novo* motif analysis with top 10 motifs (ranked by p-value) enriched within genomic sites that retain chromatin accessibility upon AU-15330 treatment in VCaP cells (HOMER hypergeometric test).



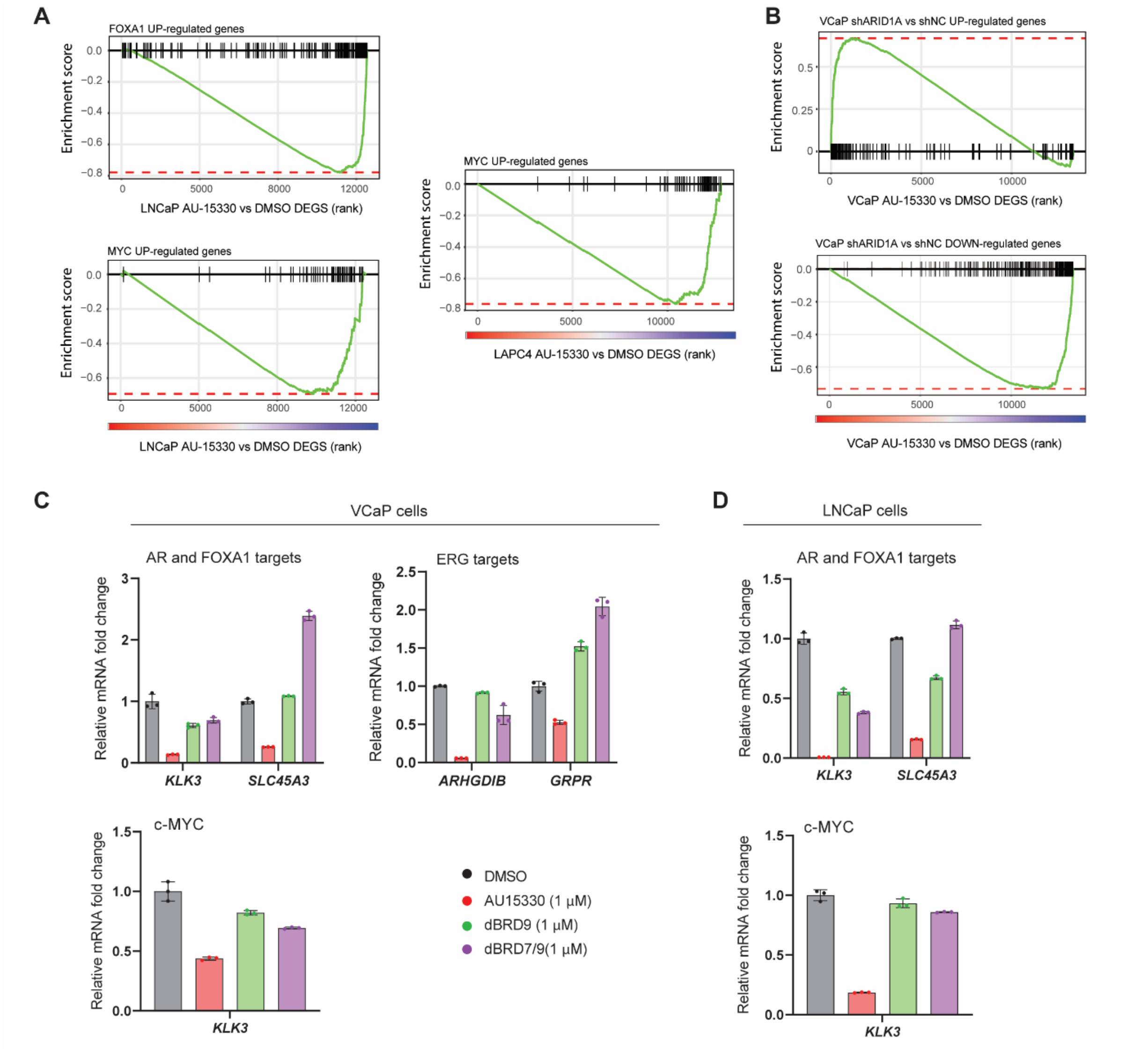
Supplementary Figure A2-4: SWI/SNF inhibition condenses chromatin at enhancer sites bound by oncogenic transcription factors AR and FOXA1 in prostate cancer cells. (A) ATAC-seq read-density heat maps from LNCaP cells treated with DMSO (solvent control) or AU-15330 for indicated durations at all genomic sites that lose physical accessibility upon AU-15330 treatment. (B) Genome-wide changes in chromatin accessibility upon AU-15330 treatment for 12 h in LNCaP cells, along with genomic annotation of sites that are lost or retained in the AU-15330-treated cells. (C) *De novo* motif analysis with top 10 motifs (ranked by p-value) enriched within AU-15330-compacted or unaltered genomic sites in LNCaP cells (HOMER, hypergeometric test). (D) Genome-wide changes in androgen receptor (AR) and forkhead box A1 (FOXA1) ChIP-seq peaks upon AU-15330 treatment (at 1 μM for 6 h) in VCaP cells stimulated with R1881, a synthetic androgen (at 1 nM for 3 h). (E) Immunoblots showing the changes in indicated chemical histone marks upon treatment with AU-15330. (F) ChIP-seq read-density heat maps for AR, FOXA1, and H3K27Ac at the compacted genomic sites in LNCaP cells after indicated durations of treatment with AU-15330 (at 1 μM). (G) Genome-wide changes in AR and FOXA1 ChIP-seq peaks upon AU-15330 treatment (at 1 μM for 6 h) in LNCaP cells stimulated with R1881 (at 1 nM for 3 h).



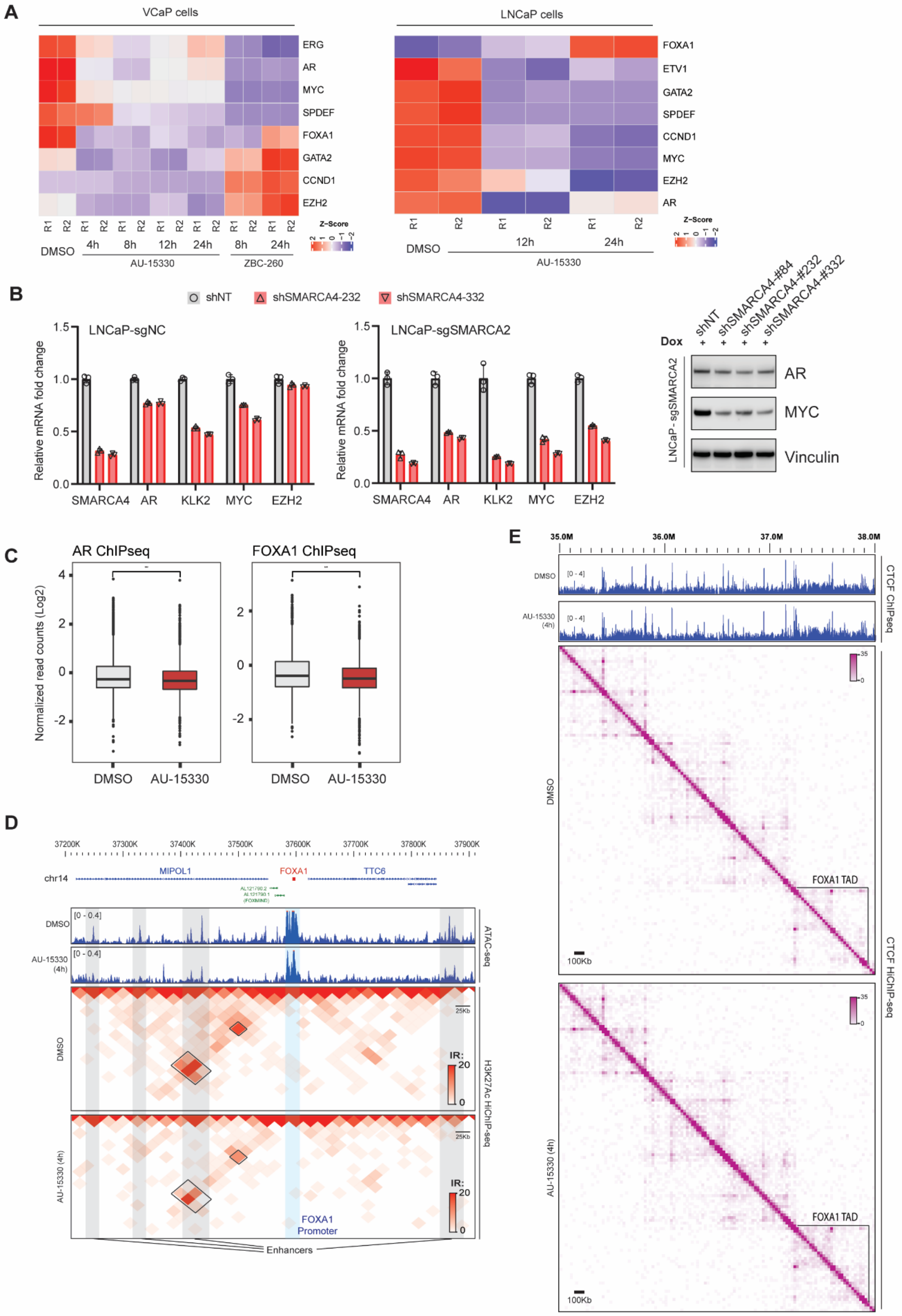
Supplementary Figure A2-5: The SWI/SNF complex is a common chromatin cofactor of the central transcriptional machinery in prostate cancer cells. (A) The overlap between AR, FOXA1, ERG (*30*), and SMARCC1 (*30*) ChIP-seq peaks in VCaP cells. (B) Genomic annotation of oncogenic transcription factor and SWI/SNF (SMARCC1) chromatin binding sites. (C) The overlap between transcription factor and SWI/SNF complex shared genomic sites (from A) and H3K27Ac ChIP-seq peaks along with the genomic annotations of the shared binding sites. (D) *Left:* volcano plot showing the AR interacting proteins identified from AR immunoprecipitation followed by mass spectrometry (*31*). Significantly enriched SWI/SNF subunits are highlighted in red. *Right:* Overlap between AR, FOXA1, and ERG interacting proteins identified from in-house or publicly-available datasets (*31*). (E) Immunoblots for indicated proteins followed by nuclear co-immunoprecipitation (IP) of AR, FOXA1, ERG, or SMARCC1 (a core SWI/SNF subunit) in VCaP and LNCaP cells after dihydrotestosterone (DHT) stimulation (at 10 nM for 3 h).



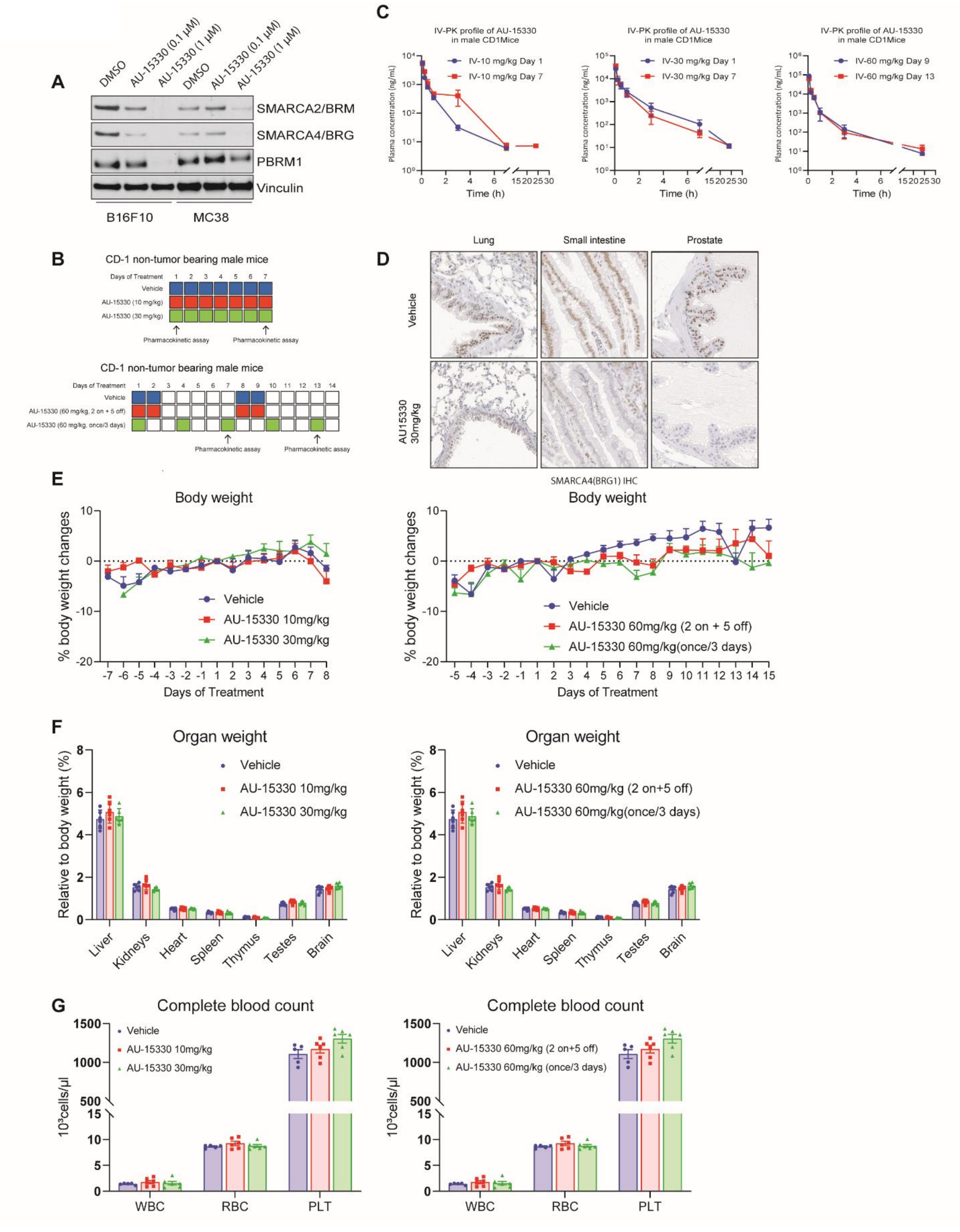
Supplementary Figure A2-6: AU-15330 does not affect genome-wide occupancy of CTCF. (A, C) Genome-wide ChIP-seq read-density heat maps and Venn diagrams for CTCF in VCaP (A) or LNCaP (C) cells treated with either DMSO or AU-15330 (1 μM) for 6 h. (B, D) Immunoblots of indicated proteins in VCaP (B) or LNCaP (D) cells treated with AU-15330 (1 μM) for increasing time durations.



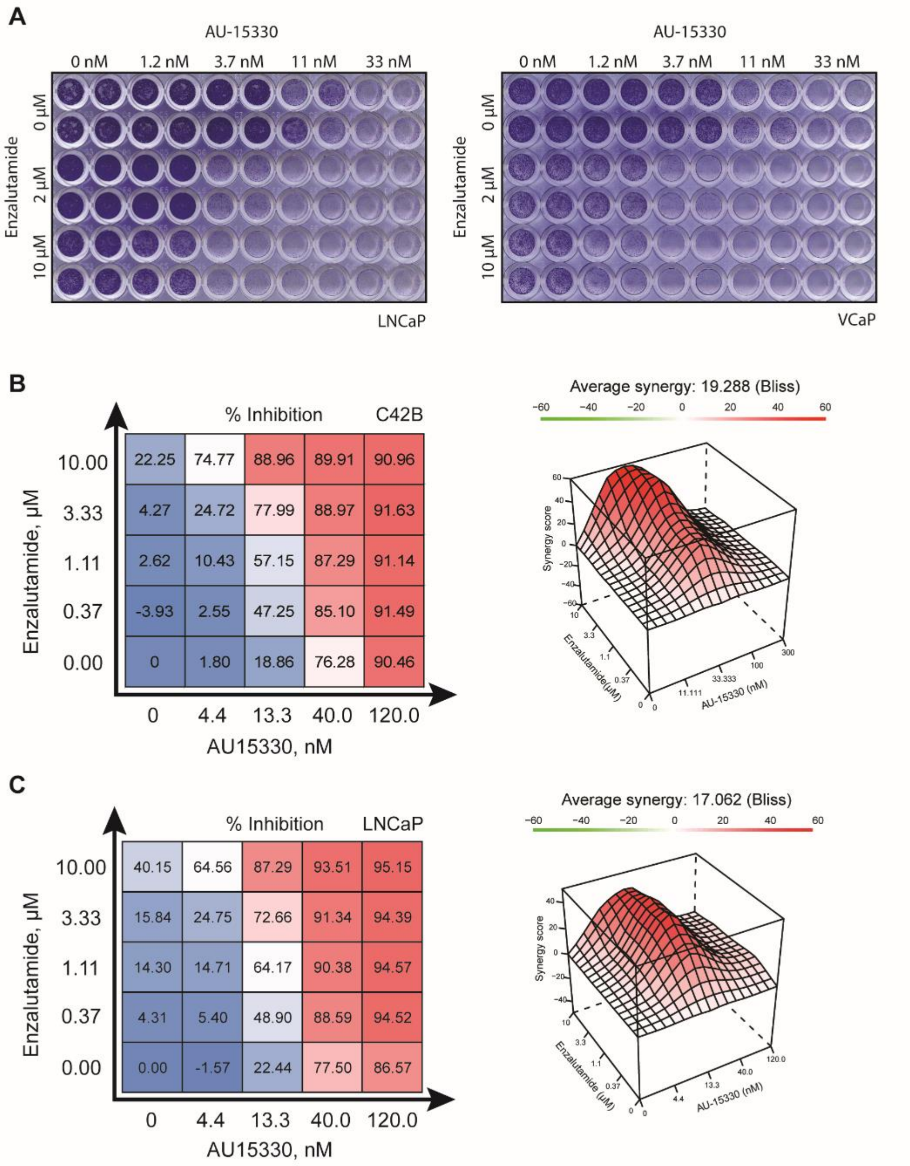
Supplementary Figure A2-7: The classical SWI/SNF complex is the primary cofactor of enhancer-binding transcription factors and is essential for enabling their oncogenic gene programs. (A, B) Gene set enrichment analysis (GSEA) of FOXA1, MYC, or ARID1A (B, data from (*30*)) regulated genes (see Methods for gene sets) in the fold change rank-ordered AU-15330 gene signature in indicated prostate cancer cells. DEGs, differentially expressed genes. (C, D) Expression of indicated genes (qPCR) in VCaP (C) or LNCaP (D) cells upon treatment with DMSO (control solvent), AU-15330, dBRD7 (BRD7 degrader), or dBRD7/9 (dual BRD7 and BRD9 degrader) at 1 µM for 24 h.


A
VCaP cells
ERG
AR
MYC
SPDEF
FOXA1
GATA2
CCND1
EZH2
R1 R2 R1 R2 R1 R2 R1 R2 R1 R2 R1 R2 R1 R2
DMSO
4h
8h
12h
24h
8h
24h
AU-15330
ZBC-260
Z-Score
2 1 0 −1 −2
LNCaP cells
FOXA1
ETV1
GATA2
SPDEF
CCND1
MYC
EZH2
AR
R1 R2 R1 R2 R1 R2
DMSO
12h
24h
AU-15330
Z-Score
2 1 0 −1 −2
B
shNT
shSMARCA4-232
shSMARCA4-332
LNCaP-sgNC
LNCaP-sgSMARCA2
Relative mRNA fold change
0.0
0.5
1.0
1.5
SMARCA4
AR
KLK2
MYC
EZH2
shNT
shSMARCA4-#84
shSMARCA4-#232
shSMARCA4-#332
Dox
+ + + +
LNCaP - sgSMARCA2
AR
MYC
Vinculin
C
AR ChIPseq
FOXA1 ChIPseq
Normalized read counts (Log2)
4
2
0
-2
DMSO
AU-15330
D
37200K
37300K
37400K
37500K
37600K
37700K
37800K
37900K
chr14
MIPOL1
FOXA1
TTC6
AL121790.2
AL121790.1
(FOXMIND)
DMSO
AU-15330 (4h)
[0 - 0.4]
ATAC-seq
25Kb
IR:
20
0
H3K27Ac HiChIP-seq
FOXA1 Promoter
Enhancers
E
35.0M
36.0M
37.0M
38.0M
DMSO
AU-15330 (4h)
[0 - 4]
CTCF ChIPseq
35
0
FOXA1 TAD
100Kb
CTCF HiChIP-seq

Supplementary Figure A2-8: SWI/SNF inhibition down-regulates the expression of oncogenic drivers through disruption of promoter and super-enhancer interactions. (A) RNA expression (RNA-seq) heat maps from VCaP or LNCaP cells treated with DMSO, AU-15330 (1 μM), or ZBC-260 (BRD4 degrader) for the noted durations. (B) RNA expression (qPCR) of indicated genes in stable CRISPR-engineered LNCaP-sgNC (control) or LNCaP-sgSMARCA2 (have SMARCA2 inactivation) cells that were treated with a non-target control shRNA or two distinct shRNAs targeting the SMARCA4 gene. *Right*, immunoblots showing expression of the indicated protein in CRISPR/shRNA-engineered LNCaP cells. (C) Normalized read density of AR or FOXA1 ChIP-seq signal at the super-enhancer sites in VCaP cells treated with DMSO or AU-15330 (1 μM) for 6 h. (D) H3K27Ac HiChIP-seq heat maps within the *FOXA1* gene locus in VCaP cells plus/minus treatment with AU-15330 (1 μM) for 4 h (bin size = 25Kb). ATAC-seq read-density tracks from the same treatment conditions are overlaid. Grey highlights mark enhancers, while the blue highlight marks the *FOXA1* promoter. (E) CTCF HiChIP-seq heat maps in a gene locus at Chr14, including the *FOXA1* topologically associating domain (TAD), in VCaP cells plus/minus treatment with AU-15330 (1 μM) for 4 h (bin size = 100Kb). CTCF ChIP-seq read-density tracks from VCaP cells plus/minus AU-15330 treatment (1 μM) for 6 h are overlaid.



Supplementary Figure A2-9: AU-15330 is well tolerated in mice and induces on-target degradation of SMARCA2, SMARCA4, and PBRM1. (A) Immunoblots of indicated proteins in B16F10 and MC38 cells treated with DMSO or AU-15330 (at 100 nM or 1 μM). Vinculin serves as a loading control. (B) Schematic outlining the design of AU-15330 *in vivo* study by using non-tumor bearing CD-1 mice. Male mice were treated with vehicle (control) or AU-15330 at the indicated concentration throughout the treatment period. (C) Immunohistochemistry staining of SMARCA4/BRG1 was carried out using lung, small intestine, and prostate sections after necropsy to show on-target efficacy of AU-15330 *in vivo*. (D) Pharmacokinetics profile of AU-15330 following intravenous (IV) injection in CD-1 mice. Mice received a single injection at indicated concentration of AU-15330, and plasma levels were determined by HPLC. (E) Animal body weight measurements showing AU-15330 does not affect animal weight. Non-tumor bearing CD-1 mice were treated with vehicle or AU-15330 at the indicated concentration throughout the treatment period. (F) Major organ weight measurements showing AU-15330 does not affect animal weight. Non-tumor bearing CD-1 mice were treated with vehicle or AU-15330 at the indicated concentration throughout the treatment period; the weights of liver, kidney, heart, spleen, thymus, testis, and brain were measured after necropsy. (G) Complete blood count showing AU-15330 does not affect the hematologic system. Non-tumor bearing CD-1 mice were treated with vehicle or AU-15330 at the indicated concentration throughout the treatment period, and whole blood was then collected and processed. WBC, white blood cells; RBC, red blood cells; PLT, platelets.



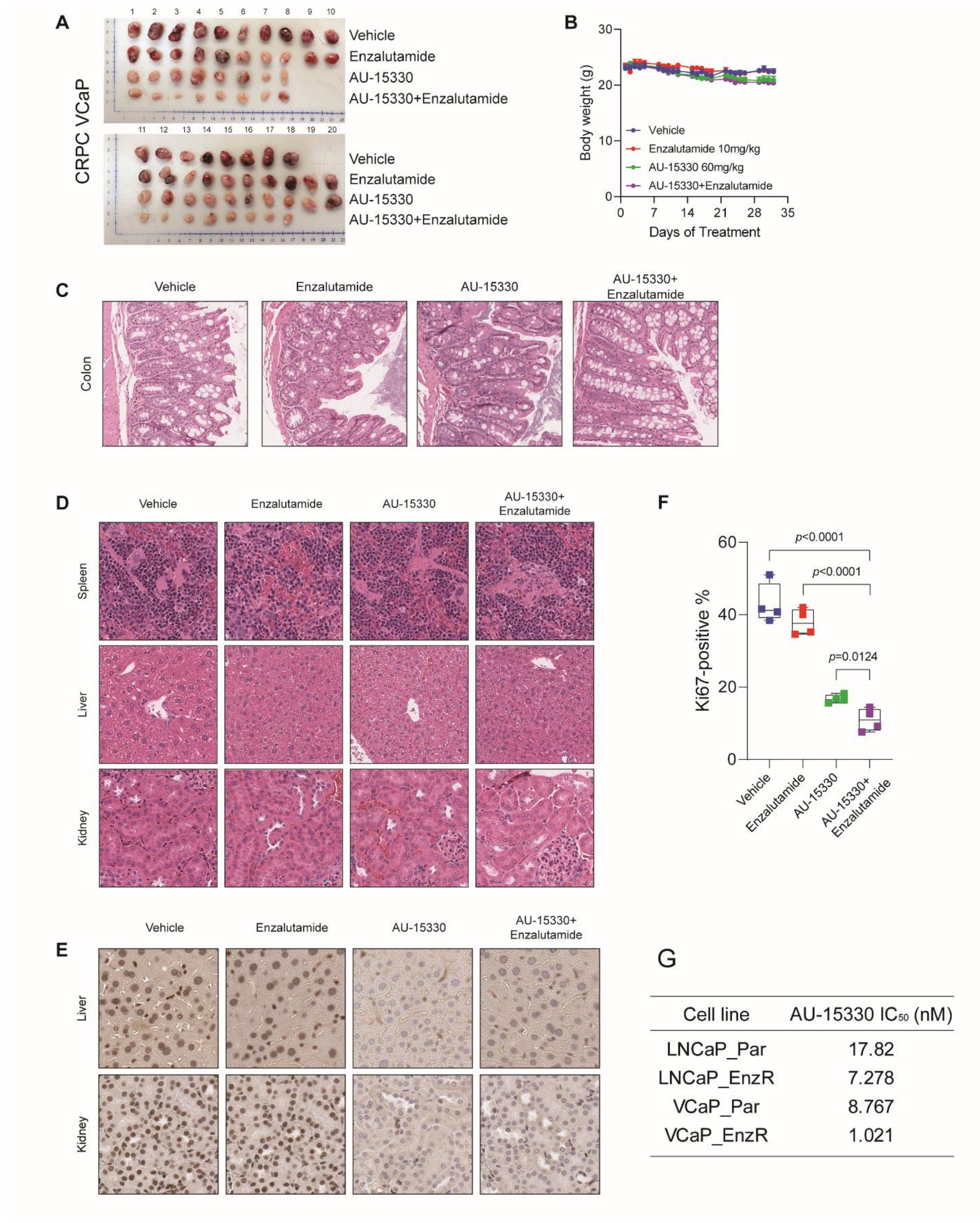
B — % Inhibition, C42B

Enzalutamide, µM	0	4.4	13.3	40.0	120.0
10.00	22.25	74.77	88.96	89.91	90.96
3.33	4.27	24.72	77.99	88.97	91.63
1.11	2.62	10.43	57.15	87.29	91.14
0.37	-3.93	2.55	47.25	85.10	91.49
0.00	0	1.80	18.86	76.28	90.46

C — % Inhibition, LNCaP

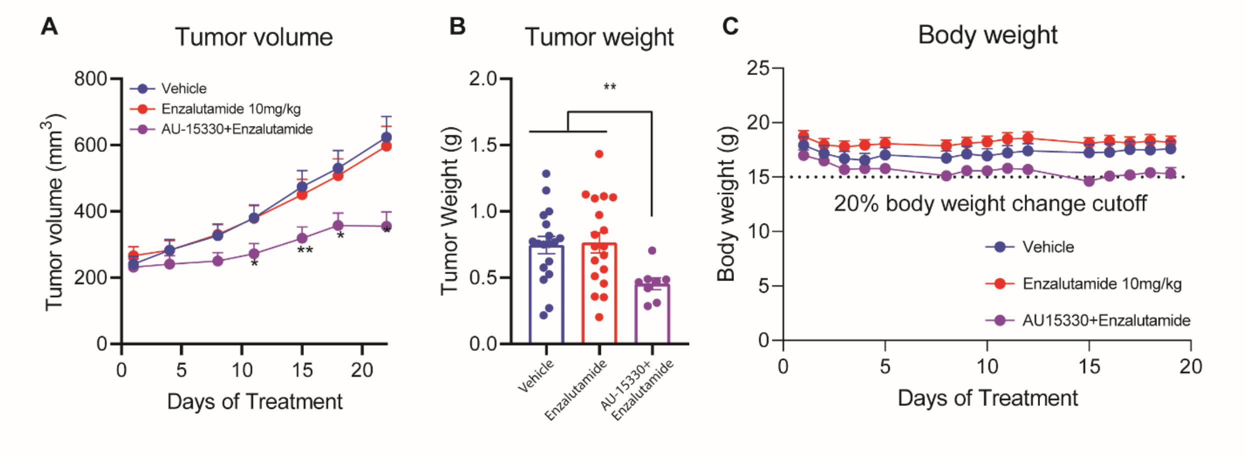
Enzalutamide, µM	0	4.4	13.3	40.0	120.0
10.00	40.15	64.56	87.29	93.51	95.15
3.33	15.84	24.75	72.66	91.34	94.39
1.11	14.30	14.71	64.17	90.38	94.57
0.37	4.31	5.40	48.90	88.59	94.52
0.00	0.00	-1.57	22.44	77.50	86.57

Supplementary Figure A2-10: SWI/SNF inhibition synergizes with the AR antagonist enzalutamide and decreases prostate cancer cell growth. (A) Crystal Violet staining showing the synergistic effect of AU-15330 and enzalutamide on colony formation in VCaP and LNCaP. (B, C) LNCaP and C4-2B cells were treated with AU-15330 and/or enzalutamide at varied concentrations to determine the effect on cancer cell growth and drug synergism assessments were carried out using the Bliss Independence method. Red peaks in the 3D plots denote synergy with the average synergy scores noted above. The mean of three biological replicates is shown on top.



Cell line	AU-15330 IC$_{50}$ (nM)
LNCaP_Par	17.82
LNCaP_EnzR	7.278
VCaP_Par	8.767
VCaP_EnzR	1.021

Supplementary Figure A2-11: Combined *in vivo* treatment with AU-15330 and enzalutamide causes tumor regression in prostate cancer xenografts without toxic effects on other organs. (A) Individual tumors from vehicle, enzalutamide, AU-15330, and AU-15330+enzalutamide groups from VCaP-CRPC (castration-resistant prostate cancer) study. (B) Animal body weight measurement showing the effect of vehicle, enzalutamide, AU-15330, and combination of AU-15330 and enzalutamide on animal weight. Tumor-bearing SCID mice were treated with vehicle, enzalutamide, AU-15330, or a combination of AU-15330 and enzalutamide at the indicated concentration throughout the treatment period, and the bodyweight was measured at the end-point. (C, D) H&E staining was carried out to examine the effect of AU-15330 *in vivo* using colon, spleen, liver, and kidney sections after necropsy. Representative images of H&E staining are shown. (E) Immunohistochemistry staining of SMARCA4/BRG1 was carried out using liver and kidney sections after necropsy to show on-target efficacy of AU-15330 *in vivo*. (F) Box plot of the percent cells with positive Ki-67 staining. Paired t-test shows significant differences between vehicle, enzalutamide, AU-15330, and a combination of AU-15330 and enzalutamide. (G) IC_{50} for AU-15330 in enzalutamide-resistant (EnzR) LNCaP and VCaP cell lines after 5 days of treatment.



Supplementary Figure A2-12: AU-15330 inhibits tumor growth of an enzalutamide-resistant patient-derived xenograft model. (A) Tumor volume measurements showing the effect of enzalutamide alone or in combination with AU-15330 in the enzalutamide-resistant MDA-PCa-146-12 patient-derived xenograft model. Caliper measurements were taken twice a week. Mean tumor volume ± s.e.m. is shown. (B) Individual tumor weights from different treatment groups with *p*-values are shown. (C) Animal body weight measurements showing the effect of vehicle, enzalutamide, and combination of AU-15330 and enzalutamide on animal weight. Tumor-bearing SCID mice were treated with vehicle, enzalutamide, or a combination of AU-15330 and enzalutamide at the indicated concentration throughout the treatment period, and the body weight was measured at the end-point.

Cell line	IC$_{50}$ (μM)	Primary Site
VCaP	0.9862	Prostate
LNCaP	2.299	Prostate
MM.1S	2.364	Multiple Myeloma
Karpas-25	2.5	Multiple Myeloma
U-937	4.983	Lymphoma
MDA Pca 2b	12.51	Prostate
WA-72-P	14.37	Breast
CHLA-10	18.18	Ewing sarcoma
DoHH2	20.45	Lymphoma
22Rv1	21.25	Prostate
WA-72-As	22.73	Breast
C4-2B	26.54	Prostate
HCC1428	28.85	Breast
LAPC-4	47.47	Prostate
SK-MEL-5	63.55	Melanoma
ZR-75-1	70.26	Breast
Capan-2	72.45	Pancreatic
HCC1500	82.6	Breast
CWR-R1	93.98	Prostate
LNCaP-AR	97.04	Prostate
BxPc-3	111.7	Pancreatic
T-47D	124.4	Breast
PC-3	216.9	Prostate
Karpas-417	244.5	Multiple Myeloma
DU 145	363.5	Prostate
RPMI 8226	432.3	Multiple Myeloma
MDA-MB-361	471.7	Breast
U266B1	599.3	Multiple Myeloma
WSU-NHL	895.5	Lymphoma
NCI-H838	1017	Lung
MDA-MB-468	1052	Breast
NCI-H716	1884	Colon
MDA-MB-231	2233	Breast
LHSR-AR	11074	Prostate
SW-1990	17608	Pancreatic
NCI-H1299	18490	Lung

MCF7	**22233**	**Breast**
UM-UC-3	**22247**	**Bladder**
COLO 205	**30152**	**Colon**
RWPE-1	**30590**	**Prostate**
A4573	**47613**	**Ewing sarcoma**
OVCAR-3	**334765**	**Ovary**
A-673	**9128754**	**Ewing sarcoma**
TC-71	**26624899**	**Ewing sarcoma**
MCF 10A	**30573722**	**Breast**
PNT2	**64951468**	**Prostate**
Panc 10.05	**75632762**	**Pancreatic**
SW-780	**80667650**	**Bladder**
MDA-MB-134	**84475928**	**Breast**
SJ-GBM2	**90703502**	**Glioblastoma**
BPH-1	**94377499**	**Prostate**
A549	**99819380**	**Lung**
Hela	**103058423**	**Cervical**
SK-MEL-2	**137745440**	**Melanoma**
HeP G2	**140435081**	**Liver**
OVCAR-5	**157477902**	**Ovary**
SCaBER	**163118968**	**Bladder**
MIA PaCa-2	**182995294**	**Pancreatic**
SK-MEL-28	**391932641**	**Melanoma**
HCT 116	**424506487**	**Colon**
DLD-1	**485611015**	**Colon**
HEK-293	**590677208**	**Kidney**
NCI-H460	**609727136**	**Lung**
COLO 320DM	**23599684364**	**Colon**

Supplementary Table 1: Anti-proliferative half-maximal inhibitory concentrations (IC50) of AU-15330 in human-derived cell lines. The panel includes both normal and cancer cell lines from 14 distinct lineages.

Antibodies:

Antigene	**Vendor**	**Catalog number**	**Application**	**Note**
BAF155	Cell Signaling Technology	11956S	Western Blot, Co-IP	1:1000
SMARCA2/BRM	Bethyl laboratories	A301-016A	Western Blot	1:1000
SMARCA4/BRG 1	Cell Signaling Technoly	52251S	Western Blot	1:1000
PBRM1	Bethyl laboratories	A301-591A	Western Blot	1:1000
Vinculin	Millipore Sigma	V9131	Western Blot	1:5000
VHL	Thermo Fisher Scientific	PA527322	Western Blot	1:1000
AR	Millipore Sigma	06-680	Western Blot, Co-IP	1:1000
ERG	Abcam	ab92513	Western Blot, Co-IP	1:1000
FOXA1	Thermo Fisher Scientific	PA5-27157	Western Blot, Co-IP	1:1000
c-Myc	Cell Signaling Technology	5605S	Western Blot	1:1000
KLK3	DAKO	A0562	Western Blot	1.430555556
YY1	Diagenode	C15410345	Western Blot	1:1000
MED1	Bethyl laboratories	A300-793A	Western Blot	1:1000
H3K27Me3	Diagenode	C15410069	Western Blot	1:1000
H3K27Ac	Cell Signaling Technology	8173	Western Blot	1:1000
H3K4me3	Cell Signaling Technology	9751	Western Blot	1:1000
H3K4Me1	Abcam	ab8895	Western Blot	1:1000
Cleaved PARP (Asp214)	Cell Signaling Technology	9541	Western Blot	1:1000
SMARCA2/BRM	Millipore sigma	HPA029981	IHC	1:100
SMARCA4/BRG 1	Abcam	108318	IHC	1:100
AR	Millipore Sigma	06-680	IHC	1:100
FOXA1	Thermo Fisher Scientific	PA5-27157	IHC	1:1000
ERG	Cell Signaling Technology	97249S	IHC	1:500
AR	Millipore/Sigma	06-680	ChIP-seq	10 μg/7-8M cells
ERG	Cell Signaling Technology	97249S	ChIP-seq	10 μg/7-8M cells

FOXA1	Thermo Fisher Scientific	PA5-27157	ChIP-seq	10 µg/7-8M cells
H3K27Ac	Abcam	ab4729	ChIP-seq	10 µg/10M cells
CTCF	Cell Signaling Technology	3418	HiChIP-seq	1.14 µg per IP
H3K4me3	Cell Signaling Technology	9751	HiChIP-seq	3.4 µg per IP
H3K27Ac	Cell Signaling Technology	8173	HiChIP-seq	0.4 µg per IP

Primers:

Name	**sequence**
AR_F	CAGTGGATGGGCTGAAAAAT
AR_R	GGAGCTTGGTGAGCTGGTAG
FOXA1_F	GAAGACTCCAGCCTCCTCAACTG
FOXA1_R	TGCCTTGAAGTCCAGCTTATGC
c-Myc_F	CGGAAGGACTATCCTGCTGC
c-Myc_R	CAAGACGTTGTGTGTTCGCC
KLK2_F	GTGTACAGTCATGGATGGGC
KLK2_R	CCCAGAATCACCCCCACAAG
KLK3_F	ACGCTGGACAGGGGGCAAAAG
KLK3_R	GGGCAGGGCACATGGTTCACT
BRG1_F	CACAGCCAAGGTTATATGTCACC
BRG1_R	TGCGGATCACCTGGGATGA
BRM_F	GACCAGCACTCCCAAGGTTAC
BRM_R	CTGGCCCGGAAGACATCTG
ACTB_F	AGGATGCAGAAGGAGATCACTG
ACTB_R	AGTACTTGCGCTCAGGAGGAG
PBRM1_F	AGGAGGAGACTTTCCAATCTTCC
PBRM1_R	CTTCGCTTTGGTGCCCTAATG

CRISPR/Cas9 sgRNAs:

Name	**Sequence**
sgNT_F	caccgACGTGGGGACATATACGTGT
sgNT_R	aaacACACGTATATGTCCCCACGTc
sgBRG1_1_F	aaacCATGGTAAGACTGGCTGCCC
sgBRG1_1_R	CACCGGGCAGCCAGTCTTACCATG
sgBRG1_2_F	aaacCATTTAACCAGAACCAGCTGC

sgBRG1_2_R	CACCGCAGCTGGTTCTGGTTAAATG
sgBRG1_3_F	aaacTCGGTGTCCGCAACAGgacc
sgBRG1_3_R	CACCggtcCTGTTGCGGACACCGA
sgBRM_1_F	CACCGCTGCAAGCTGCAGCGTTTCG
sgBRM_1_R	aaacCGAAACGCTGCAGCTTGCAGC
sgBRM_2_F	aaacACCAGATGGTCTGTTGTAGTC
sgBRM_2_R	CACCGACTACAACAGACCATCTGGT
sgBRM_3_F	aaacCGGGCCAGCATTTTATAAGC
sgBRM_3_R	CACCGCTTATAAAATGCTGGCCCG

Short hairpin RNAs:

Target	**Vendor**	**Catalog number**
sh_BRM_1	Dharmacon	V3SH11252-227095613
sh_BRM_2	Dharmacon	V3SH11252-227388488
sh_BRM_3	Dharmacon	V3SH11252-227492834
sh_BRG1_1	Dharmacon	V3SH11252-224944534
sh_BRG1_2	Dharmacon	V3SH11252-225289582
sh_BRG1_3	Dharmacon	V3SH11252-226665682

Compounds:

Name	**Vendor**	**Catalog number**
AU-15330	Aurigene	n/a
AU-15330 dead analog	Aurigene	n/a
ZBC-260	medchemexpress	HY-101519
ARV-825	Cayman Chemical Company	21109
VZ-185	Tocris / R&D Systems	6936
dBRD9	Tocris / R&D Systems	6606

Supplementary Table 2: Antibodies and reagents used in the work summarized in chapter 3. Information of antibodies, qPCR primers, CRISPR/Cas9 single guide RNAs (sgRNAs) sequences, short hairpin RNAs (shRNAs), and compounds that were used in the SWI/SNF degrader study.