Chimeric AR Motifs and Neo-Enhancer Expansion in Prostate Cancer Development

by

Sanjana Eyunni

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
2025

Doctoral Committee:

Professor Arul M. Chinnaiyan, Co-Chair
Assistant Professor Abhijit Parolia, Co-Chair
Assistant Professor Marcin Cieslik
Professor Costas Lyssiotis
Professor Jason R. Spence

Sanjana Eyunni

seyunni@umich.edu

ORCID iD:0009-0006-5101-6506

© Sanjana Eyunni 2025

Dedication

To my family and friends, for their unwavering support, love, and encouragement; and my scientific mentors, for their guidance and inspiration. Thank you for shaping who I am today.

Acknowledgements

As I sit at my desk, reflecting on the last five years, I am overwhelmed with gratitude for the journey that has brought me to this moment. Reaching July 29, 2025, marks not only a milestone in my academic path but also a testament to the unwavering love, support, and guidance of many remarkable people. This journey has never been mine alone, and it is a privilege to recognize and thank those who have walked beside me.

As quintessential as it may sound, my passion for Biology began early, thanks to an extraordinary woman named Ms. Kalpana Puri, my high school biology teacher. Her teachings inspired me to pursue a career in biology. However, back then, I did not understand what it meant to do research, let alone pursue a career in it. This, however, changed during my master's degree, where I got a preliminary glimpse into the world of a research laboratory. I worked in Dr. Ravanan Palaniyandi's lab, under the mentorship of Dr. Ahmed Abdullah, a PhD student at that time. I watched him perform exciting experiments – albeit from a distance. However, this was about to change.

In 2017, I was accepted as a Study Abroad Scholar at the University of Michigan (UoM) to pursue my master's thesis. Here, I joined the lab of Dr. Arul Chinnaiyan, who would later become my PhD advisor and significantly alter the trajectory of my scientific career. I arrived in Detroit on a cold December evening and was warmly received by Sunita Shankar and her husband, Chandan Sinha, who were the primary reason for my research stint at UoM. This was my first professional research experience, and I am profoundly grateful to Sunita, Lakshmi, Seema, and Abhijit for nurturing my scientific spirit. Six months later, I moved to Nebraska, where I made wonderful memories with Mansi, Shaista, and Sanchita. However, I was keen on coming back to UoM, and

hence applied to the Program in Biomedical Sciences, which I was accepted into in the fall of 2019.

Returning to Michigan as a PhD student felt like coming full circle, and being able to continue my journey in Arul's lab has been an honor. Arul is a visionary. I have been in awe of him and his work since I joined, and I am grateful to him for allowing me to be a part of his scientific empire. Arul has built a scientific institute with some of the most brilliant minds, and learning from them has been the joy of my PhD. His ideas, vision, and ability to narrate complex stories are lessons that I will forever hold close to me. It is a privilege to be associated with his name and his mentorship. Through Arul's lab, I also met other incredibly inspiring scientists who shaped the future of my PhD.

The most notable of them all is Dr. Abhijit Parolia. He has had the most significant influence on my scientific career, single-handedly. When I joined Arul's lab, Abhijit was a graduate student, and I worked under his tutelage on the FOXA1 project. Very soon, he would become my dear friend, mentor, and closest confidant in the lab. Working with Abhijit — first as a fellow graduate student and then as my official PhD mentor — has shaped me indelibly as a scientist. Abhijit's unwavering patience, his wisdom on diverse scientific topics, and his utmost dedication to science have been my constitution for a fruitful PhD. I joined the PhD program with very little knowledge, but within a year, he changed that for me. He has taught me to find joy in fundamental biology, the craft of careful experimentation, the importance of reading, and the resilience required in science. He often encouraged me in the righteous pursuit of a PhD — seeking knowledge and ever more knowledge. When my project lagged with few results for years, it was his encouragement and confidence in FOXA1 that kept me going. I am very grateful to have been one of the founding members of his lab and sincerely thank him for showing me what an incredible PhD journey could look like.

The Chinnaiyan and Parolia labs are filled with amazing scientists, colleagues, and friends. Caleb, Tongchen, Yihan, Hannah, and I created fond memories together as graduate students, and I am very grateful to them for being there as we navigated the PhD journey together. I was fortunate to learn from and work with many talented senior lab members, including Lanbo Xiao, Yuanyuan Qiao, Jie Luo, Eleanor Young, Jean Tien, and Yuping Zhang. My thesis involved extensive mouse prostate pathology, and for this, Dr. Rahul Mannan and his team deserve full credit. Rahul is one of the kindest people I have met, and his energy fills every room he enters. He was gracious with his time and patiently analyzed over a hundred histology slides with me. Through the co-mentorship between two labs, I am fortunate to have two scientific families. In PARLAB, I am grateful to Qiuyang, Michaela, Calvin, and Miloni for all their contributions to my project, as well as for being wonderful people to spend my day with in the lab. During the last five years, mentoring students in the lab has been one of the most fulfilling aspects of my PhD. My mentoring journey began with Mustapha Jaber and James George, who helped us conduct numerous critical experiments for the FOXA1 project. They were such kind and hardworking students, and I am very grateful to them for letting me mentor them. Watching them succeed in their medical careers is heartwarming, and every time they visit us, it always brings back fond memories. I am grateful to Qiuyang Zhang, Hamzah Hakkani, and Matthew Pang, who have been pillars of the FOXA1 project. Qiuyang has been a tour de force in our group, taking the lead on various techniques in the lab, including maintaining our animal inventory, organoid work, and numerous immunostainings. Matt joined our group as an undergraduate student, but his dedication and perseverance in learning and contributing have been reflected in the numerous hours he spends in the lab, despite a busy school schedule. Hamzah joined the group as a pre-med student and had a significant impact on it in a short period of time. His tireless work ethic, his approach to life, and his positive attitude made him a lab favorite. Teaching them allowed me to evolve as a mentor and made me realize the power of a team. As Phil Jackson famously once said, “The strength of the team is each individual member. The strength of each member is the team”.

None of the work we do would be possible without the amazing administrative support in the center. I would first like to thank Xuhong Cao, our lab manager, without whom this massive operation would cease to exist. She is a role model for what hard work and dedication look like and has gone out of her way numerous times to ensure that our projects are on track. Next, I would like to thank Hui for helping to coordinate my thesis meetings, Jyoti Athanikar for her efforts in submitting numerous support letters, Vicki Auken for assisting me with grant submissions, and Stephanie for proofreading and submitting our manuscripts. Next, I would like to acknowledge Fengyun Su and Rui Wang, who are the powerhouses of the sequencing team and have helped generate numerous libraries for our project. A special mention goes out to Melissa Dunn, Chia-Mei (Grace Huang), Parker Schanen, and Calvin Burns for ensuring we never run out of lab supplies. Lastly, I would like to thank all the other members of the Chinnaiyan and Parolia lab for making my time here so memorable.

I would also like to thank my thesis committee, Dr. Marcin Cieslik, Dr. Costas Lyssiotis, and Dr. Jason Spence, who took the time to provide valuable feedback on my projects and support letters for award and grant submissions. I would especially like to thank Dr. Sriram Venneti, who has been an unofficial mentor and has supported many of my award applications. Finally, I would like to thank the members of the Pathology department, Dr. Zaneta Nikolovska-Coleska, who interviewed me for my PhD admission; Laura Labut for always keeping me on track during my PhD program; Dr. Jeff Rual and Dr. Simon Hogan, directors of the Pathology program; and all the Pathology students. I would like to thank my PhD cohort mates, Agamjot, Brian, Alec, and Mohammed, with whom I spent numerous hours studying together and navigating the new life of being in a PhD program. During this time, I had meaningful conversations with Sahiti Marella, who became a dear friend over the years.

Reflecting on this journey, I would like to highlight one significant eureka moment in my project that will forever be etched in my memory. Up until July 19th, 2023, I had performed over 100 necropsies on mouse prostate tissues and saw no phenotypes. However, on July 19th, I was in the

microscope room, ready to see another set of slides with no phenotypes. I recall examining the anterior lobe of the mouse prostate and noticing something I had never seen before. Big nuclei, mitotic cells, and, in general, a tissue that looked abnormal to me. Thankfully, Dr. Rahul Mannan, our pathologist, was in the room at the time, and he immediately joined me to confirm my observations. I immediately called Abhijit to share the joyful news, and that day changed the course of my project. That moment made me realize the value of patience and perseverance in the scientific field.

Next, I would like to thank my friends, who have been an integral part of my life. Leaving India meant saying goodbye to cherished friends and making new ones on a different continent. I am grateful to my childhood friends—Dhathri, Shrenya, Vandana, and Sneha—whose friendship set the early standard for loyalty and support. My college years were brighter thanks to Rosheni, Vaishnavi, Sravya, and Aishu. Meghana, my soul-sister, has been a constant light, checking in on me through the busiest and loneliest days, even when I struggled to keep in touch. My enormous "framily" of friends from home, whose love knows no distance, has made me feel supported and cared for, ensuring my parents are surrounded by warmth even in my absence. Viraat, my childhood friend, has been a steadfast confidant; I miss our long talks and shared curiosity. Nidhi, with her enduring support (including organizing my bachelorette party in India), has been a constant presence. Vivek, my closest confidant and "triplet" in life and values, has been an instrumental pillar in my journey. Although conversations with him are sometimes rare, I always end them feeling better about myself and the world. I have had the privilege of witnessing his approach to life, his dedication to his work, and his kindness to people around him. He has given me and my family an immense amount of love for which I will be forever grateful. In Ann Arbor, I was fortunate to form new bonds—chief among them, with my friend-turned-mentor, Abhijit, and his wonderful wife, Rini. Rini always created a safe space during our many outings together, talking where we shared conversations about life, the immigrant experience, and so much more. Together with their new baby boy, Miraan, they have generously welcomed us into their lives, sharing celebrations and everyday joys.

Finally, I would like to extend my love to my family. My gratitude stretches across continents to my family in both India and the U.S. To my aunt, uncles, and cousins here in California and Atlanta, thank you for your support and care along the way. To my grandparents, Vijaya Lakshmi and Dr. RHG Rau—thank you for your unconditional love and for inspiring me, whether in the pursuit of knowledge or the art of gardening. My grandfather, who holds a PhD himself, has an unending desire to acquire knowledge, a testament to the fact that learning knows no age. He will often write me emails and ask me questions about my project, which I fondly treasure. I have always watched my grandmother care and love with so much intention, and I hope to imbibe that quality moving forward. My parents in India—Ranga and Sasi Eyunni—are the pillars of our house. They sent me abroad with encouragement, trusting me to follow a path unfamiliar to them. They built a home filled with love and warmth that remains my anchor to this day. They have supported me and my siblings in our distinct career choices, giving us the freedom to explore life while instilling a sense of purpose and responsibility in the choices we make. As I grow up, I reflect on the values they instilled in us and hope to be half as good a parent as they were to us. Time spent with them during their recent visit to Ann Arbor remains among my most cherished memories. Growing up in a house with my grandparents and parents has taught me what it means to love unconditionally and to give generously without expecting anything in return.

My siblings, Karan and Namrata, are my comfort and champions; their unwavering support and our unique bond are a reassuring source of strength for me. Karan, my older brother, is our protector and the calming presence in our lives. He set the standard for what a sibling bond should look like, always taking care of us and fiercely encouraging us to pursue our dreams. Namrata, my twin sister, is my partner in all happiness. For anyone who knows me, understands that my life revolves around her. She has always been my rock, motivating me on bad days and understanding me in all my elements. Sanchit, my twin sister's fiancé, is the newest addition to the family, and I am so grateful to him for being such a wonderful partner to my sister. He is an extremely kind and warm person, and I have enjoyed our trips and conversations together. My in-laws—Mira, Shekharam, Radha, and Aishwarya—have embraced me with warmth and love and have been

sincerely invested in my work, celebrating each of my milestones with pride. I am very grateful to them for raising a son like my husband and providing me with a second home that has given me so much joy.

Finally, to Aditya, my husband, my love, my greatest supporter, and my home. Since our first meeting sixteen years ago, he has been a constant source of joy and stability in my life. Sharing the PhD journey with him in Ann Arbor has been an incomparable blessing. His patience, empathy, and boundless curiosity ignite my enthusiasm for learning. I will always cherish our numerous running and gym escapades, reading marathons, nature expeditions, and shared joys of cooking food together. I owe a lot of who I am today to him. He inspired us to adopt a vegan lifestyle, a decision that has shaped my values as an individual to this day. His affection towards me, my family, and those around him is a testament to his kind and loving personality.

To everyone who helped me along this journey—mentors, friends, and family—your belief and encouragement have brought me to this day. Thank you for carrying me through the challenges and for celebrating the moments of joy. This accomplishment is as much yours as it is mine.

Table of Contents

Dedication ii

Acknowledgements iii

List of Tables xiv

List of Figures xv

List of Appendices xviii

Abstract xix

Chapter 1: Beyond the Genome: The Chromatin Regulome Instructs Cellular Behavior 1

The morphogenesis of the prostate gland 1

Disease evolution and clinical management of prostate cancer 3

Prostate cancer incidence 3

PCa diagnosis and grading 3

Clinical management 5

The mutational roots of prostate cancer 7

Prostate cancer is a transcription factor addicted disease 7

Cancer-specific epigenetic proteins reprogram AR to its oncogenic sites 9

A pioneer factor dictates organ lineage specification and drives cancer 10

Chapter acknowledgements 11

References 11

Chapter 2: NSD2 is a Requisite Subunit of the AR/FOXA1 Neo-Enhanceosome in Promoting Prostate Tumorigenesis 19

Abstract 19

Introduction 20

Results 20

Functional CRISPR screen reveals NSD2 as an AR co-activator 20

NSD2 activates neo-enhancers with chimeric AR-half motifs 25

NSD1 and NSD2 independently enable oncogenic AR activity 29

NSD1/2 dual PROTAC preferentially kills AR+ prostate cancer 33

Discussion 35

References 37

Materials and Methods 44

Chapter acknowledgements 54

Author contributions 54

Chapter 3: Divergent FOXA1 Mutations Drive Prostate Tumorigenesis and Therapy-Resistant Cellular Plasticity 56

Abstract 56

Introduction 57

Results 59

FOXA1 Class 1 mutation with Trp53 deletion drives high-grade invasive adenocarcinoma 59

FOXA1 Class 1 mutants co-activate AR and mTORC1 tumorigenic pathways 66

FOXA1 Class 2 mutations trigger castration-associated intra-luminal plasticity 73

FOXA1 Class 2 mutants activate KLF5/AP-1 neo-enhancers to drive castration resistance 78

Discussion 81

References 85

Materials and Methods 102

Chapter acknowledgements 116

Author contributions 116

Chapter 4: Discussion 118

AR activity is reprogrammed to enable malignant phenotypes during prostate tumorigenesis 118

Targeting cancer-specific epigenetic cofactors to disrupt oncogenic lineage programs 120

Paralog co-targeting as a novel approach for cancer therapy 121

Distinct classes of FOXA1 mutations activate divergent oncogenic programs..... 123

Metastatic cancers reactivate features of their embryonic past 124

Chapter acknowledgements 126

References..... 126

Appendices.....131

List of Tables

Supplementary Table 1: Antiproliferative half-maximal inhibitory concentration (IC50) of LLC0150 across human-derived normal and cancer cell lines 178
Supplementary Table 2: Primer sequences used in the study from Chapter 2 185
Supplementary Table 3: siRNA sequences used in the study from Chapter 2 186
Supplementary Table 4: sgRNA sequences used in the study from Chapter 2 187
Supplementary Table 5: Putative FOXA1 Class 1 mutant gene targets in mouse prostate organoids 188

List of Figures

Figure 2-1: Epigenetics-focused CRISPR screen shows NSD2 as an AR coactivator 21
Figure 2-2: NSD2 expands the AR neo-enhancer circuitry to include chimeric AR half-sites 24
Figure 2-3: NSD1 and NSD2 independently enable oncogenic AR activity 28
Figure 2-4: LLC0150 is an NSD1/2 PROTAC with preferential cytotoxicity in AR-driven PCa 32
Figure 2-5: Schema depicting NSD2's role in loading the AR enhanceosome at tumor-enriched chimeric AR neo-enhancer elements 36
Figure 3-1: FOXA1 Class 1 mutants drive high-grade invasive adenocarcinoma in the background of p53 loss 61
Figure 3-2: FOXA1 Class 1 mutants co-activate AR and mTORC1 oncogenic signaling 65
Figure 3-3: FOXA1 Class 2 mutants trigger castration-associated intra-luminal plasticity 72
Figure 3-4: FOXA1 Class 2 mutants activate KLF5/AP-1 neo-enhancer circuitry to drive therapy-resistance 77
Figure 3-5: Schema depicting distinct oncogenic mechanisms of FOXA1 Class 1 and Class 2 mutations in driving prostate tumorigenesis or therapy-resistant progression 84
Supplementary Figure A2-1: Generation and characterization of the endogenous mCherry-PSA AR reporter lines 132
Supplementary Figure A2-2: NSD2 transcript and protein expression in primary patient specimens 134
Supplementary Figure A2-3: NSD2 and H3K36me2 expression in patient tumors and prostate cancer cell lines 136

Supplementary Figure A2-4: Motif characterization of the NSD2-enabled AR neo-cistrome in prostate cancer cells 138
Supplementary Figure A2-5: Molecular characterization of the NSD2-rescued prostate cancer cells 140
Supplementary Figure A2-6: GREAT neighboring genes and pathway enrichment analyses ... 142
Supplementary Figure A2-7: Fragment-based NSD2–AR coimmunoprecipitation and characterization of the NSD paralog knockout prostate cancer cells 144
Supplementary Figure A2-8: Mechanistic characterization of the NSD1/2 PROTAC degrader LLC0150 146
Supplementary Figure A2-9: Transcriptomic effect and drug synergism of LLC0150 in prostate cancer cells 148
Supplementary Figure A2-10: Efficacy assessment of LLC0150 in prostate cancer organoids and xenografts 150
Supplementary Figure A3-1: Generation and characterization of FOXA1 Class 1 mutant transgenic mice 153
Supplementary Figure A3-2: High-grade invasive carcinoma in FOXA1 R265-71del+/+;Trp53f/f animals 155
Supplementary Figure A3-3: Single-cell transcriptomic characterization of FOXA1 R265-71del+/+;Trp53f/f tumor tissues 157
Supplementary Figure A3-4: FOXA1 Class 1 tumors are enriched for hyper-proliferative and oncogenic signatures 159
Supplementary Figure A3-5: FOXA1 Class 1 mutants reprogram AR and activate mTORC1 signaling 161
Supplementary Figure A3-6: NSD2 reprograms AR at chimeric AR half-sites 163
Supplementary Figure A3-7: FOXA1 Class 2 mutants cannot initiate pre-cancerous lesions ... 165
Supplementary Figure A3-8: Cistromically dominant FOXA1 Class 2 mutants activate stemness gene programs 167
Supplementary Figure A3-9: FOXA1 Class 2 mutants drive higher stemness potential 169

Supplementary Figure A3-10: FOXA1 Class 2 mutants pioneer neo-enhancers co-bound by KLF5 and AP-1 family proteins 171
Supplementary Figure A3-11: FOXA1 Class 2 mutants activate KLF5 in clinical datasets..... 173
Appendix Figure C-1: Schematic of the construct used to engineer the FOXA1 transgenic mouse lines 175

List of Appendices

Appendix A: Supplementary Figures for Chapter 2……131

Appendix B: Supplementary Figures for Chapter 3……152

Appendix C: Supplementary Notes for Chapter 3 ……174

Abstract

While cancer has traditionally been viewed as a genetic disease, the role of epigenetics, which literally translates to "above genetics," has revealed another layer of complexity. Somatic and germline mutations, in collaboration with reversible histone and DNA chemical modifications, reprogram chromatin architecture and modulate gene expression. Multiprotein complexes deposit and interpret these epigenetic marks, functioning as readers, writers, or erasers of epigenetic information. Notably, many epigenetic proteins are mutated or aberrantly expressed in cancer, leading to the recognition of "nonmutational epigenetic reprogramming" as a new hallmark of human cancers. Today, we reframe cancer as a result of interactions between the genome and epigenome, a theme that is the focus of my thesis. Here, we uncover the genomic and epigenomic contributions underlying oncogenic androgen receptor (AR) reprogramming, prostate cancer development, and therapy resistance, with a particular emphasis on understanding the interplay between two key proteins: FOXA1 and NSD2.

The cellular identity of a prostate cell is determined by the expression of AR and its downstream signaling. In a normal prostate luminal epithelial cell, AR drives terminal differentiation, i.e., its activity impedes proliferation. However, upon transformation, AR activates malignant and proliferative gene programs, largely mediated through extensive reprogramming of AR's chromatin occupancy. Here, using a CRISPR-screening approach, we identified NSD2, an H3K36 dimethyltransferase, as a critical epigenetic cofactor responsible for reprogramming AR at its tumor-specific enhancers. Notably, NSD2 expression is abnormally induced in prostate cancer, where it collaborates with FOXA1 to reprogram AR binding at enhancer sites that harbor chimeric FOXA1:AR half-motifs. Loss of NSD2 inactivates cancer hallmark features of invasion and migration and fosters increased dependency on a paralogous gene called NSD1. Hence, we developed a dual NSD1/2 PROTAC degrader that exhibits preferential cytotoxicity in AR-dependent prostate cancer. Altogether, we characterize NSD2 as an essential AR neo-

enhanceosome subunit that enables its oncogenic activity, and position NSD1/2 as viable co-targets in advanced prostate cancer.

NSD2-dependent reprogramming of AR at chimeric half-sites relies on FOXA1's pioneering activity. FOXA1 decompacts chromatin to recruit lineage-specific transcription factors, such as AR in the prostate epithelia, and chromatin remodeling proteins. Earlier, our lab found FOXA1 to be recurrently altered in over 35% of metastatic castration-resistant prostate cancer (mCRPC) cases in Caucasian White men, which recur within three distinct structural classes with distinct clinical incidence and gain-of-function. More recently, FOXA1 was reported to be mutated in over 40% of primary prostate cancer in Chinese patients, positioning FOXA1 as a commonly mutated oncogene in this disease. Yet, its tumorigenic potential and pathobiology remain unexplored *in vivo*. As a central component of this thesis, we have developed and characterized the first in-field transgenic mouse models that conditionally overexpress FOXA1 Class 1 or Class 2 mutant in the mouse prostate luminal epithelia. Our findings reveal that FOXA1 Class 1 mutations, that disrupt the wing2 sub-secondary region, in a *Trp53*-null background drive high-grade, AR-dependent luminal prostate adenocarcinoma. In contrast, Class 2 mutations are unable to drive disease formation. Instead, they induce intra-luminal plasticity by reprogramming differentiated cells into an AR-positive, luminal progenitor state that is androgen-insensitive. Mechanistically, Class 1 mutants co-activate mTORC1 signaling and reprogram AR at chimeric AR half-sites in concert with aberrantly gained NSD2. On the other hand, Class 2 mutations activate stemness genes and decompact over 40,000 new enhancer elements, bound by stemness-associated transcription factors, like KLF5 and AP1. This instructs an androgen-insensitive luminal progenitor cell fate, making the Class 2-mutant mouse prostate tissues relatively resistant to atrophy upon androgen withdrawal through castration. Collectively, these findings establish FOXA1 as a key driver of prostate cancer development, showcasing how specific alterations in this single oncogene can either trigger tumorigenesis in early disease or enable resistance to treatment in more advanced stages.

In summary, this doctoral thesis elucidates the functional and phenotypic impacts of FOXA1 mutations in prostate cancer and identifies NSD2 as a cancer-specific epigenetic regulator that collaborates with FOXA1 to drive AR-dependent oncogenesis. Notably, this work presents the

first evidence of FOXA1-driven prostate adenocarcinoma formation in murine models and further substantiates the structural classification of distinct FOXA1 alterations within its coding region or genomic locus. Furthermore, this thesis characterizes collaborative reprogramming of the prostate cancer epigenome by the NSD1 and NSD2 paralogous enzymes, emphasizing paralog co-targeting as a promising therapeutic strategy for advanced, lethal forms of the disease.

Chapter 1

Beyond the Genome: The Chromatin Regulome Instructs Cellular Behavior

While the history of cancer has been closely tied to alterations in the DNA sequence of somatic or germ cells, the fields of epigenetics and epigenomics, which literally translate to "above genetics," have revealed another layer of complexity. Epigenetic modifications, including histone methylation, acetylation, sumoylation, and DNA methylation, can alter gene expression and chromatin structure, thereby affecting cell fate decisions without changing the underlying DNA sequence [1,2]. This is particularly evident during development, where histone modifications govern gene activation and repression, chromatin architecture, and the expression of lineage-defining and body-patterning transcription factors, such as the Hox family genes [3–5]. Multiprotein complexes interpret these epigenetic marks, functioning as readers, writers, or erasers [1]. The dynamic and malleable nature of these marks allows for significant cellular plasticity and adaptation, both during development and in response to stress, such as injury or cancer. Notably, many epigenetic proteins are mutated or aberrantly expressed in cancer, leading to the recognition of "nonmutational epigenetic reprogramming" as a new hallmark of many cancers [6]. These findings have spurred the development of numerous preclinical and clinical drugs targeting epigenetic proteins, such as histone deacetylases, methyltransferases, and histone acetyltransferases, underscoring the clinical significance of the epigenome. Today, we rethink cancer as a result of interactions between the genome and epigenome. My thesis focuses on uncovering the genomic and epigenomic contributions underlying prostate cancer development, with a particular emphasis on two key proteins: FOXA1 and NSD2.

The morphogenesis of the prostate gland

The prostate gland is an accessory male reproductive organ located at the base of the bladder. Its name is derived from the Greek word "prostates," meaning "one who stands before" or "protector," a reference to its anatomical position in front of the bladder [7]. The primary role of the prostate

gland is to secrete prostatic fluid, which contributes to sperm viability and the formulation of the ejaculate fluid [8]. The prostate gland develops from the urogenital sinus through an intricate process of budding, elongation, branching, and ductal morphogenesis [9–11]. These distinct stages are regulated by the timed expression of androgens and their nuclear hormone receptor, the Androgen Receptor (AR), at embryonic day 12-13 [12]. AR is expressed in both the epithelial and stromal compartments of the prostate, and during development, stromal AR is particularly critical for prostate determination. Importantly, paracrine signaling between epithelium and stroma is integral to proper prostate development and function [9,12]. While the distinction between stromal and epithelial AR signaling remains less explored, most research has focused on its role in the epithelial compartment, where it helps maintain the differentiated, secretory phenotype and overall homeostasis of the prostate gland.

While AR is central to the establishment and maintenance of prostate cell identity, its activity at lineage-defining cis-regulatory elements—also known as enhancers—is critically dependent on the pioneer transcription factor FOXA1 [13]. FOXA1 can bind compacted chromatin, displace H1 linker histones to decompact the underlying DNA sequence, thereby enabling the recruitment of AR and other chromatin remodeling factors [14]. Notably, FOXA1 is expressed at the onset of prostate development, even before AR is detectable in the epithelium, and is essential for establishing the chromatin landscape required for subsequent AR-driven transcriptional programs [15].

As the prostate gland develops, AR expression increases in both the epithelial and stromal compartments. Upon androgen binding, AR translocates to the nucleus, where it binds to FOXA1-primed enhancers containing a 15-bp palindromic androgen response element (ARE) and activates gene expression programs necessary for prostate differentiation and function [16]. This interplay ensures that androgen signaling is precisely targeted to genes essential for prostate morphogenesis and secretory activity. The mature, terminally differentiated prostate gland comprises a bilayered epithelium with a discontinuous outer layer of basal cells and an inner layer of AR/FOXA1+ secretory luminal cells interspersed with neuroendocrine cells [17,18]. Basal cells, marked by cytokeratins such as CK5 and p63, provide structural support and progenitor potential, while

luminal cells, characterized by the expression of CK8/CK18 and high levels of AR, are responsible for secreting prostatic fluid [19].

The adult human prostate gland is organized into three distinct zones – central, transitional, and peripheral – whereas in mice, it comprises four lobes: the dorsal, ventral, lateral, and anterior [20,21]. Notably, approximately 75% of human prostate adenocarcinoma originates in the peripheral zone and is characterized by the hyperproliferation of AR+/FOXA1+ luminal epithelial cells and the loss of basal cells, resulting in a strong luminal phenotype [18]. Since nearly all prostate tumors express AR, recent efforts have focused on understanding the mechanisms of AR-positive prostate cancer and identifying novel therapeutic targets, beyond AR itself. In this thesis, I will uncover the molecular mechanisms governing FOXA1-driven prostate tumorigenesis by using genetically engineered mouse models that carry human-relevant FOXA1 mutations.

Disease evolution and clinical management of prostate cancer

Prostate cancer incidence

Prostate cancer (PCa) is the second-leading cause of cancer-related deaths in North America. In 2025 alone, the American Cancer Society predicts approximately 300,000 new cases, with 35,000 resulting in new deaths [22]. Globally, there are more than 1 million new diagnosed cases annually, with the deaths exceeding 350,000 [23]. Racial disparities significantly contribute to PCa incidence [18]. For instance, African and Caribbean men have the highest risk of early, aggressive disease relative to the white population, while Asian men living in Asia have a lower risk of disease [24,25]. Additionally, approximately 10% of PCa cases have a substantial familial inheritance, where germline mutations in DNA-repair and mismatch repair genes, such as BRCA2, BRCA1, ATR, MSH2, and a key prostate-specific transcription factor (TF) called HOXB13 (G84E), strongly dictate the early onset of aggressive PCa [26–30].

PCa diagnosis and grading

PCa diagnosis and staging are guided by a multi-parametric approach that utilizes digital rectal examinations (DRE), blood-based biochemical screening (prostate-specific antigen, PSA), histologic assessments of tissue architecture, and various imaging modalities [31]. Physical rectal

examination of the prostate gland using DRE along with PSA screening is the frontline of PCa diagnosis [31]. In case of abnormalities in the DRE or elevated PSA levels (>3ng/ml), a prostate biopsy is often used to confirm the presence of cancer. The Gleason Grade Group system, developed by Dr. Jonathan Epstein and colleagues at Johns Hopkins, USA, serves as the benchmark for assessing tumor aggressiveness through visual microscopic inspection of the tumor tissue obtained through multiple transrectal or transperineal core biopsies [18,32]. This system relies on the microscopic architecture of prostate glands, as originally described by the Gleason scoring method, to differentiate between poorly differentiated (aggressive) tumors and well-differentiated tumors. Pathologists evaluate prostate tissue samples and identify the two most common architectural patterns of cancer, each assigned a grade from 3 to 5. The Gleason score is calculated by adding the two most common grades, resulting in scores ranging from 6 (least aggressive) to 10 (most aggressive). Further, the Gleason scores are grouped into Grade Groups 1 through 5. Grade Group 1 (Gleason score 6) represents the least aggressive cancers, often suitable for active surveillance. Grade Groups 2 and 3 (Gleason score 7) indicate an intermediate risk, with tumors in Grades 3+4 being less aggressive than those in Grades 4+3. Grade Groups 4 and 5 (Gleason scores 8 to 10) correspond to high-risk, aggressive disease that typically requires more intensive treatment [33,34].

Prostate cancer develops through a series of genetic, epigenetic, and cellular changes that transform a normal, luminal epithelial cell into cancer[35]. High-grade prostatic intraepithelial neoplasia (HGPIN) is recognized as the most established precursor, characterized by the abnormal proliferation of glandular luminal epithelial cells, and is often found adjacent to carcinoma [18,35]. This includes enlarged atypical nuclei with multiple nucleoli and increased expression of cell cycle genes. A notable pathological distinction between HGPINs and acinar adenocarcinomas is the absence of p63+/CK5+ basal cells in the latter, along with intense staining of luminal cell-specific cytokeratin 8/18, and elevated levels of cancer-related proteins, such as α-methylacyl-CoA racemase (AMACR)[36]. While most prostate adenocarcinomas are well differentiated, more aggressive disease subtypes evolve *de novo* or under therapy pressures to gain poorly differentiated features, including small-cell neuroendocrine prostate cancer [37]. While prostate acinar adenocarcinoma that develops in the prostate gland accounts for 95% of cases, ductal and

intraductal PCa (DPC and IDCP) are less frequent, yet more aggressive disease forms that are characterized by papillary and cribriform formations.

Clinical management

A clinical observation in the 1940s by Dr. Charles Huggins revealed that removing testicular androgens, either through orchiectomy or the administration of estrogens, caused significant regression of prostate tumors in men with advanced disease [38]. This discovery was the first to demonstrate the essential role of androgens in prostate cancer and laid the foundation for hormone-based therapies, i.e., androgen deprivation therapies (ADT), also referred to as castration, that remain central to prostate cancer management today [39]. However, the biology of androgens and the discovery of the androgen receptor (AR) occurred later in the 1960s, leading to the development of more potent, second-generation anti-androgen drugs that are used in advanced disease [40–42]. PCa is a heterogeneous disease, and treatment options are often dictated by the disease stage, histopathological features (acinar, ductal, or neuroendocrine), and biochemical monitoring.

Approximately 80–85% of patients are diagnosed with localized disease. [18,43] The management of primary localized prostate cancer is tailored based on risk stratification, incorporating factors such as PSA level, Gleason score, and tumor stage. For low-risk, localized disease, active surveillance is often preferred, especially in older patients or those with significant comorbidities [44]. Definitive treatments for localized or locally advanced disease include radical prostatectomy and radiation therapy, while patients who have intermediate- or high-risk disease are given radiation therapy in combination with ADT for 18–36 months [45,46]. Advances in imaging, such as multiparametric MRI [47] and PSMA PET [48], have improved staging accuracy and patient selection for curative treatments. These treatments often lead to a decline in PSA levels (<0.1 ng/ml). While prostatectomy and radiation therapy cure about 70% of patients, about 20–30% of patients will relapse with a higher chance of metastatic disease, as indicated by a biochemical recurrence or elevated PSA levels, and are subsequently administered ADT therapies like luteinizing hormone-releasing hormone (LHRH) agonists or antagonists—also referred to as medical castration [49]. These therapies inhibit the testicular production of androgen by ablating the pituitary-testis signaling axis.

For castration-sensitive prostate cancer (CSPC)—particularly in the metastatic setting (mCSPC)—management has evolved from ADT monotherapy to combination regimens. Current standards recommend ADT combined with second-generation androgen receptor signaling inhibitors (ARSIs) such as abiraterone or enzalutamide, both of which have demonstrated significant survival benefits. Abiraterone works by inhibiting CYP17, a key enzyme in androgen biosynthesis, thereby reducing androgen production from the testes, adrenal glands, and even within the tumor itself [50,51]. Enzalutamide, on the other hand, directly inhibits the androgen receptor, blocking androgen binding and its nuclear translocation, thereby impeding its transcriptional activity [52]. In patients with high-volume metastatic disease, the addition of docetaxel chemotherapy to ADT and an ARSI (triplet therapy) can further improve outcomes. The integration of these agents into treatment regimens has significantly advanced the standard of care for men with advanced prostate cancer, offering improved survival and delayed disease progression.

However, despite the initial treatment response to castration, nearly all patients will progress to an androgen-independent disease, also referred to as castration-resistant prostate cancer (CRPC). Here, a combination of therapies, including chemotherapy (docetaxel, cabazitaxel), ARSIs (enzalutamide, abiraterone, apalutamide), PARP inhibitors (especially for those with DNA repair gene mutations), and PSMA-targeted radioligand therapy, is administered [18,53]. The choice of therapy is increasingly informed by prior treatments and molecular profiling, with a focus on personalizing care to optimize both survival and quality of life.

A distinctive feature of primary, mCSPC, and mCRPC tumors is their continued reliance on AR and its signaling. In CRPC, castrate levels of testosterone are sufficient to maintain cell growth and proliferation. However, continuous treatment with ADT and ARSI therapy has coerced prostate cancer cells to exit their lineage and transdifferentiate into an AR-negative disease subtype known as neuroendocrine prostate cancer (NEPC) [54–56]. This phenomenon is known as lineage plasticity, specifically intercellular plasticity, which is observed in approximately 20% of CRPC patients. This process of transdifferentiation renders most current therapies ineffective, owing to the aggressive and AR-independent nature of the tumor. Understanding the molecular drivers of NEPC will aid in the development of novel therapies for this aggressive disease.

The mutational roots of prostate cancer

PCa exemplifies the idea of lineage addiction, wherein AR serves as the master regulator of prostate cell identity [39]. Extensive sequencing efforts of both primary and metastatic castration-resistant prostate cancer tumors has revealed that PCa progresses through a multi-step process involving the accumulation of diverse clonal and subclonal genetic alterations, ultimately leading to the development of an AR-positive, luminal-type disease. The genomic landscape of PCa is primarily characterized by structural rearrangements and copy number alterations, rather than single-nucleotide variations.

Prostate cancer is a transcription factor–addicted disease

One of the hallmarks of primary PCa in Caucasian men is the presence of the *TMPRSS2-ERG* gene fusion, which is observed in approximately 50% of cases [57–59]. Here, an AR-regulated promoter, *TMPRSS2*, is fused to a member of the ETS family (ERG, ETV1, and ETV4), making it the first report of gene fusions in solid tumors. Multi-institutional efforts and next-generation sequencing technologies have led to a more comprehensive genomic map of PCa [60]. DNA sequencing studies from The Cancer Genome Atlas (TCGA, primary PCa), MSK-IMPACT (mCSPC and mCRPC), and SU2C-PCF (mCRPC) have broadened our understanding of PCa evolution. Here, in addition to the TMPRSS2-ERG gene fusion, recurrent mutations/alterations in transcription factors and chromatin remodelers, including FOXA1 Class 1 (a pioneer transcription factor and co-factor of AR) and SPOP (an E3 ubiquitin ligase of AR), as well as amplifications of oncogenes such as MYC (chromosome 8q gain), PIK3CA, CCND1 and loss of tumor suppressor genes like PTEN, NKX3.1 (chromosome 8p loss), TP53, and CDK12, were observed in the primary disease [61–66]. These alterations not only initiate tumorigenesis but also help shape the evolving chromatin landscape, increasing accessibility at regulatory regions and priming the cancer genome for further transcriptional dysregulation.

In metastatic disease, particularly castration-resistant prostate cancer, there is a significantly higher incidence of bi-allelic losses of tumor suppressor genes, such as PTEN, TP53, RB1, and CDK12, which deprive the cell of its proliferative gatekeeper proteins [67,68]. Growth-promoting gain-of-function alterations in the PI3K pathway (PIK3CA and AKT1) are seen in about 6% of advanced

tumors. Additionally, activation of WNT signaling in approximately 18% of mCRPC cases, driven by the loss of APC and gain of CTNNB1, promotes aggressive metastatic phenotypes. While the genome of a PCa cell is relatively stable compared to many other tumor types, a unique subset of DDR or mismatch repair-deficient tumors harbors loss-of-function mutations in BRCA2 (6-12.5%), ATM (5-7%), and CDK12 [67,68]. These tumors, owing to their instability, have a higher burden of neo-antigens and hence benefit from immune checkpoint blockade therapies. Finally, metastatic tumors have an increased incidence of FOXA1 Class 2 mutations, which activate WNT signaling to drive metastasis [64].

While alterations in transcription factors serve as key drivers for PCa, epigenetic and chromatin remodelling proteins help demarcate and reshape chromatin accessibility for transcriptional activation. For instance, loss of CHD1, a chromatin remodeling protein, is observed in approximately 15% of primary PCa cases, where it redistributes AR to occupy enhancer elements that drive its oncogenic program [65, 69]. Mutations in epigenetic modifier proteins, such as KMT2C and KMT2D, are also detected, albeit at lower frequencies [67]. In addition to these, sustained or aberrant expression of proteins belonging to the BRD family of proteins [70], histone methyl transferases such as NSD2 [71,72], and the SWI/SNF chromatin remodelling complex [73] are essential for the growth, survival, or invasion of PCa cells.

These genomic insights reveal that PCa is fundamentally a transcription factor–addicted malignancy, with AR being the central node of the disease. Under homeostatic conditions, AR instructs terminal differentiation of prostate epithelial cells. While AR is rarely mutated in the primary disease [61], PCa cells reinforce AR activity at all stages of disease, particularly under therapy pressure through copy amplifications of AR and its super-enhancer, mutations in the AR ligand-binding domains, and AR splice variants (AR-V7) that enable ligand-independent activation of AR [39,74,75]. Additionally, recurrent alterations in co-factors or regulators of AR, such as FOXA1, SPOP, HOXB13, and NCOR1/2, highlight the importance of sustained AR signaling in PCa. These findings raise a pertinent question: If AR drives terminal differentiation, i.e., proliferative arrest, why do cancer cells restore AR activity?

Notably, seminal studies have highlighted that AR chromatin binding and activity are significantly reprogrammed during transformation [76]. Upon binding to its ligand, dihydrotestosterone (DHT), AR dimerizes and translocates into the nucleus where it binds as a homodimer at its 15-bp palindromic AR response element (ARE) [39]. These sites constitute its normal cistrome. However, upon transformation, AR is reprogrammed to new genomic sites that drive its oncogenic program. Hence, rather than enforcing differentiation, AR signaling becomes hijacked to drive proliferation and survival. However, the mechanisms for this reprogramming remained elusive. While FOXA1 and HOXB13 were deemed essential for this reprogramming, both proteins are also expressed in a normal prostate, suggesting that other cancer-specific co-factors likely collaborate with lineage-defining pioneer transcription factors, such as FOXA1 and HOXB13, to drive cancer-specific reprogramming of AR [76]. Here, we have identified NSD2 and FOXA1 as essential proteins required for reprogramming AR activity in prostate cancer. In the following chapters, I will highlight the role of NSD2, a cancer-specific epigenetic co-factor of AR (Chapter 2), as well as the driver roles of distinct FOXA1 mutations (Chapter 3) in driving prostate tumorigenesis.

Cancer-specific epigenetic proteins reprogram AR to its oncogenic sites

Cancer has traditionally been viewed as a genetic disease that is driven by the accumulation of somatic mutations in oncogenes and tumor suppressor genes that disrupt normal cellular processes such as growth, division, and apoptosis. However, it is now clear that the epigenome—the collection of chemical modifications to DNA and histone proteins — plays an equally critical role in tumorigenesis and cancer progression [1,2]. Epigenetic alterations, including DNA methylation, histone modifications, and chromatin remodeling, can profoundly influence gene expression programs, cellular identity, and therapeutic response. In prostate cancer and many other malignancies, these epigenetic changes are not merely passive consequences of genetic mutations but are active drivers that cooperate with genetic alterations to shape cancer cell behavior. While there are diverse families of epigenetic proteins, including histone methyltransferases, histone acetyltransferases, DNA methyltransferases, and chromatin remodelers, this thesis focuses on one such histone methyltransferase, NSD2.

The NSD (Nuclear receptor-binding SET Domain) family, comprising NSD1, NSD2 (also known as WHSC1 or MMSET), and NSD3, has been demonstrated to have essential roles in development

and cancer [77]. These enzymes catalyze the methylation of histone H3 at lysine 36 (H3K36), a modification associated with active transcription and open chromatin states. Chapter 2 of this thesis focuses on NSD2, also known as MMSET, which was initially identified as a driver oncogene in multiple myeloma. NSD2 translocations (t4;14) are observed in approximately 15-20% of multiple myelomas, where NSD2 is positioned under the regulation of an IgH enhancer, driving aggressive disease phenotypes [78,79]. Later, NSD2 was also identified as a driver of metastasis in prostate cancer, consistent with its elevated expression in mCRPC relative to the primary disease [71,72]. Interestingly, NSD2 expression is the highest in neuroendocrine prostate cancer, where it is proposed to counteract the repressive functions of EZH2. However, this remains to be fully characterized [80]. In Chapter 2, we mechanistically demonstrate that NSD2 serves as a cancer-specific co-factor of AR, where it reprograms AR away from its normal binding sites to its oncogenic sites.

A pioneer factor dictates organ lineage specification and drives cancer

FOXA1, also known as hepatocyte nuclear factor 3-alpha (HNF3α), is a member of the forkhead box (FOX) family of pioneer transcription factors, distinguished by their butterfly-like winged helix DNA-binding domain [81,82]. This structural feature enables FOXA1 to decompact chromatin, thereby facilitating the recruitment of lineage-specific transcription factors and chromatin remodeling proteins. Initially identified in the liver, where it regulates genes essential for hepatic differentiation, FOXA1 has since been recognized as a master regulator of cell fate in multiple organs, including the breast and prostate [13]. In the breast, FOXA1 plays a crucial role in mammary gland development, working in concert with the estrogen receptor to establish luminal cell identity [83]. In the prostate gland, FOXA1 is indispensable for ductal morphogenesis, epithelial differentiation, and function, as it reshapes the chromatin landscape and enables AR-driven gene expression, thereby dictating prostate lineage identity [15].

While FOXA1 is essential for normal development of the prostate, recent work from our lab demonstrated that FOXA1 is recurrently mutated in over 35% of prostate cancers in Caucasian men [64]. Notably, this number increases to 41% in Asian men, where FOXA1 surpasses the TMPRSS2-ERG gene fusion as the dominant oncogene [84]. FOXA1 mutations are also detected in breast cancer, where they reprogram estrogen receptor activity to drive tumorigenesis [64,85,86]. While the primary paper from our lab [64] defined three distinct structural classes of

FOXA1 alterations (Classes 1, 2, and 3) in PCa, Chapter 3 of this thesis focuses on uncovering the pathobiology and driver roles of these mutations using the first-in-field genetically engineered mouse models that harbor the distinct FOXA1 mutations. Here, by integrating multiomics data from single-cell RNA/ATAC sequencing, bulk RNA-sequencing, and ChIP-sequencing on mouse prostate tissues and tissue-derived 3D organoids, we functionally demonstrate how distinct FOXA1 mutations drive divergent PCa phenotypes [87].

Chapter acknowledgements:

This chapter was written by Sanjana Eyunni. Overarching ideas and concepts were integrated from prior individual discussions with Abhijit Parolia and Arul Chinnaiyan. Figure 1-1 was adapted from a similar summary figure in Eyunni S, et al. Divergent FOXA1 mutations drive prostate tumorigenesis and therapy-resistant cellular plasticity. *Science*. 2025. doi: 10.1126/science.adv2367. Rephrasing of select paragraphs was done using Perplexity AI, though all ideas and drafts were developed solely by Sanjana Eyunni.

References:

1. Arrowsmith CH, Bountra C, Fish PV, Lee K, Schapira M. Epigenetic protein families: a new frontier for drug discovery. Nat Rev Drug Discov. 2012;11: 384–400.

2. Strahl BD, Allis CD. The language of covalent histone modifications. Nature. 2000;403: 41–45.

3. Cheedipudi S, Genolet O, Dobreva G. Epigenetic inheritance of cell fates during embryonic development. Front Genet. 2014;5: 19.

4. Agger K, Cloos PAC, Christensen J, Pasini D, Rose S, Rappsilber J, et al. UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. Nature. 2007;449: 731–734.

5. Cantone I, Fisher AG. Epigenetic programming and reprogramming during development. Nat Struct Mol Biol. 2013;20: 282–289.

6. Hanahan D. Hallmarks of cancer: New dimensions. Cancer Discov. 2022;12: 31–46.

7. Prostate - Etymology, Origin & Meaning. In: etymonline [Internet]. [cited 14 Jul 2025]. Available: https://www.etymonline.com/word/prostate

8. Verze P, Cai T, Lorenzetti S. The role of the prostate in male fertility, health and disease. Nat Rev Urol. 2016;13: 379–386.
9. Cunha GR, Vezina CM, Isaacson D, Ricke WA, Timms BG, Cao M, et al. Development of the human prostate. Differentiation. 2018;103: 24–45.
10. 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.
11. Cunha GR, Donjacour AA, Cooke PS, Mee S, Bigsby RM, Higgins SJ, et al. The endocrinology and developmental biology of the prostate. Endocr Rev. 1987;8: 338–362.
12. Pointis G, Latreille MT, Mignot TM, Janssens Y, Cedard L. Regulation of testosterone synthesis in the fetal mouse testis. J Steroid Biochem. 1979;11: 1609–1612.
13. Augello MA, Hickey TE, Knudsen KE. FOXA1: master of steroid receptor function in cancer. EMBO J. 2011;30: 3885–3894.
14. 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.
15. 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.
16. Lv S, Song Q, Chen G, Cheng E, Chen W, Cole R, et al. Regulation and targeting of androgen receptor nuclear localization in castration-resistant prostate cancer. J Clin Invest. 2021;131: e141335.
17. van Leenders GJLH, Schalken JA. Epithelial cell differentiation in the human prostate epithelium: implications for the pathogenesis and therapy of prostate cancer. Crit Rev Oncol Hematol. 2003;46 Suppl: S3–10.
18. Rebello RJ, Oing C, Knudsen KE, Loeb S, Johnson DC, Reiter RE, et al. Prostate cancer. Nat Rev Dis Primers. 2021;7: 9.
19. Toivanen R, Shen MM. Prostate organogenesis: tissue induction, hormonal regulation and cell type specification. Development. 2017;144: 1382–1398.
20. Oliveira DSM, Dzinic S, Bonfil AI, Saliganan AD, Sheng S, Bonfil RD. The mouse prostate: a basic anatomical and histological guideline. Bosn J Basic Med Sci. 2016;16: 8–13.

21. Henry GH, Malewska A, Joseph DB, Malladi VS, Lee J, Torrealba J, et al. A Cellular Anatomy of the Normal Adult Human Prostate and Prostatic Urethra. Cell Rep. 2018;25: 3530–3542.e5.
22. Key Statistics for Prostate Cancer. [cited 12 Jul 2025]. Available: https://www.cancer.org/cancer/types/prostate-cancer/about/key-statistics.html
23. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74: 229–263.
24. Shenoy D, Packianathan S, Chen AM, Vijayakumar S. Do African-American men need separate prostate cancer screening guidelines? BMC Urol. 2016;16: 19.
25. Zeigler-Johnson CM, Spangler E, Jalloh M, Gueye SM, Rennert H, Rebbeck TR. Genetic susceptibility to prostate cancer in men of African descent: implications for global disparities in incidence and outcomes. Can J Urol. 2008;15: 3872–3882.
26. Hemminki K. Familial risk and familial survival in prostate cancer. World J Urol. 2012;30: 143–148.
27. Storebjerg TM, Høyer S, Kirkegaard P, Bro F, LuCamp Study Group, Ørntoft TF, et al. Prevalence of the HOXB13 G84E mutation in Danish men undergoing radical prostatectomy and its correlations with prostate cancer risk and aggressiveness. BJU Int. 2016;118: 646–653.
28. Karlsson R, Aly M, Clements M, Zheng L, Adolfsson J, Xu J, et al. A population-based assessment of germline HOXB13 G84E mutation and prostate cancer risk. Eur Urol. 2014;65: 169–176.
29. McLeod DG, Srivastava S. Re: Inherited DNA-repair gene mutations in men with metastatic prostate cancer. Eur Urol. 2017;71: 692.
30. Pritchard CC, Mateo J, Walsh MF, De Sarkar N, Abida W, Beltran H, et al. Inherited DNA-repair gene mutations in men with metastatic prostate cancer. N Engl J Med. 2016;375: 443–453.
31. Mottet N, Bellmunt J, Bolla M, Briers E, Cumberbatch MG, De Santis M, et al. EAU-ESTRO-SIOG guidelines on Prostate Cancer. Part 1: Screening, diagnosis, and local treatment with curative intent. Eur Urol. 2017;71: 618–629.
32. Prostate Cancer Grading System. [cited 12 Jul 2025]. Available: https://pathology.jhu.edu/urologic/prostate-cancer-grading-system

33. 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.

34. Humphrey PA. Gleason grading and prognostic factors in carcinoma of the prostate. Mod Pathol. 2004;17: 292–306.

35. Miyai K, Divatia MK, Shen SS, Miles BJ, Ayala AG, Ro JY. Clinicopathological analysis of intraductal proliferative lesions of prostate: intraductal carcinoma of prostate, high-grade prostatic intraepithelial neoplasia, and atypical cribriform lesion. Hum Pathol. 2014;45: 1572–1581.

36. Epstein JI, Egevad L, Humphrey PA, Montironi R, Members of the ISUP Immunohistochemistry in Diagnostic Urologic Pathology Group. Best practices recommendations in the application of immunohistochemistry in the prostate: report from the International Society of Urologic Pathology consensus conference. Am J Surg Pathol. 2014;38: e6–e19.

37. Yamada Y, Beltran H. Clinical and biological features of neuroendocrine prostate cancer. Curr Oncol Rep. 2021;23: 15.

38. Huggins C. EFFECT OF ORCHIECTOMY AND IRRADIATION ON CANCER OF THE PROSTATE. Ann Surg. 1942;115: 1192–1200.

39. Mills IG. Maintaining and reprogramming genomic androgen receptor activity in prostate cancer. Nat Rev Cancer. 2014;14: 187–198.

40. 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.

41. Wong YNS, Ferraldeschi R, Attard G, de Bono J. Evolution of androgen receptor targeted therapy for advanced prostate cancer. Nat Rev Clin Oncol. 2014;11: 365–376.

42. 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.

43. 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.

44. Active surveillance for 21 the management of localized prostate cancer: Guideline recommendations.

45. Zattoni F, Maresca S, Dal Moro F, Bednarova I, Randazzo G, Basso G, et al. Abbreviated versus multiparametric prostate MRI in active surveillance for prostate-cancer patients: Comparison of accuracy and clinical utility as a decisional tool. Diagnostics (Basel). 2023;13. doi:10.3390/diagnostics13040578

46. Biradar M, Assistant Professor, Department of Radiology, Grant Government Medical College, Mumbai, INDIA. Diagnostic accuracy of Multiparametric MRI in detection of prostate cancer compared with histopathology obtained by MRI directed TRUS guided cognitive fusion biopsy are T2WI, DWI and DCE enough for Indian scenario. MedPulse International Journal of Radiology. 2021;20: 10–12.

47. Standard Operating Procedure for Multiparametric Magnetic Resonance Imaging in the Diagnosis, Staging and Management of Prostate Cancer - American Urological Association. [cited 14 Jul 2025]. Available: https://www.auanet.org/guidelines-and-quality/guidelines/other-clinical-guidance/mri-of-the-prostate-sop

48. Hofman MS, Lawrentschuk N, Francis RJ, Tang C, Vela I, Thomas P, et al. Prostate-specific membrane antigen PET-CT in patients with high-risk prostate cancer before curative-intent surgery or radiotherapy (proPSMA): a prospective, randomised, multicentre study. Lancet. 2020;395: 1208–1216.

49. El-Amm J, Aragon-Ching JB. The changing landscape in the treatment of metastatic castration-resistant prostate cancer. Ther Adv Med Oncol. 2013;5: 25–40.

50. Attard G, Belldegrun AS, de Bono JS. Selective blockade of androgenic steroid synthesis by novel lyase inhibitors as a therapeutic strategy for treating metastatic prostate cancer. BJU Int. 2005;96: 1241–1246.

51. Li Z, Bishop AC, Alyamani M, Garcia JA, Dreicer R, Bunch D, et al. Conversion of abiraterone to D4A drives anti-tumour activity in prostate cancer. Nature. 2015;523: 347–351.

52. Schalken J, Fitzpatrick JM. Enzalutamide: targeting the androgen signalling pathway in metastatic castration-resistant prostate cancer. BJU Int. 2016;117: 215–225.

53. 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.

54. Beltran H, Tomlins S, Aparicio A, Arora V, Rickman D, Ayala G, et al. Aggressive variants of castration-resistant prostate cancer. Clin Cancer Res. 2014;20: 2846–2850.

55. Tang DG. Understanding and targeting prostate cancer cell heterogeneity and plasticity. Semin Cancer Biol. 2022;82: 68–93.

56. Watson PA, Arora VK, Sawyers CL. Emerging mechanisms of resistance to androgen receptor inhibitors in prostate cancer. Nat Rev Cancer. 2015;15: 701–711.

57. Carver BS, Tran J, Chen Z, Carracedo-Perez A, Alimonti A, Nardella C, et al. ETS rearrangements and prostate cancer initiation. Nature. 2009;457: E1; discussion E2–3.

58. 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.

59. Kumar-Sinha C, Tomlins SA, Chinnaiyan AM. Recurrent gene fusions in prostate cancer. Nat Rev Cancer. 2008;8: 497–511.

60. 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.

61. Cancer Genome Atlas Research Network. The Molecular Taxonomy of Primary Prostate Cancer. Cell. 2015;163: 1011–1025.

62. 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.

63. 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.

64. 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.

65. 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.

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. 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.

68. 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.
69. 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.
70. Asangani IA, Dommeti VL, Wang X, Malik R, Cieslik M, Yang R, et al. Therapeutic targeting of BET bromodomain proteins in castration-resistant prostate cancer. Nature. 2014;510: 278–282.
71. Aytes A, Giacobbe A, Mitrofanova A, Ruggero K, Cyrta J, Arriaga J, et al. NSD2 is a conserved driver of metastatic prostate cancer progression. Nat Commun. 2018;9: 5201.
72. Li JJ, Vasciaveo A, Karagiannis D, Sun Z, Chen X, Socciarelli F, et al. NSD2 maintains lineage plasticity and castration-resistance in neuroendocrine prostate cancer. bioRxivorg. 2023. doi:10.1101/2023.07.18.549585
73. Xiao L, Parolia A, Qiao Y, Bawa P, Eyunni S, Mannan R, et al. Targeting SWI/SNF ATPases in enhancer-addicted prostate cancer. Nature. 2022;601: 434–439.
74. 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.
75. Shen M, Demers L-K, Bailey SD, Labbé DP. To bind or not to bind: Cistromic reprogramming in prostate cancer. Front Oncol. 2022;12: 963007.
76. 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.
77. Topchu I, Pangeni RP, Bychkov I, Miller SA, Izumchenko E, Yu J, et al. The role of NSD1, NSD2, and NSD3 histone methyltransferases in solid tumors. Cell Mol Life Sci. 2022;79: 285.
78. Popovic R, Martinez-Garcia E, Giannopoulou EG, Zhang Q, Zhang Q, Ezponda T, et al. Histone methyltransferase MMSET/NSD2 alters EZH2 binding and reprograms the myeloma epigenome through global and focal changes in H3K36 and H3K27 methylation. PLoS Genet. 2014;10: e1004566.
79. Lauring J, Abukhdeir AM, Konishi H, Garay JP, Gustin JP, Wang Q, et al. The multiple myeloma associated MMSET gene contributes to cellular adhesion, clonogenic growth, and tumorigenicity. Blood. 2008;111: 856–864.

80. Asangani IA, Ateeq B, Cao Q, Dodson L, Pandhi M, Kunju LP, et al. Characterization of the EZH2-MMSET histone methyltransferase regulatory axis in cancer. Mol Cell. 2013;49: 80–93.
81. Zaret KS, Carroll JS. Pioneer transcription factors: establishing competence for gene expression. Genes Dev. 2011;25: 2227–2241.
82. Cirillo LA, Zaret KS. Specific interactions of the wing domains of FOXA1 transcription factor with DNA. J Mol Biol. 2007;366: 720–724.
83. Seachrist DD, Anstine LJ, Keri RA. FOXA1: A pioneer of nuclear receptor action in breast cancer. Cancers (Basel). 2021;13: 5205.
84. 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.
85. Fu X, Pereira R, De Angelis C, Veeraraghavan J, Nanda S, Qin L, et al. FOXA1 upregulation promotes enhancer and transcriptional reprogramming in endocrine-resistant breast cancer. Proc Natl Acad Sci U S A. 2019. doi:10.1073/pnas.1911584116
86. Arruabarrena-Aristorena A, Maag JLV, Kittane S, Cai Y, Karthaus WR, Ladewig E, et al. FOXA1 Mutations Reveal Distinct Chromatin Profiles and Influence Therapeutic Response in Breast Cancer. Cancer Cell. 2020. doi:10.1016/j.ccell.2020.08.003
87. Eyunni S, Mannan R, Zhang Y, Young E, Zhang Q, Luo J, et al. Divergent FOXA1 mutations drive prostate tumorigenesis and therapy-resistant cellular plasticity. Science. 2025; eadv2367.

Chapter 2

NSD2 is a Requisite Subunit of the AR/FOXA1 Neo-Enhanceosome in Promoting Prostate Tumorigenesis[1]

Abstract

The androgen receptor (AR) is a ligand-responsive transcription factor that binds at enhancers to drive terminal differentiation of the prostatic luminal epithelia. By contrast, in tumors originating from these cells, AR chromatin occupancy is extensively reprogrammed to drive hyper-proliferative, metastatic, or therapy-resistant phenotypes, the molecular mechanisms of which remain poorly understood. Here, we show that the tumor-specific enhancer circuitry of AR is critically reliant on the activity of Nuclear Receptor Binding SET Domain Protein 2 (NSD2), a histone 3 lysine 36 di-methyltransferase. NSD2 expression is abnormally gained in prostate cancer cells, and its functional inhibition impairs AR trans-activation potential through partial off-loading from over 40,000 genomic sites, which is greater than 65% of the AR tumor cistrome. The NSD2-dependent AR sites distinctly harbor a chimeric AR-half motif juxtaposed to a FOXA1 element. Similar chimeric motifs of AR are absent at the NSD2-independent AR enhancers and instead contain the canonical palindromic motifs. Meta-analyses of AR cistromes from patient tumors uncovered chimeric AR motifs to exclusively participate in tumor-specific enhancer circuitries, with a minimal role in the physiological activity of AR. Accordingly, NSD2 inactivation attenuated hallmark cancer phenotypes that were fully reinstated upon exogenous NSD2 re-expression. Inactivation of NSD2 also engendered increased dependency on its paralog NSD1, which independently maintained AR and MYC hyper-transcriptional programs in cancer cells. Concordantly, a dual NSD1/2 PROTAC degrader, called LLC0150, was preferentially cytotoxic in AR-dependent prostate cancer as well as NSD2-altered hematologic malignancies. Altogether, we identify NSD2 as a novel subunit of the AR *neo*-enhanceosome that wires prostate cancer gene expression programs, positioning NSD1/2 as viable paralog co-targets in advanced prostate cancer.

[1]Contents of this chapter were previously published as part of the following manuscript: Parolia A, **Eyunni S**, et al. NSD2 is a requisite subunit of the AR/FOXA1 neo-enhanceosome in promoting prostate tumorigenesis. *Nature Genetics*. 2024;56: 2132-2143.

Introduction

Functional CRISPR screen reveals NSD2 as an AR co-activator

Conventional plasmid-based reporter systems fail to capture intricate epigenetic or chromatin-level regulation of gene expression as they lack the native histone composition or higher-order chromosomal structure. Thus, we engineered an endogenous AR reporter system by using the CRISPR/Cas9 and homologous recombination methodologies. We edited the *KLK3* gene (also known as prostate-specific antigen, PSA) locus in AR-driven LNCaP cells to knock-in the mCherry coding sequence directly downstream of the endogenous promoter and fused in-frame via an endopeptidase sequence to the *KLK3* gene (**Figure. 2-1A** and **Supplementary Figure. A2-1A-C**). In the monoclonal reporter cell line, akin to PSA, mCherry expression is directly regulated by the AR transcriptional complex (**Supplementary Figure. A2-1D-F**) and, most importantly, captures chromatin or epigenetic-level changes in AR transactivation potential. Like PSA, mCherry expression was attenuated upon pharmacologic inhibition of co-activators like BRD4[16], SWI/SNF[18], or P300/CBP[19], while increasing upon inhibition of the repressive PRC2/EZH2 complex[32,33] (**Figure. 2-1B** and **Supplementary Figure. A2-1G**). Using these endogenous AR reporter cell lines, we carried out a functional CRISPR screen, wherein we treated the cells with a custom single guide RNA (sgRNA) library targeting druggable transcriptional cofactors[34] (**Supplementary Figure. A2-1H**) for 8 days, stimulated with DHT for 16h, and FACS-sorted into $mCherry^{HIGH}$ and $mCherry^{LOW}$ populations. Genomic sgRNAs were sequenced and the ratio of normalized counts in $mCherry^{LOW}$ to $mCherry^{HIGH}$ cell populations was used to rank individual sgRNAs. Here, ranked alongside BRD4[16] and TRIM24[35,36], we identified NSD2 as an AR coactivator (**Figure. 2-1C** and **Supplementary Figure. A2-1I**). In contrast, subunits of the PRC2 complex, namely EZH2 and JARID2, that repress AR activity[32] were enriched in the $mCherry^{HIGH}$ cells. Validating the screening results, siRNA-mediated knockdown of NSD2 attenuated the expression of PSA/KLK3 in PCa cell lines (**Figure. 2-1D**).

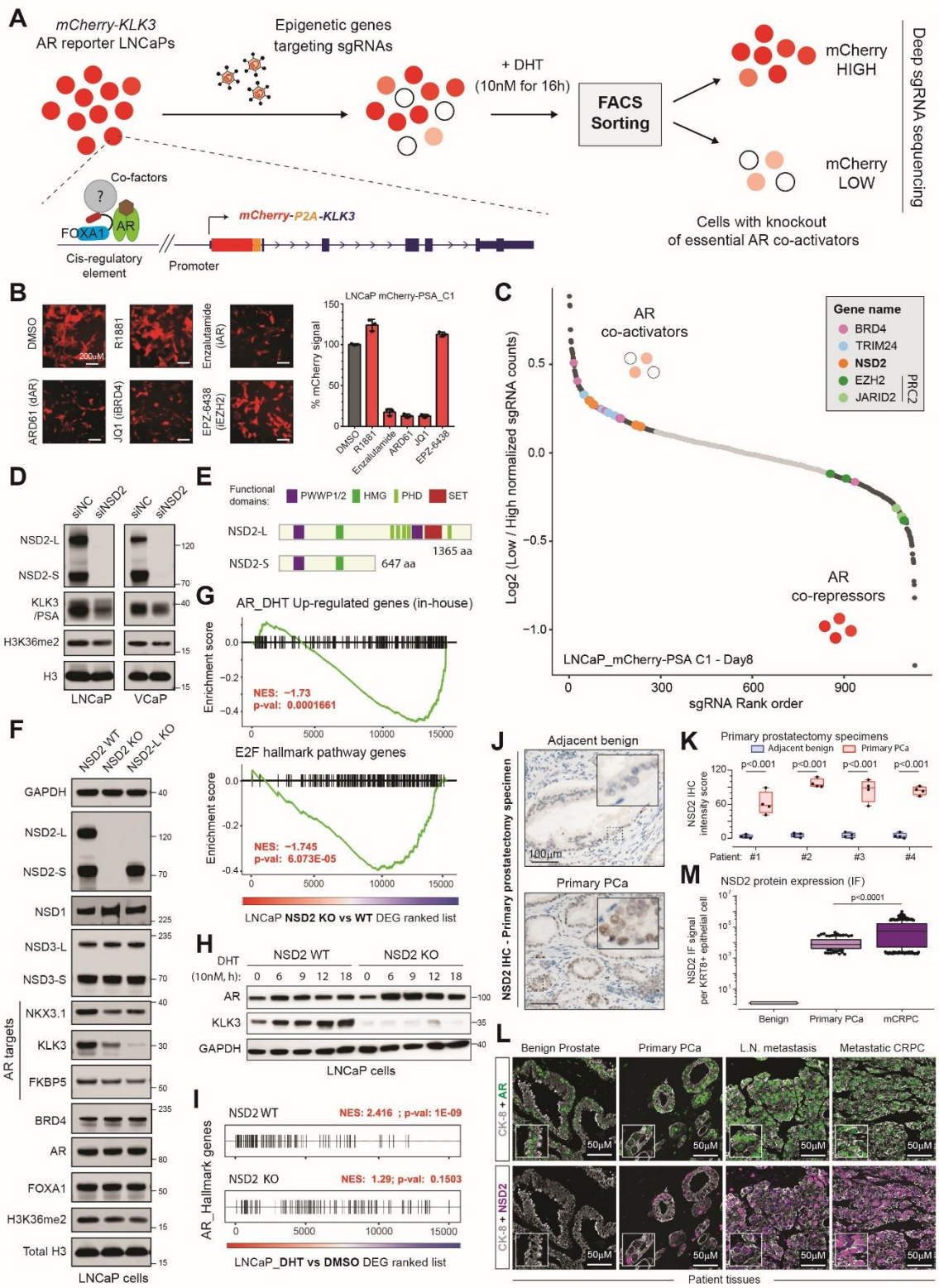

A
mCherry-KLK3
AR reporter LNCaPs
Epigenetic genes
targeting sgRNAs
+ DHT
(10nM for 16h)
FACS
Sorting
mCherry
HIGH
mCherry
LOW
Deep sgRNA sequencing
Co-factors
FOXA1
AR
Cis-regulatory
element
Promoter
mCherry-P2A-KLK3
Cells with knockout
of essential AR co-activators
B
DMSO
R1881
Enzalutamide
(iAR)
ARD61 (dAR)
JQ1 (iBRD4)
EPZ-6438
(iEZH2)
200μM
LNCaP mCherry-PSA_C1
% mCherry signal
C
AR
co-activators
AR
co-repressors
Gene name
BRD4
TRIM24
NSD2
EZH2
JARID2
PRC2
Log2 (Low / High normalized sgRNA counts)
LNCaP_mCherry-PSA C1 - Day8
sgRNA Rank order
D
siNC
siNSD2
NSD2-L
NSD2-S
KLK3
/PSA
H3K36me2
H3
LNCaP
VCaP
E
Functional
domains:
PWWP1/2
HMG
PHD
SET
1365 aa
647 aa
F
NSD2 WT
NSD2 KO
NSD2-L KO
GAPDH
NSD1
NSD3-L
NSD3-S
AR targets
NKX3.1
KLK3
FKBP5
BRD4
AR
FOXA1
Total H3
LNCaP cells
G
AR_DHT Up-regulated genes (in-house)
Enrichment score
NES: −1.73
p-val: 0.0001661
E2F hallmark pathway genes
NES: −1.745
p-val: 6.073E-05
LNCaP NSD2 KO vs WT DEG ranked list
H
DHT
(10nM, h):
LNCaP cells
I
AR_Hallmark genes
NSD2 WT
NES: 2.416 ; p-val: 1E-09
NSD2 KO
NES: 1.29; p-val: 0.1503
LNCaP_DHT vs DMSO DEG ranked list
J
Adjacent benign
Primary PCa
NSD2 IHC - Primary prostatectomy specimen
100μm
K
Primary prostatectomy specimens
Adjacent benign
Primary PCa
NSD2 IHC
intensity score
p<0.001
Patient: #1
#2
#3
#4
M
NSD2 protein expression (IF)
p<0.0001
NSD2 IF signal
per KRT8+ epithelial cell
Benign
Primary PCa
mCRPC
L
Benign Prostate
Primary PCa
L.N. metastasis
Metastatic CRPC
CK-8 + AR
CK-8 + NSD2
50μM
Patient tissues

Figure 2-1: Epigenetics-focused CRISPR screen shows NSD2 as an AR coactivator. A) Schematic of the epigenetic-targeted CRISPR screen using LNCaP mCherry-KLK3 AR reporter lines. B) *Left*: mCherry immunofluorescence images of LNCaP reporters treated with labeled epigenetic drugs. *Right*: Barplot showing quantification of the mCherry signal from treated reporter cells normalized to the DMSO treatment (n = 3 biological replicates). Mean+/- SEM are shown. Scale bar: 200μm. C) sgRNA enrichment rank plot based on guide RNA ratio in mCherry-LOW to mCherry-HIGH cells. D) Immunoblots of listed proteins upon treatment with control (siNC) or NSD2-targeting (siNSD2) siRNAs. Total H3 is used as loading control. LNCaP lysates were collected at Day 15. VCaP lysates were collected at Day 10 or 15 post treatment. E) Representative protein map of NSD2-Long (NSD2-L) and NSD2-Short (NSD2-S) isoforms. HMG: High mobility group; PHD: Plant homeodomain. F) Immunoblots of noted proteins in CRISPR-mediated stable knock-out (KO) of both NSD2 isoforms or NSD2-L alone. Total H3 is used as loading control. G) GSEA plots for AR and E2F up-regulated genes using the fold change rank-ordered genes from the NSD2 knock-out (KO) vs wild-type (WT) LNCaP cells. DEGS, differentially expressed genes. (n = 2 biological replicates; GSEA enrichment test). H) Immunoblots of listed proteins in NSD2 KO LNCaP cells stimulated with 10nM DHT. I) GSEA plots of AR hallmark genes in NSD2 wild-type (WT) vs knock-out (KO) LNCaP cells using the fold change rank-ordered genes from DHT (10nM for 24h) vs DMSO treatment. DEGS, differentially expressed genes (n = 2 biological replicates; GSEA enrichment test). J) Representative immunohistochemistry (IHC) images of NSD2 in prostatectomy patient specimens. Scale bar:100μm. K) NSD2 signal intensity from IHC staining in panel j (n = 4 patient tumors; two-sided t-test). Box plot center, median; box, quartiles 1-3, whiskers, quartiles 1-3 ±1.5*interquartile range, dot, outliers. L) Representative multiplex immunofluorescence (IF) images of KRT8, AR, and NSD2 in benign prostate, primary prostate cancer, or metastatic castration-resistant prostate cancer (mCRPC) patient specimens. Scale bar:50μm. M) Quantification of NSD2 IF signal intensity per KRT8+ luminal epithelial cell from images in panel l (two-sided t-test; normal=39, primary PCa=145, mCRPC=381 nuclei). Box plot center, median; box, quartile 1-3; whiskers, 10th and 90th percentile; dot, outliers.

The NSD2 gene templates two splice isoforms producing a long, catalytically active form (hereafter referred to as NSD2-L) as well as a truncated shorter isoform (called NSD2-S) containing only the reader and protein-protein interacting PWWP and HMG domains, respectively. We found both NSD2 isoforms to be robustly expressed in PCa cells (**Figure. 2-1E** and **Supplementary Figure. A2-2A**). Deletion of NSD2-L alone strongly attenuated the expression of AR target genes in LNCaP cells, which was comparable to complete loss of the NSD2 protein (**Figure. 2-1F**). The transcriptomic analysis further showed global AR activity to be significantly dampened in NSD2-deficient LNCaP cells with a parallel loss in hyper-proliferative gene expression programs (**Figure. 2-1G**). AR and NSD2 transcriptional activities were also positively correlated in primary prostate tumors from the TCGA cohort (**Supplementary Figure. A2-2B;** R=0.68, p=2.2e-16). Notably, there was no change in the abundance of AR transcript or protein itself in NSD2-deleted cells (**Figure. 2-1F** and **Supplementary Figure. A2-2C**), yet stimulation with DHT failed to significantly up-regulate the expression of AR target genes (**Figure. 2-1H,I** and **Supplementary Figure. A2-2C,D**).

To date, several studies have implicated NSD2 in PCa[37–41]; however, it is worth noting that these studies were focused on the AR-negative disease. Using tissue microarrays, these studies showed NSD2 protein to be elevated in cancer specimens, showing a stage-wise increase from primary to metastatic castration-resistant or neuroendocrine PCa[39,40]. Building on these findings, in primary prostatectomy specimens we found NSD2 levels to be undetectable in the normal or adjacent benign foci with marked gain in expression in malignant cells (**Figure. 2-1J,K**). Consistent with this, in single-cell RNA-seq data from patient tumors, we found the *NSD2* transcript to be exclusively expressed in the AR-positive luminal epithelial cells (**Supplementary Figure. A2-2E**). Pseudo-bulk analyses confirmed NSD2 expression to be markedly elevated in the matched tumor vs the normal luminal compartment (n=18), and NSD2 expression positively correlated with Gleason score of the primary disease (**Supplementary Figure. A2-2F,G**). Multiplex immunofluorescence in additional prostatectomy and patient tumor specimens further confirmed the KRT8+/AR+ normal epithelial cells to have no detectable expression of NSD2, which was robustly expressed in the transformed epithelial cells (**Figure. 2-1L,M** and **Supplementary Figure. A2-3A-E**). Altogether, this data suggests that NSD2 is abnormally expressed in the transformed prostate luminal epithelial cells, wherein its methyltransferase function is critical for maintaining transcriptional activity of the AR complex.

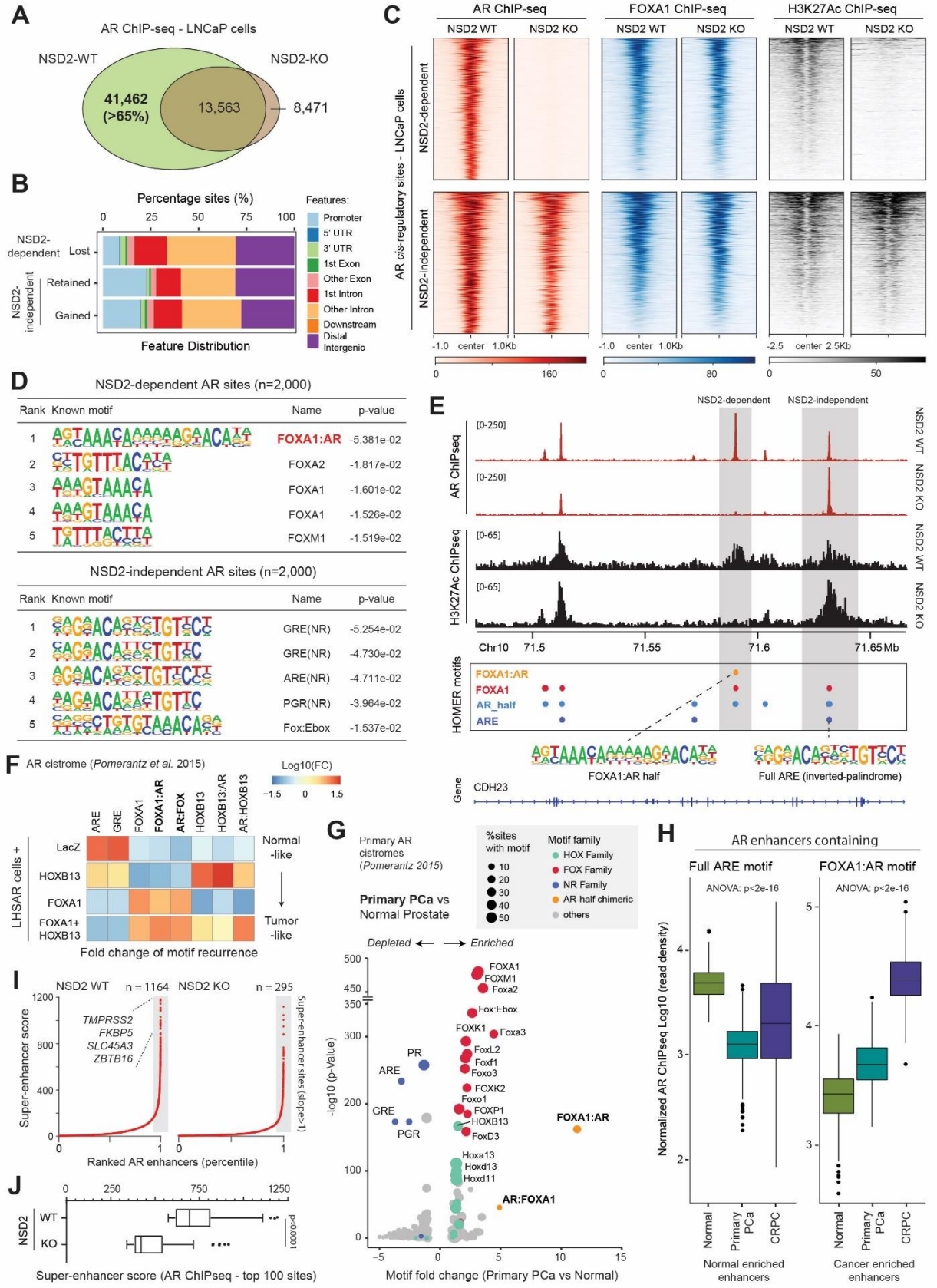

A
AR ChIP-seq - LNCaP cells
NSD2-WT
NSD2-KO
41,462 (>65%)
13,563
8,471
B
Percentage sites (%)
0 25 50 75 100
NSD2-dependent
Lost
NSD2-independent
Retained
Gained
Feature Distribution
Features:
Promoter
5' UTR
3' UTR
1st Exon
Other Exon
1st Intron
Other Intron
Downstream
Distal Intergenic
C
AR ChIP-seq
FOXA1 ChIP-seq
H3K27Ac ChIP-seq
NSD2 WT
NSD2 KO
AR cis-regulatory sites - LNCaP cells
NSD2-dependent
NSD2-independent
-1.0 center 1.0Kb
-2.5 center 2.5Kb
0
160
80
50
D
NSD2-dependent AR sites (n=2,000)
Rank Known motif Name p-value
1 FOXA1:AR -5.381e-02
2 FOXA2 -1.817e-02
3 FOXA1 -1.601e-02
4 FOXA1 -1.526e-02
5 FOXM1 -1.519e-02
NSD2-independent AR sites (n=2,000)
Rank Known motif Name p-value
1 GRE(NR) -5.254e-02
2 GRE(NR) -4.730e-02
3 ARE(NR) -4.711e-02
4 PGR(NR) -3.964e-02
5 Fox:Ebox -1.537e-02
E
NSD2-dependent
NSD2-independent
AR ChIPseq
H3K27Ac ChIPseq
[0-250]
[0-65]
Chr10 71.5 71.55 71.6 71.65 Mb
HOMER motifs
FOXA1:AR
FOXA1
AR_half
ARE
FOXA1:AR half
Full ARE (inverted-palindrome)
Gene
CDH23
F
AR cistrome (Pomerantz et al. 2015)
Log10(FC)
−1.5 0 1.5
ARE
GRE
FOXA1
FOXA1:AR
AR:FOX
HOXB13
HOXB13:AR
AR:HOXB13
LHSAR cells +
LacZ
HOXB13
FOXA1
FOXA1+ HOXB13
Normal -like
Tumor -like
Fold change of motif recurrence
G
Primary AR cistromes (Pomerantz 2015)
Primary PCs vs Normal Prostate
%sites with motif
10
20
30
40
50
Motif family
HOX Family
FOX Family
NR Family
AR-half chimeric
others
Depleted
Enriched
-log10 (p-Value)
FOXA1
FOXM1
Foxa2
Fox:Ebox
FOXK1
Foxa3
FoxL2
Foxf1
Foxo3
FOXK2
Foxo1
FOXP1
HOXB13
FoxD3
Hoxa13
Hoxd13
Hoxd11
PR
ARE
GRE
PGR
FOXA1:AR
AR:FOXA1
Motif fold change (Primary PCa vs Normal)
H
AR enhancers containing
Full ARE motif
FOXA1:AR motif
ANOVA: p<2e-16
Normalized AR ChIPseq Log10 (read density)
Normal
Primary PCa
CRPC
Normal enriched enhancers
Cancer enriched enhancers
I
NSD2 WT n = 1164
NSD2 KO n = 295
Super-enhancer score
TMPRSS2
FKBP5
SLC45A3
ZBTB16
Super-enhancer sites (slope<1)
Ranked AR enhancers (percentile)
J
NSD2
WT
KO
p<0.0001
Super-enhancer score (AR ChIPseq - top 100 sites)

Figure 2-2: NSD2 expands the AR neo-enhancer circuitry to include chimeric AR half-sites. A) Venn diagram showing overlaps of AR ChIP-seq peaks in NSD2 wild-type (WT) and knock-out (KO) LNCaP cell lines. B) Genomic location of NSD2-dependent and independent AR sites defined from the overlap analysis in panel a. C) ChIP-seq read-density heatmaps of AR, FOXA1, and H3K27ac at top 1,000 AR enhancer sites in LNCaP NSD2 WT and KO cell lines. D) Top five known HOMER motifs enriched within NSD2-dependent and independent AR sites in LNCaP cells (HOMER, hypergeometric test). E) ChIP-seq read-density tracks of AR and H3K27ac in NSD2 WT and KO LNCaP cell lines. HOMER motifs detected within AR peaks are shown below with grey boxes highlighting NSD2 dependent and independent AR elements. F) Fold change heatmap of HOMER motifs enrichment within AR binding sites specific to HOXB13, FOXA1, or FOXA1+HOXB13 overexpression in LHSAR cells (data from Pomerantz et al, 2015). G) Fold change and significance of HOMER motifs enriched within primary prostate cancer-specific AR sites over normal AR enhancers (data from Pomerantz et al., 2015; HOMER, hypergeometric test). H) AR ChIP-seq read density boxplot at sites containing the ARE or the FOXA1:AR chimeric motif in primary normal and tumor patient samples (normal prostate, n=7; primary prostate cancer, n=13; mCRPC, n=15). In box plots, the center line shows the median, box edges mark quartiles 1-3, and whiskers span quartiles 1-3 ± 1.5 * interquartile range (one-way ANOVA). I) Rank-ordered plot of AR super-enhancers (HOMER ROSE algorithm) in NSD2 WT and KO LNCaP cells with select known AR target genes shown. J) Boxplot of AR super-enhancer scores (HOMER ROSE algorithm) of top 100 *cis*-elements in NSD2 WT or KO LNCaP cells (two-sided t-test). Box plot centre, median; box, quartiles 1-3; whiskers, quartiles 1-3 ± 1.5 × interquartile range; dot, outliers.

NSD2 activates neo-enhancers with chimeric AR-half motifs

Given that NSD2 loss had no impact on the abundance of the AR protein, we next profiled AR binding on chromatin. AR ChIP-seq in NSD2-deficient LNCaP cells showed a dramatic and complete off-loading of the AR protein from over 40,000 genomic sites that comprise over 65% of the tumor cistrome (**Figure. 2-2A**). The majority of the lost sites (i.e., NSD2-dependent) were within intronic or intergenic regions associated with *cis*-regulatory DNA elements (**Figure. 2-2B**). At these sites there was no change in the binding of FOXA1 upon NSD2 inactivation. Yet, disruption of AR binding was sufficient to trigger loss of the H3K27ac mark that demarcates active enhancers (**Figure. 2-2C**). In contrast, AR remains bound at over 20,000 genomic sites independent of NSD2, which also retained the H3K27ac active mark in the NSD2-null PCa cells (**Fig. 2-2A-C**). Next, ChIP-seq-based profiling of the chemical chromatin state showed NSD2-dependent AR sites to have higher abundance of H3K36me2 as well as active enhancer-associated H3K4me1/2 and H3K27ac modifications compared to the NSD2-independent elements (**Supplementary Figure. A2-3D-F**). Contrastingly, NSD2-independent AR sites had higher levels of the PRC2/EZH2 catalyzed repressive H3K27me3 mark. More importantly, NSD2 inactivation led to a significant decrease in H3K36me2 levels at the NSD2-dependent sites, reducing it to the levels at NSD2-independent sites in the wild-type cells, with a parallel increase in H3K27me3 (**Supplementary Figure. A2-3G,H**).

Motif analyses (HOMER[42]) of the NSD2-dependent AR sites identified a chimeric motif comprising a FOXA1 element juxtaposed to the AR-half site (called FOXA1:AR-half motif) as the most significantly enriched DNA sequence (**Figure. 2-2D**), with 40% of these enhancers harboring this motif. In contrast, NSD2-independent AR sites, a large fraction of which showed increased AR binding upon NSD2 inactivation, housed the palindromic androgen response element (ARE) (**Figure. 2-2D**). These distinct modes of AR DNA interaction were evident within a Chr10 gene locus, wherein the loss of NSD2 completely disrupted AR binding at the FOXA1:AR-half motif, without affecting AR's interaction with a canonical ARE element in *cis*-proximity (**Figure. 2-2E**). Furthermore, custom motif analyses showed enrichment of other transcription factor motifs, including HOXB13 and ETS, within 25bp of AR-half elements detected within the NSD2-dependent AR sites (**Supplementary Figure. A2-4A**). We also custom-assembled chimeric AR-half motifs with FOXA1 and HOXB13 elements (in both 5' and 3' confirmations, see Methods) and interrogated their recurrence in published AR ChIP-seq data derived from non-cancerous LHSAR cells[5]. Here, we found the overexpression of FOXA1 and HOXB13 alone, or in combination, to markedly shift the AR cistrome away from full AREs (normal-like) towards chimeric AR-half elements in the tumor-like state (**Figure. 2-2F** and **Supplementary Figure. A2-4B**). Strikingly, motif analysis of the AR cistromes generated from patient specimens[5,6] revealed the FOXA1:AR-half motif to be exclusively detected in the tumor-specific AR enhancer circuitries, with such chimeric motifs being essentially absent at normal AR sites (**Figure. 2-2G** and **Supplementary Figure. A2-4C-F**). In these analyses, we also found palindromic AREs to be depleted within cancer-specific enhancers (**Figure. 2-2G**). Concordantly, AR ChIP-seq signal at ARE sites was strongest in normal prostate tissues, while enhancers containing the chimeric FOXA1:AR-half motif had higher AR binding in PCa specimens (**Figure. 2-2H**). H3K27ac ChIP-seq signal from matched tumors showed similar redistribution, with FOXA1:AR-half sites being strongly activated in mCRPC tumors (**Supplementary Figure. A2-4G**).

In tumor cells, the aberrant expression of oncogenes is frequently amplified through dense clusters of closely spaced enhancers, often referred to as super-enhancers[43]. NSD2 inactivation resulted in the loss of over 75% of the AR-bound super-enhancers in PCa (**Figure. 2-2I**), including those that are hijacked by activating translocations[44] (**Supplementary Figure. A2-4H**). The residual super-enhancers also showed a significant decrease in the enhancer strength in the NSD2-null relative to

the wild-type LNCaP cells (**Figure. 2-2J**). Altogether, these data suggest that, upon ectopic expression, NSD2 assists oncogenic transcription factors (namely FOXA1 and HOXB13) in expanding the AR enhancer circuitry to include chimeric AR-half sites that constitute over two-thirds of PCa AR cistromes.

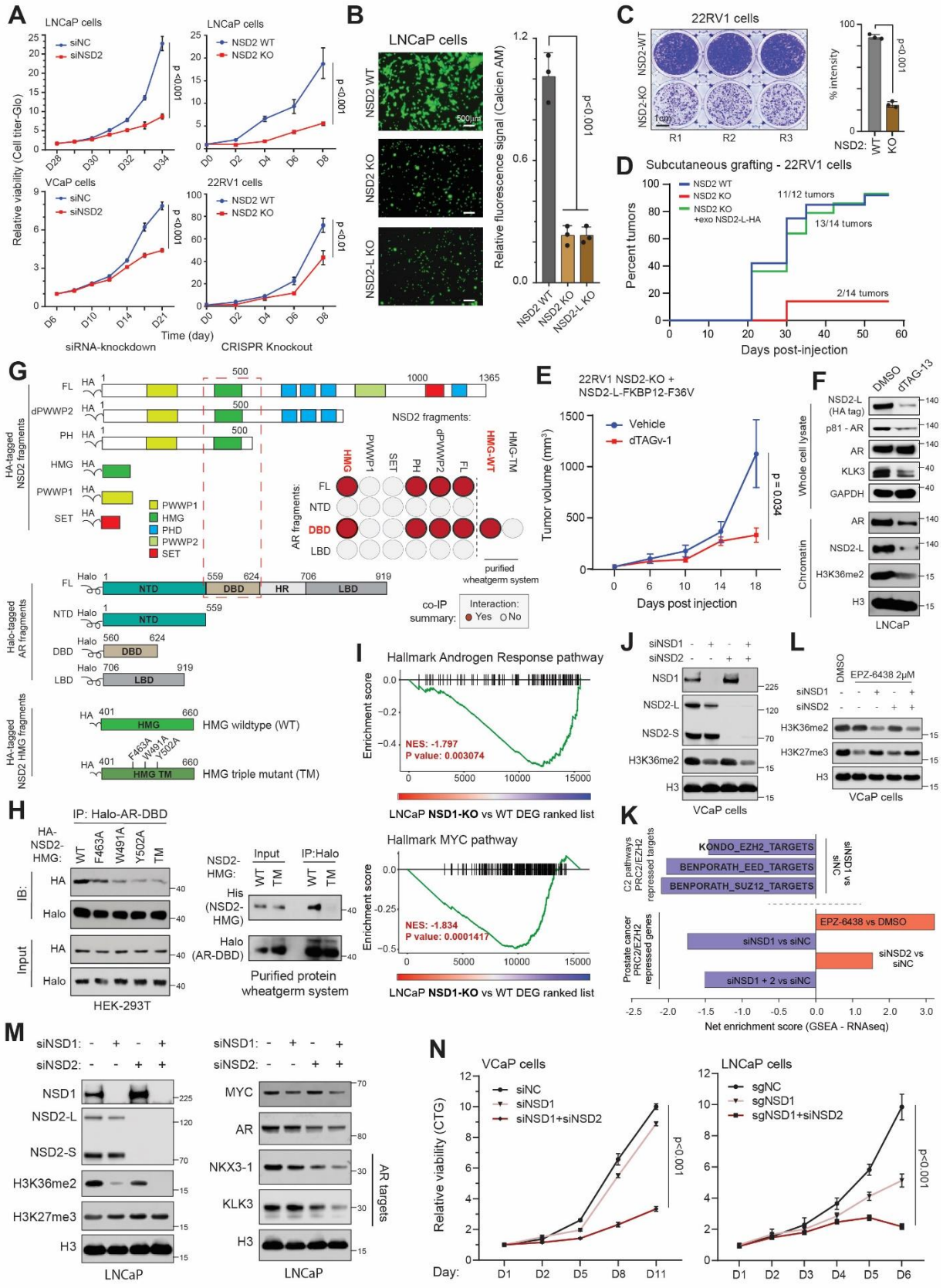

A
LNCaP cells
siNC
siNSD2
Relative viability (Cell titer-Glo)
p<0.001
LNCaP cells
NSD2 WT
NSD2 KO
p<0.001
VCaP cells
siNC
siNSD2
p<0.001
22RV1 cells
NSD2 WT
NSD2 KO
p<0.01
Time (day)
siRNA-knockdown
CRISPR Knockout
B
LNCaP cells
NSD2 WT
NSD2 KO
NSD2-L KO
500μm
Relative fluorescence signal (Calcien AM)
p<0.001
NSD2 WT
NSD2 KO
NSD2-L KO
C
22RV1 cells
NSD2-WT
NSD2-KO
1cm
R1
R2
R3
% intensity
p<0.001
NSD2:
WT
KO
D
Subcutaneous grafting - 22RV1 cells
NSD2 WT
NSD2 KO
NSD2 KO +exo NSD2-L-HA
11/12 tumors
13/14 tumors
2/14 tumors
Percent tumors
Days post-injection
G
HA-tagged NSD2 fragments
FL
dPWWP2
PH
HMG
PWWP1
SET
PWWP1
HMG
PHD
PWWP2
SET
NSD2 fragments:
HMG
PWWP1
SET
PH
dPWWP2
FL
HMG-WT
HMG-TM
AR fragments:
FL
NTD
DBD
LBD
purified wheatgerm system
co-IP summary:
Interaction:
Yes
No
Halo-tagged AR fragments
NTD
DBD
LBD
HA-tagged NSD2 HMG fragments
HMG wildtype (WT)
HMG triple mutant (TM)
F463A
W491A
Y502A
E
22RV1 NSD2-KO + NSD2-L-FKBP12-F36V
Vehicle
dTAGv-1
Tumor volume (mm³)
p = 0.034
Days post injection
F
DMSO
dTAG-13
NSD2-L (HA tag)
p81 - AR
AR
KLK3
GAPDH
Whole cell lysate
AR
NSD2-L
H3K36me2
H3
Chromatin
LNCaP
H
IP: Halo-AR-DBD
HA-NSD2-HMG:
WT
F463A
W491A
Y502A
TM
IB:
HA
Halo
Input
HEK-293T
NSD2-HMG:
Input
IP:Halo
His (NSD2-HMG)
Halo (AR-DBD)
Purified protein wheatgerm system
I
Hallmark Androgen Response pathway
Enrichment score
NES: -1.797
P value: 0.003074
LNCaP NSD1-KO vs WT DEG ranked list
Hallmark MYC pathway
NES: -1.834
P value: 0.0001417
LNCaP NSD1-KO vs WT DEG ranked list
J
siNSD1
siNSD2
NSD1
NSD2-L
NSD2-S
H3K36me2
H3
VCaP cells
L
DMSO
EPZ-6438 2μM
siNSD1
siNSD2
H3K36me2
H3K27me3
H3
VCaP cells
K
C2 pathways PRC2/EZH2 repressed targets
KONDO_EZH2_TARGETS
BENPORATH_EED_TARGETS
BENPORATH_SUZ12_TARGETS
siNC vs siNSD1
Prostate cancer PRC2/EZH2 repressed genes
EPZ-6438 vs DMSO
siNSD1 vs siNC
siNSD2 vs siNC
siNSD1 + 2 vs siNC
Net enrichment score (GSEA - RNAseq)
M
siNSD1:
siNSD2:
NSD1
NSD2-L
NSD2-S
H3K36me2
H3K27me3
H3
LNCaP
MYC
AR
NKX3-1
KLK3
H3
AR targets
LNCaP
N
VCaP cells
siNC
siNSD1
siNSD1+siNSD2
Relative viability (CTG)
p<0.001
Day:
LNCaP cells
sgNC
sgNSD1
sgNSD1+siNSD2
p<0.001

Figure 2-3: NSD1 and NSD2 independently enable oncogenic AR activity. A) *Left*: Growth curves of cells treated with control (siNC) or NSD2-targeting siRNAs (n=6 biological replicates; two-sided t-test). *Right*: Growth curves of NSD2 knockout (KO) or wildtype (WT) cells (n = 3 biological replicates; two-sided t-test). Mean+/-SEM are shown. B) *Left*: Boyden chamber images of NSD2 KO and WT cells. Scale bar: 500µm. *Right*: Quantification of fluorescence signal (n=3 biological replication; one-way ANOVA+Tukey's test). Mean+/-SEM are shown. C) *Left*: Representative images of NSD2 KO and WT 22RV1 cell colonies (n=3 biological replicates). Scale bar: 1cm. *Right*: Staining intensity of cell colonies (two-sided t-test). Mean+/-SEM are shown. D) Reverse Kaplan-Meier plot of tumor grafting of 22RV1 WT, NSD2 KO, or NSD2 KO + NSD2-L cells. E) Tumor volumes of 22RV1 NSD2-KO+NSD2-L-FKBP12^{F36V} xenografts +/- dTAGv-1 treatment. Mean ± SEM is shown (n = 10 biological replicates; two-sided t-test). F) Immunoblots of listed proteins in whole-cell or chromatin fractions of LNCaP NSD2-FKBP12^{F36V} cells +/- dTAG-13. G) Schematic of co-immunoprecipitation (coIP) protein fragments. Dashed red box marks interacting domains. *Inset*: AR-NSD2 co-IP interaction summary. Red circles, interaction. Gray circles, no detectable binding. H) *Left*: co-IP immunoblots of AR DNA-binding domain (DBD) with HA-NSD2-HMG mutants. TM, triple mutant. *Right*: co-IP immunoblots of wheatgerm-purified Halo-AR-DBD with His-NSD2-HMG fragments. Input fractions are shown as control. I) GSEA plots for AR and MYC target genes in NSD1 KO vs WT LNCaP cells. DEGS, differentially expressed genes (n=2 biological replicates; GSEA enrichment test). J) Immunoblots of labeled proteins upon treatment with siNC or NSD1 and/or NSD2 targeting siRNAs (siNSD1 or siNSD2). H3 is a loading control. K) *Top*: GSEA enrichment scores of EZH2/PRC2-repressed genesets in siNSD1 versus siNC-treated cells. *Bottom*: GSEA enrichment scores of prostate cancer-specific EZH2 signature in siNSD1 and/or siNSD2 vs siNC treated cells. L) Immunoblots of noted proteins in siNSD1 and/or siNSD2 treated cells +/- EPZ-6438. M) Immunoblot of listed proteins in siNC or siNSD1 and/or siNSD2 treated cells. N) *Left*: Growth curves of cells treated with siNC, siNSD1, or siNSD1+NSD2. *Right*: Growth curves of control (sgNC) or NSD1-deficient (sgNSD1) cells +/- siNSD2 treatment. (n=5 biological replicates; two-sided t-test). Mean+/-SEM are shown.

NSD1 and NSD2 independently enable oncogenic AR activity

Given NSD2 inactivation resulted in disruption of the cancer-specific AR cistrome, we set out to phenotypically characterize NSD2-deficient PCa cells. Here, siRNA/shRNA knockdown or CRISPR knockout of NSD2 significantly impaired hyper-proliferative ability of AR-positive PCa cell lines (**Figure. 2-3A** and **Supplementary Figure. A2-5A,B**). NSD2-deficient cells also lost their ability to invade through Matrigel (**Figure. 2-3B**) and form colonies starting from single cells in clonogenic assays (**Figure. 2-3C** and **Supplementary Figure. A2-5C**). NSD2-null 22RV1 cells also lost their ability to graft when injected subcutaneously in NOD/SCID mice (**Figure. 2-3D**). Strikingly, exogenous reintroduction of NSD2-L restored the xenografting potential (green line, **Figure. 2-3D**), with resulting tumors growing at a rate comparable to those established with the parental wild-type cells (**Supplementary Figure. A2-5D**). NSD2-L re-expression also restored the invasive ability (**Supplementary Figure. A2-5E**), along with restoring the expression of KLK3 (**Supplementary Figure. A2-5F**). In the same experiment, re-expression of NSD2 variant lacking the SET domain (dSET) failed to rescue KLK3 (**Supplementary Figure. A2-5G**), while

expression of the hyper-active NSD2-E1099K SET-domain mutant[29] completely restored KLK3 levels (**Supplementary Figure. A2-5H**). Next, we engineered the NSD2-null 22RV1 cells to stably express dTAG-version of the NSD2-L protein fused to the $FKBP12^{F36V}$ tag[45], which is rapidly degraded upon treatment with an FKBP12 degrader (**Supplementary Figure. A2-5I**). The 22RV1 NSD2 KO + NSD2-L-$FKBP12^{F36V}$ cells successfully grafted and grew to form tumors *in vivo*; however, dosing of host animals with an FKBP12 degrader significantly diminished the growth of tumor xenografts (**Figure. 2-3E** and **Supplementary Figure. A2-5J**). Even at the molecular level, degradation of the exogenous NSD2-L-$FKBP12^{F36V}$ protein resulted in lower levels of KLK3 and chromatin-bound (i.e., p-S81) AR without a decrease in total AR expression (top panel, **Figure. 2-3F**). Concordantly, chromatin fractionation in these cells showed a marked loss of AR binding with a parallel decrease in H3K36me2 upon NSD2-L-$FKBP12^{F36V}$ degradation (bottom panel, **Figure. 2-3F**). In LNCaP NSD2-dTAG models, NSD2 degradation led to downregulation of multiple AR target genes in a time-dependent manner (**Supplementary Figure. A2-5K**). Also, other genes encoded in *cis*-proximity of NSD2-dependent AR sites (see Methods) were similarly down-regulated in the NSD2-deficient LNCaP cells (**Supplementary Figure. 2-6A**), and were enriched for oncogenic KRAS, angiogenesis, and G2-M checkpoint pathways (**Supplementary Figure. A2-6B**). In contrast, genes associated with NSD2-independent AR sites were enriched for developmental pathways and AR/NKX3-1 signaling (**Supplementary Figure. A2-6C**). These findings position NSD2 as a molecular "switch" that activates oncogenic AR cistrome and enables hallmark cancer properties.

Next, in size exclusion chromatography we found NSD2 to co-elute with higher order AR transcriptional complexes (**Supplementary Figure. A2-7A**). NSD2 also co-precipitated with AR in several PCa cell lines (**Supplementary Figure. A2-7B**). As previously reported[46], using fragment-based co-immunoprecipitation, we confirmed the high mobility group box (HMG-box) domain of NSD2 to interact with the DNA binding domain (DBD) of AR (**Figure. 2-3G** and **Supplementary Figure. A2-7C,D**).Furthermore, alanine substitution of three highly conserved HMG-box residues (i.e., F463/W491/Y502A) individually or together (triple mutant, **Figure. 2-3G**) disrupted its interaction with the AR-DBD in the ectopic HEK293T as well as cell-free purified wheatgerm extract systems (**Figure. 2-3H**). This suggests that NSD2 directly, and independent of DNA, interacts with the AR DBD through its HMG-box domain, which is notably absent in other NSD family histone methyltransferases.

Despite a striking loss of neoplastic features, NSD2-deficient PCa cells remained viable. Thus, we speculated if NSD2 paralogs could sustain AR activity through alternative mechanisms. To test this, we knocked-out NSD1 or NSD3 individually in LNCaP and assessed its transcriptional impact. Unlike NSD3, NSD1 loss significantly attenuated the AR and MYC gene programs (**Figure. 2-3I** and **Supplementary Figure. A2-7E,F**), in addition to reducing the AR protein levels (**Supplementary Figure. A2-7G**). NSD1 inactivation also diminished hyper-proliferative gene pathways (like E2F and G2M; **Supplementary Figure. A2-7H**) and had the strongest reduction in H3K36me2 levels upon a single-gene loss (**Figure. 2-3J** and **Supplementary Figure. A2-7G**), positioning NSD1 as the predominant H3K36 di-methyltransferase in PCa cells. NSD-catalyzed H3K36me2 mark was recently shown to sterically hinder loading of the H3K27 residue into catalytic pocket of the EZH2 enzyme[47], and NSD1 was reported to primarily antagonize the repressive PRC2/EZH2 complex[48]. Consistently, we found NSD1 loss to trigger a significant increase in EZH2/PRC2 activity, with several repressed target gene signatures being significantly downregulated upon NSD1 knockdown in VCaP cells (top panel, **Figure. 2-3K**). This was also confirmed using a PCa-specific PRC2 gene signature (bottom panel, **Figure. 2-3K**). Concordantly, treatment of siNSD1 cells with EPZ-6438 had substantially higher residual levels of the H3K27me3 mark relative to the control as well as NSD2-inactivated VCaP cells (**Figure. 2-3L**). In these experiments, loss of NSD2 alone had little to no effect on EZH2/PRC2 activity (**Figure. 2-3J-L**). These results position NSD1 as the primary writer of the H3K36me2 histone mark that counterbalances the EZH2/PRC2 repressive complex in PCa cells to maintain the hyper-transcriptional AR and MYC gene programs.

Interestingly, the loss of NSD2 led to a marked increase in NSD1 levels in PCa cells (**Figure. 2-3J,M**), likely suggesting that NSD1 could sustain residual oncogenic AR activity in these cells. Parallel inactivation of NSD1 and NSD2 in PCa cells resulted in the strongest decrease in H3K36me2 levels and AR target gene expression (**Figure. 2-3J,M**), triggering an accumulation of apoptotic marker cleaved-PARP (**Supplementary Figure. A2-7I**). Consistently, combined NSD1 and NSD2 inhibition resulted in significant cytotoxicity in AR-positive PCa cells (**Figure. 2-3N**), while inactivation of either genes alone had little to no tumor-killing effect in prostatic cell lines (**Supplementary Figure. A2-7J**). Altogether, these data suggest that NSD1 and NSD2, through distinct mechanisms, promote a hyper-transcriptional chromatin state or enable oncogenic AR activity, respectively, in PCa cells.

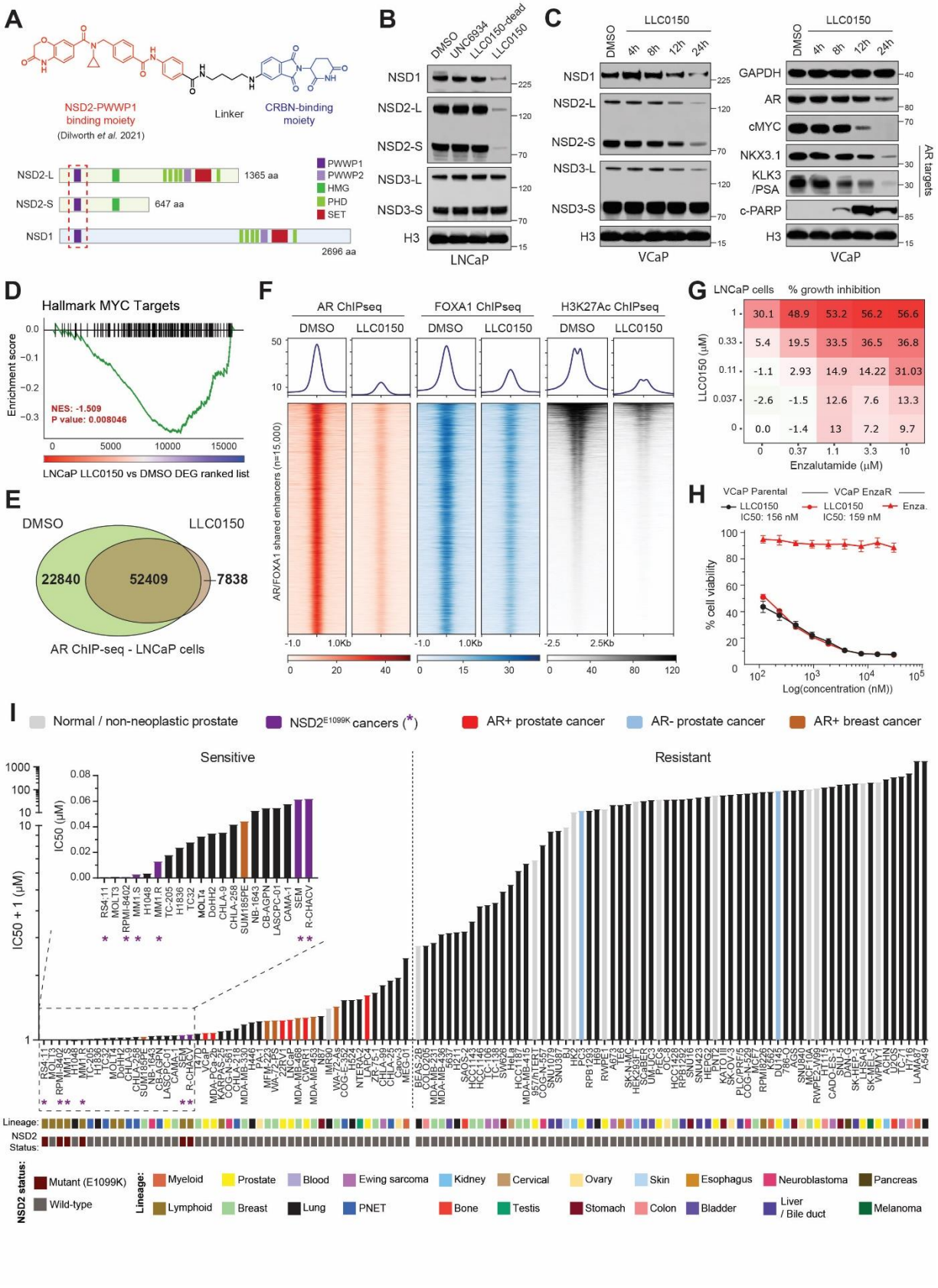

A
NSD2-PWWP1 binding moiety
(Dilworth et al. 2021)
Linker
CRBN-binding moiety
NSD2-L
1365 aa
NSD2-S
647 aa
NSD1
2696 aa
PWWP1
PWWP2
HMG
PHD
SET
B
DMSO
UNC6934
LLC0150-dead
LLC0150
NSD1
NSD2-L
NSD2-S
NSD3-L
NSD3-S
H3
LNCaP
C
LLC0150
DMSO
4h
8h
12h
24h
NSD1
NSD2-L
NSD2-S
NSD3-L
NSD3-S
H3
VCaP
GAPDH
AR
cMYC
NKX3.1
KLK3 /PSA
c-PARP
AR targets
D
Hallmark MYC Targets
Enrichment score
NES: -1.509
P value: 0.008046
LNCaP LLC0150 vs DMSO DEG ranked list
E
DMSO
LLC0150
22840
52409
7838
AR ChIP-seq - LNCaP cells
F
AR ChIPseq
FOXA1 ChIPseq
H3K27Ac ChIPseq
DMSO
LLC0150
AR/FOXA1 shared enhancers (n=15,000)
G
LNCaP cells
% growth inhibition
LLC0150 (μM)
Enzalutamide (μM)
H
VCaP Parental
VCaP EnzaR
LLC0150
IC50: 156 nM
LLC0150
IC50: 159 nM
Enza.
% cell viability
Log(concentration (nM))
I
Normal / non-neoplastic prostate
NSD2E1099K cancers (*)
AR+ prostate cancer
AR- prostate cancer
AR+ breast cancer
Sensitive
Resistant
IC50 + 1 (μM)
IC50 (μM)
Lineage:
NSD2 Status:
NSD2 status:
Mutant (E1099K)
Wild-type
Lineage:
Myeloid
Prostate
Blood
Ewing sarcoma
Kidney
Cervical
Ovary
Skin
Esophagus
Neuroblastoma
Pancreas
Lymphoid
Breast
Lung
PNET
Bone
Testis
Stomach
Colon
Bladder
Liver / Bile duct
Melanoma

Figure 2-4: LLC0150 is an NSD1/2 PROTAC with preferential cytotoxicity in AR-driven prostate cancer. A) Structure of LLC0150 and schema of NSD1 and NSD2 functional domains. LLC0150-binding PWWP1 domain is highlighted using a dashed red box. HMG: High mobility group; PHD: Plant homeodomain. **B)** Immunoblots of listed proteins in LNCaP cells treated with UNC6934 (warhead), LLC0150-dead (epimer control) or LLC0150 for 12h at 1uM. Total histone H3 is used as a loading control. **C)** Immunoblots of listed proteins in VCaP cells treated with LLC0150 (2uM) for increasing time durations. Total histone H3 is used as a loading control. **D)** GSEA plots of MYC target genes using the fold change rank-ordered genes from LLC0150 vs DMSO treated LNCaP cells. DEGS, differentially expressed genes (n = 2 biological replicates; GSEA enrichment test). **E)** Venn diagram showing the overlap of AR ChIP-seq peaks in LNCaP cells treated with LLC0150 (2uM for 48h) or DMSO as control. **F)** ChIP-seq read-density heatmaps of AR, FOXA1, and H3K27ac at enhancers that are co-bound by AR and FOXA1 in LNCaP cells plus/minus treatment with LLC0150 (2uM for 48h). **G)** Percent growth inhibition (Cell-titer Glo) of LNCaP cells upon co-treatment with varying concentrations of LLC0150 and enzalutamide. **H)** Dose-response curves of LLC0150 or enzalutamide in parental or enzalutamide-resistant VCaP cells. Data are presented as mean +/- SEM (n=2 biological replicates). Serving as a control, enzalutamide dose-response curve credentials the enzalutamide-resistant VCaP cell line. **I)** IC50 rank-order plot of over 110 human-derived normal or cancer cell lines after 5 days of treatment with LLC0150. AR+ prostate cancer models are highlighted in red, and NSD2-mutant hematologic cell lines are shown in purple as well as marked with an asterisk (*). Each cell line's originating tissue lineages and known NSD2 alteration status are shown below.

NSD1/2 dual PROTAC preferentially kills AR+ prostate cancer

Following a medicinal chemistry campaign, we developed a proteolysis targeting chimera (PROTAC) compound, called LLC0150, which co-targets NSD1 and NSD2 (**Figure. 2-4A**). LLC0150 links an NSD2 PWWP-domain binding warhead[49] to a cereblon E3-ligase-recruiting moiety pomalidomide. Treatment with LLC0150 triggers degradation of NSD1 and NSD2, while sparing NSD3 (**Fig. 2-4B**), in a proteasome and cereblon-dependent manner (**Supplementary Figure. A2-8A**). LLC0150 had no effect on other PWWP-domain-containing proteins, but showed partial neo-substrate activity (**Supplementary Figure. A2-8B**). This first-generation PROTAC also had poor solubility and pharmacokinetic properties for *in vivo* use. In line with our genetic data, in PCa cells, NSD1/2 co-degradation with LLC0150 triggered a decrease in the expression of AR and MYC, as well as their downstream gene targets (**Figure. 2-4C,D**). Acute loss of NSD1 and NSD2 in LLC0150-treated LNCaP cells resulted in impaired AR and FOXA1 chromatin binding (**Figure. 2-4E,F** and **Supplementary Figure. A2-8C**), with a parallel loss of H3K27ac activation mark at shared AR/FOXA1 enhancer sites (**Figure. 2-4F**). Treatment with LLC0150 also diminished the chromatin-bound AR fraction and the H3K36me2 histone mark (**Supplementary Figure. A2-8D**). In absence of NSD1/2, DHT-induced expression of AR target genes was significantly weakened (**Supplementary Figure. A2-8E**). LLC0150 treatment also

markedly disrupted the assembly and activity of AR super-enhancers in LNCaP cells (**ESupplementary Figure. A2-8F**)

Global transcriptomic analyses of LNCaP and VCaP cells treated with LLC0150 further showed a significant attenuation of proliferative pathways with a parallel induction of apoptotic signaling (**Supplementary Figure. A2-9A**). This was confirmed via massive accumulation of cleaved-PARP in the LLC0150-treated AR-positive PCa cell lines (**Figure. 2-4C** and **Supplementary Figure. A2-9B**). AR-positive PCa cell lines were considerably more sensitive to treatment with LLC0150 relative to the AR-negative disease models, immortalized normal, as well as primary prostate epithelial cells (**Supplementary Figure. A2-9C**). Inactive epimer control of LLC0150 (labeled as LLC0150-dead) did not affect the NSD1/2 levels or the viability of PCa cells (**Figure. 2-4B** and **Supplementary Figure. A2-9D**). Notably, LLC0150 showed marked synergy with enzalutamide—an AR antagonistic drug—in killing LNCaP and VCaP cells (**Figure. 2-4G and Supplementary Figure. A2-9E,F**). More impressively, LLC0150 also retained cytotoxicity in cell line models that had acquired resistance to enzalutamide (**Figure. 2-4H** and **Supplementary Figure. A2-9G**). Similarly, several models of AR-positive mCRPC organoids that robustly express NSD2 (**Supplementary Figure. A2-10A**), showed significant attenuation of growth upon treatment with LLC0150 in a dose-dependent manner (**Supplementary Figure. A2-10B-D**).

Next, we characterized the cytotoxic effect of LLC0150 in a panel of over 110 human-derived normal and cancer cell lines originating from 22 different lineages (**Supplementary Table 1**). As expected, hematologic cancers harboring activating NSD2 mutations emerged as the most sensitive to treatment with LLC0150 (IC50 ranging from 0.274 - 69.68 nM), which was immediately followed by AR-positive PCa cell lines (shown in red, **Figure. 2-4I**). Notably, AR-positive disease models showed preferential cytotoxicity to NSD1/2 combined loss relative to AR-negative disease models as well as a host of normal cell lines. As proof of concept, we next performed direct intratumoral injection of LLC0150 in mice bearing VCaP xenograft tumors (**Supplementary Figure. A2-10E**). LLC0150 triggered marked degradation of NSD1/2 in tumor xenografts with a parallel loss in proliferative and gain of apoptotic markers (**Supplementary Figure. A2-10F,G**). Altogether, this data suggests that combined loss of NSD1 and NSD2 leads to a dramatic, almost complete, loss of the H3K36me2 histone mark and disruption of the

AR/FOXA1 neo-enhancer circuitry, resulting in apoptotic PCa cell death. This positions NSD1/2 paralogs as a targetable digenic dependency in AR-driven, therapy-resistant PCa.

Discussion

Most targeted therapies following surgical resection or radiation of primary PCa inhibit the androgen/AR signaling axis[4]. However, how the pro-differentiation AR pathway in normal physiology gets reprogrammed to serve as the central oncogene in PCa remains largely unknown. Global AR chromatin binding profiles are markedly different between normal and transformed prostate epithelia[5,6,13–15,50], and FOXA1 and HOXB13 have been implicated in driving AR's reprogramming upon transformation[5,14]. However, both FOXA1 and HOXB13 are also expressed in the normal epithelial cells, raising the possibility for additional cofactors to underlie the recruitment of AR to PCa-specific enhancer elements. Here, in a functional CRISPR screen, we identify NSD2 as a coactivator of the AR/FOXA1 enhanceosome. NSD2 is exclusively expressed in PCa cells, wherein it enables functional binding of AR at chimeric AR-half motifs, which majorly comprise the AR neo-enhancer circuitries. Consequently, NSD2 inactivation abolishes hallmark cancer phenotypes, while its re-expression in deficient cells restores neoplastic features. This positions NSD2 as a neo-coactivator of AR that assists transcription factors, like FOXA1, HOXB13, and ETS, in redistributing AR on the chromatin, thereby unlocking its oncogenic gene programs.

Intriguingly, in motif analyses of the PCa-specific AR cistromes, we also found significant depletion of the canonical ARE elements. Despite magnitude-folds increase in AR abundance in mCRPC, its loading at *cis*-regulatory elements comprising only palindromic AREs was significantly diminished. Also, the full ARE-containing sites were particularly inactivated in mCRPC tumors as evidenced by the loss of H3K27ac. This raises an intriguing possibility for the AR transcriptional activity stemming from a subset of canonical elements to rather impede tumor formation and/or progression, which is consistent with the physiological role of AR as a pro-differentiation factor. In fact, hyper-stimulation of AR activity has anti-proliferative effects in PCa cells[51], and bipolar androgen therapy involving cyclical inhibition and hyperactivation of AR is being currently tested in advanced patients[52,53]. These are exciting areas for further research.

We further found the loss of NSD2 in PCa cells to up-regulate NSD1, and co-inactivation of both NSD1/2 paralogs to be acutely cytotoxic. We uncovered that NSD1 and NSD2, through disparate mechanisms converge on wiring and maintaining the oncogenic AR gene program. While NSD2 directly binds to AR and stabilizes the AR enhanceosome at *de novo* neo-enhancer elements, NSD1 functions as the primary writer enzyme for the H3K36me2 mark that antagonizes the PRC2/EZH2 repressive complex[32,33]. We envision the NSD2 function to evolve from enabling oncogenic AR activity in primary AR-dependent PCa to additionally supporting NSD1 in counter-balancing the canonical repressive PRC2 activity in the metastatic castration-resistant disease. Thus, the loss of NSD2 creates an increased dependency on NSD1 in AR-addicted PCa cells, positioning the NSD1/2 paralogs as targetable co-vulnerabilities in advanced disease. Here, we also characterized a dual PROTAC of NSD1 and NSD2 that confirmed co-degradation of these proteins to result in apoptotic cell death in AR-positive PCa. Notably both NSD1 and NSD2 are recurrently altered in hematological malignancies where they function as driver oncogenes. Accordingly, we found LLC0150 to have the highest potency in NSD2-altered cancers. This highlights the potential application of this compound in studying and treating these tumors.

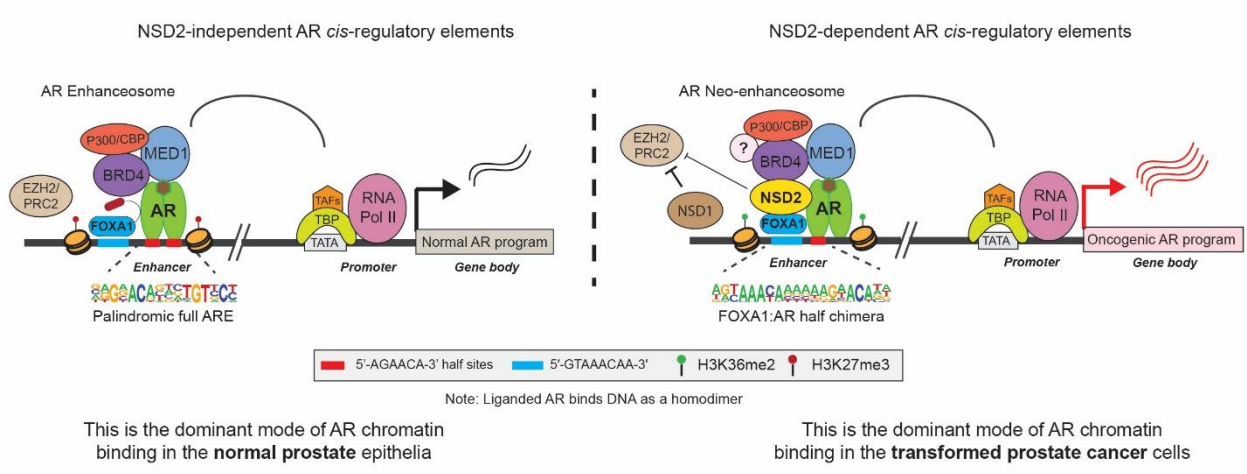


Figure 2-5: Schema depicting NSD2's role in loading the AR enhanceosome at tumor-enriched chimeric AR neo-enhancer elements. Chromatin loading of AR in prostate epithelial cells follows two distinct modes of DNA interactions: **Left**, NSD2-independent binding at *cis*-elements harboring the canonical, 15 bp palindromic androgen response elements (AREs) that are predominantly found in the physiological/normal enhancer circuitry, and **Right**: NSD2-dependent loading at *cis*-regulatory elements harboring chimeric AR-half motifs juxtaposed to the FOXA1 sequence that distinctively constitute the prostate cancer-specific enhancer/super-enhancer (i.e., AR neo-enhancer) circuitries. NSD1, partly supported by NSD2, counteracts repressive

activity of the PRC2/EZH2 complex, thus further amplifying AR/MYC gene expression programs in metastatic castration resistant prostate cancer (mCRPC) cells.

In summary, we identify and characterize NSD2 as an essential coactivator of the AR neo-enhanceosome that is exclusively expressed in PCa cells. NSD2 directly binds to AR and enables its loading at *cis*-regulatory elements harboring chimeric AR-half motifs, comprising over 65% of the malignant AR cistrome. We coalesce these mechanistic insights to propose that AR has two distinct modes of interacting with chromatin: 1) NSD2-independent binding at *cis*-elements harboring canonical full AREs that are predominantly found in the physiological enhancer circuitry, and 2) NSD2-dependent binding at *cis*-regulatory elements harboring chimeric AR-half motifs (like FOXA1:AR-half) that distinctively constitute the cancer-specific enhancer circuitries of AR (**Figure. 2-5**). Furthermore, we uncover NSD1 and NSD2 as a digenic dependency in AR-positive PCa, and develop an NSD1/2 dual PROTAC degrader that shows preferential cytotoxicity in AR-positive PCa. Our findings warrant a focused development of new NSD-targeting therapeutics and evaluation of their efficacy and safety in preclinical and clinical studies.

References:

1. Heinlein, C. A. & Chang, C. Androgen receptor in prostate cancer. *Endocr. Rev.* **25**, 276–308 (2004).

2. Nadal, M. *et al.* Structure of the homodimeric androgen receptor ligand-binding domain. *Nat. Commun.* **8**, 14388 (2017).

3. Shaffer, P. L., Jivan, A., Dollins, D. E., Claessens, F. & Gewirth, D. T. Structural basis of androgen receptor binding to selective androgen response elements. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 4758–4763 (2004).

4. Mills, I. G. Maintaining and reprogramming genomic androgen receptor activity in prostate cancer. *Nat. Rev. Cancer* **14**, 187–198 (2014).

5. Pomerantz, M. M. *et al.* The androgen receptor cistrome is extensively reprogrammed in human prostate tumorigenesis. *Nat. Genet.* **47**, 1346–1351 (2015).

6. Pomerantz, M. M. *et al.* Prostate cancer reactivates developmental epigenomic programs during metastatic progression. *Nat. Genet.* **52**, 790–799 (2020).

7. Wong, Y. N. S., Ferraldeschi, R., Attard, G. & de Bono, J. Evolution of androgen receptor targeted therapy for advanced prostate cancer. *Nat. Rev. Clin. Oncol.* **11**, 365–376 (2014).

8. Robinson, D. *et al.* Integrative clinical genomics of advanced prostate cancer. *Cell* **161**, 1215–1228 (2015).

9. Abida, W. *et al.* Genomic correlates of clinical outcome in advanced prostate cancer. *Proc. Natl. Acad. Sci. U. S. A.* (2019) doi:10.1073/pnas.1902651116.

10. Sahu, B. *et al.* Dual role of FoxA1 in androgen receptor binding to chromatin, androgen signalling and prostate cancer. *EMBO J.* **30**, 3962–3976 (2011).

11. Parolia, A. *et al.* Distinct structural classes of activating FOXA1 alterations in advanced prostate cancer. *Nature* **571**, 413–418 (2019).

12. Armenia, J. *et al.* The long tail of oncogenic drivers in prostate cancer. *Nat. Genet.* **50**, 645–651 (2018).

13. Berchuck, J. E. *et al.* The Prostate Cancer Androgen Receptor Cistrome in African American Men Associates with Upregulation of Lipid Metabolism and Immune Response. *Cancer Res.* **82**, 2848–2859 (2022).

14. Grbesa, I. *et al.* Reshaping of the androgen-driven chromatin landscape in normal prostate cells by early cancer drivers and effect on therapeutic sensitivity. *Cell Rep.* **36**, 109625 (2021).

15. Augello, M. A. *et al.* CHD1 Loss Alters AR Binding at Lineage-Specific Enhancers and Modulates Distinct Transcriptional Programs to Drive Prostate Tumorigenesis. *Cancer Cell* **35**, 603–617.e8 (2019).

16. Asangani, I. A. *et al.* Therapeutic targeting of BET bromodomain proteins in castration-resistant prostate cancer. *Nature* **510**, 278–282 (2014).

17. Rasool, R. U. *et al.* CDK7 Inhibition Suppresses Castration-Resistant Prostate Cancer through MED1 Inactivation. *Cancer Discov.* **9**, 1538–1555 (2019).

18. Xiao, L. *et al.* Targeting SWI/SNF ATPases in enhancer-addicted prostate cancer. *Nature* **601**, 434–439 (2022).

19. Welti, J. *et al.* Targeting the p300/CBP Axis in Lethal Prostate Cancer. *Cancer Discov.* **11**, 1118–1137 (2021).

20. Faivre, E. J. *et al.* Selective inhibition of the BD2 bromodomain of BET proteins in prostate cancer. *Nature* **578**, 306–310 (2020).

21. Lasko, L. M. *et al.* Discovery of a selective catalytic p300/CBP inhibitor that targets lineage-specific tumours. *Nature* **550**, 128–132 (2017).

22. Jin, L. *et al.* Therapeutic Targeting of the CBP/p300 Bromodomain Blocks the Growth of Castration-Resistant Prostate Cancer. *Cancer Res.* **77**, 5564–5575 (2017).

23. Schmitges, F. W. *et al.* Histone methylation by PRC2 is inhibited by active chromatin marks. *Mol. Cell* **42**, 330–341 (2011).

24. Hansen, K. H. *et al.* A model for transmission of the H3K27me3 epigenetic mark. *Nat. Cell Biol.* **10**, 1291–1300 (2008).

25. Poulin, M. B. *et al.* Transition state for the NSD2-catalyzed methylation of histone H3 lysine 36. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 1197–1201 (2016).

26. Malgeri, U. *et al.* Detection of t(4;14)(p16.3;q32) chromosomal translocation in multiple myeloma by reverse transcription-polymerase chain reaction analysis of IGH-MMSET fusion transcripts. *Cancer Res.* **60**, 4058–4061 (2000).

27. Chesi, M. *et al.* The t(4;14) translocation in myeloma dysregulates both FGFR3 and a novel gene, MMSET, resulting in IgH/MMSET hybrid transcripts. *Blood* **92**, 3025–3034 (1998).

28. Vo, J. N. *et al.* The genetic heterogeneity and drug resistance mechanisms of relapsed refractory multiple myeloma. *Nat. Commun.* **13**, 3750 (2022).

29. Jaffe, J. D. *et al.* Global chromatin profiling reveals NSD2 mutations in pediatric acute lymphoblastic leukemia. *Nat. Genet.* **45**, 1386–1391 (2013).

30. Li, J. *et al.* NSD2-E1099K Mutation Leads to Glucocorticoid-Resistant B Cell Lymphocytic Leukemia in Mice. *Blood* **136**, 3–4 (2020).

31. Kuo, A. J. *et al.* NSD2 links dimethylation of histone H3 at lysine 36 to oncogenic programming. *Mol. Cell* **44**, 609–620 (2011).

32. Xiao, L. *et al.* Epigenetic Reprogramming with Antisense Oligonucleotides Enhances the Effectiveness of Androgen Receptor Inhibition in Castration-Resistant Prostate Cancer. *Cancer Res.* **78**, 5731–5740 (2018).

33. Ku, S. Y. *et al.* Rb1 and Trp53 cooperate to suppress prostate cancer lineage plasticity, metastasis, and antiandrogen resistance. *Science* **355**, 78–83 (2017).

34. Shi, J. *et al.* Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. *Nat. Biotechnol.* **33**, 661–667 (2015).

35. Groner, A. C. *et al.* TRIM24 Is an Oncogenic Transcriptional Activator in Prostate Cancer. *Cancer Cell* **29**, 846–858 (2016).

36. Fong, K.-W., Zhao, J. C., Song, B., Zheng, B. & Yu, J. TRIM28 protects TRIM24 from SPOP-mediated degradation and promotes prostate cancer progression. *Nat. Commun.* **9**, 5007 (2018).

37. Yang, P. *et al.* Histone methyltransferase NSD2/MMSET mediates constitutive NF-κB signaling for cancer cell proliferation, survival, and tumor growth via a feed-forward loop. *Mol. Cell. Biol.* **32**, 3121–3131 (2012).

38. Ezponda, T. *et al.* The histone methyltransferase MMSET/WHSC1 activates TWIST1 to promote an epithelial-mesenchymal transition and invasive properties of prostate cancer. *Oncogene* **32**, 2882–2890 (2013).

39. Asangani, I. A. *et al.* Characterization of the EZH2-MMSET histone methyltransferase regulatory axis in cancer. *Mol. Cell* **49**, 80–93 (2013).

40. Aytes, A. *et al.* NSD2 is a conserved driver of metastatic prostate cancer progression. *Nat. Commun.* **9**, 5201 (2018).

41. Li, N. *et al.* AKT-mediated stabilization of histone methyltransferase WHSC1 promotes prostate cancer metastasis. *J. Clin. Invest.* **127**, 1284–1302 (2017).

42. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).

43. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).

44. Tomlins, S. A. *et al.* Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. *Nature* **448**, 595–599 (2007).

45. Nabet, B. *et al.* The dTAG system for immediate and target-specific protein degradation. *Nat. Chem. Biol.* **14**, 431–441 (2018).

46. Kang, H.-B. *et al.* The histone methyltransferase, NSD2, enhances androgen receptor-mediated transcription. *FEBS Lett.* **583**, 1880–1886 (2009).

47. Finogenova, K. *et al.* Structural basis for PRC2 decoding of active histone methylation marks H3K36me2/3. *Elife* **9**, (2020).

48. Drosos, Y. *et al.* NSD1 mediates antagonism between SWI/SNF and polycomb complexes and is required for transcriptional activation upon EZH2 inhibition. *Mol. Cell* **82**, 2472–2489.e8 (2022).

49. Dilworth, D. *et al.* A chemical probe targeting the PWWP domain alters NSD2 nucleolar localization. *Nat. Chem. Biol.* **18**, 56–63 (2022).

50. Shen, M., Demers, L.-K., Bailey, S. D. & Labbé, D. P. To bind or not to bind: Cistromic reprogramming in prostate cancer. *Front. Oncol.* **12**, 963007 (2022).

51. Schweizer, M. T. *et al.* Effect of bipolar androgen therapy for asymptomatic men with castration-resistant prostate cancer: results from a pilot clinical study. *Sci. Transl. Med.* **7**, 269ra2 (2015).

52. Markowski, M. C. *et al.* Clinical Efficacy of Bipolar Androgen Therapy in Men with Metastatic Castration-Resistant Prostate Cancer and Combined Tumor-Suppressor Loss. *Eur Urol Open Sci* **41**, 112–115 (2022).

53. Markowski, M. C. *et al.* A Multicohort Open-label Phase II Trial of Bipolar Androgen Therapy in Men with Metastatic Castration-resistant Prostate Cancer (RESTORE): A Comparison of Post-abiraterone Versus Post-enzalutamide Cohorts. *Eur. Urol.* **79**, 692–699 (2021).

54. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).

55. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

56. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

57. Korotkevich, G. *et al.* Fast gene set enrichment analysis. *bioRxiv* (2016) doi:10.1101/060012.

58. Wickham, H. Ggplot2. *Wiley Interdiscip. Rev. Comput. Stat.* **3**, 180–185 (2011).

59. Wickham, H. & Pedersen, T. L. gtable: Arrange grobs in tables. *R package version 0.3. 0*.

60. Blighe, K., Rana, S. & Lewis, M. EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling. doi: 10.18129/B9. bioc. Preprint at (2019).

61. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

62. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

63. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, (2021).

64. Institute, B. 'Picard Toolkit', Broad institute, GitHub repository. *Picard Toolkit*.

65. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).

66. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

67. Kent, W. J., Zweig, A. S., Barber, G., Hinrichs, A. S. & Karolchik, D. BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* **26**, 2204–2207 (2010).

68. Zhu, L. J. Integrative analysis of ChIP-chip and ChIP-seq dataset. *Methods Mol. Biol.* **1067**, 105–124 (2013).

69. Yu, G., Wang, L.-G. & He, Q.-Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382–2383 (2015).

70. Wang, Q. *et al.* Exploring Epigenomic Datasets by ChIPseeker. *Curr Protoc* **2**, e585 (2022).

71. Bailey, T. L., Johnson, J., Grant, C. E. & Noble, W. S. The MEME Suite. *Nucleic Acids Res.* **43**, W39–49 (2015).

72. Grant, C. E. & Bailey, T. L. XSTREME: Comprehensive motif analysis of biological sequence datasets. *bioRxiv* 2021.09.02.458722 (2021) doi:10.1101/2021.09.02.458722.

73. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–5 (2016).

74. Phanstiel, D. H., Boyle, A. P., Araya, C. L. & Snyder, M. Sushi: Tools for visualizing genomics data. *R package version*.

75. Killian, T. & Gatto, L. Exploiting the DepMap cancer dependency data using the depmap R package. *F1000Res.* **10**, 416 (2021).

Methods

Ethical Statement

All experiments detailed in this manuscript were performed in compliance with the Institutional Review Board (IRB) and the Institutional Animal Care and Use Committee (IACUC) at the University of Pennsylvania and the University of Michigan.

Animal procurement: Animal studies were approved by the Institutional Animal Care and Use Committee (IACUC) at the University of Pennsylvania and/or the University of Michigan. Animal use and care were in strict compliance with institutional guidelines, and all experiments conformed to the relevant regulatory standards by the universities. NOD SCID or NCI SCID/NCr athymic nude mice were obtained from the Jackson Laboratory (strain code: 005557) and Charles River (strain code: 561). All *in vivo* experiments were initiated with male mice aged 5-8 weeks. All mice were housed in a pathogen-free animal barrier facility and all in vivo experiments were initiated with male mice aged 5-8 weeks. All mice were maintained under the conditions of pathogen-free, 12 hours light/12 hours dark cycle, temperatures of 18-23°C, and 40-60% humidity.

Statement on use of human specimens: Prostate tumor patient tissues were acquired from the University of Michigan (U-M) pathology archives. These tissues were utilized for Immunohistochemistry and multiplex Immunofluorescence experiments to assess for Cytokeratin-8 and NSD2 expression in tumor or adjacent normal prostate cells. Formalin-fixed paraffin embedded specimens from the archives were used upon approval by the U-M Institutional Review Board and does not require patient consent.

Cell lines

Most cell lines were purchased from the American Type Culture Collection (ATCC) and were cultured following ATCC protocols. For all experiments, LNCaP and 22RV1 cells were grown in RPMI 1640 medium (Gibco) and VCaP cells in DMEM with Glutamax (Gibco) medium supplemented with 10% fetal bovine serum (FBS; Invitrogen). HEK293FT cells were grown in DMEM (Gibco) medium with 10% FBS. All cells were grown in a humidified 5% CO2 incubator at 37 °C. Mycoplasma and cell line genotyping were performed once a fortnight and every month respectively at the University of Michigan Sequencing Core using Profiler Plus (Applied Biosystems). Results from these were compared with corresponding short tandem repeat profiles in the ATCC database to authenticate their identity.

Antibodies

For immunoblotting, the following antibodies were used: NSD1 (NeuroMab: 75-280, 1:1000); NSD2 (Abcam:ab75359, 1:1000); NSD3 (Cell Signaling Technologies: 92056S, 1:1000); KLK3/PSA (Dako:A0562, 1:1000); FKBP5(Cell Signaling Technologies: 12210, 1:1000); NKX3-1 (Cell Signaling Technologies:83700S, 1:1000); FOXA1 N-terminal (Cell Signaling Technologies: 58613S; Sigma-Aldrich: SAB2100835, 1:1000); FOXA1 C-terminal (Thermo Fisher Scientific: PA5-27157, 1:1000); AR (Millipore: 06-680, 1:1000); AR (Abcam: ab133273, ab108341, 1:1000); H3 (Cell Signaling Technologies: 3638S, 1:2500); GAPDH (Cell Signaling Technologies: 3683, 1:2500); H3K27me3(Millipore: 07-449, 1:2500); H3K36me2 (Cell Signaling Technologies: 2901S, Abcam: ab9049, 1:2500); H3K27ac (Active Motif, Cat#39336, Cat#39133, 1:2500); Phospho-AR (Ser-81) (Millipore, Cat# 07-1375-EMD, 1:1000); HALO (Thermo Scientific ,Cat# G9281, 1:1000); HA (Cell Signaling Technologies, Cat# 3724S, 1:1000); His (Cell Signaling Technologies, Cat#2365S, 1:1000). ChIP-seq assays were performed using the following antibodies: FOXA1 (Thermo Fisher Scientific: PA5-27157); AR (Millipore: 06-680);

H3K4me1 (Abcam: ab8895); H3K4me2 (CST: C64G9); H3K36me2 (Abcam: ab9049); H3K27me3 (EMD: 07-449), and H3K27ac (Active Motif, Cat#39336).

Cell free protein-protein interaction studies

In vitro protein expression was carried out by cloning the desired expression cassettes downstream of a Halo- or His-tag to produce fusion proteins. Briefly, AR-DBD was subcloned in pFN21K containing Halo tag, and NSD2-HMGa was cloned in pcDNA4c containing His tag. After cloning, the fusion proteins were expressed using the cell-free transcription and translation system (Cat. # L4140, Promega) following the manufacturer's protocol. For each reaction, protein expression was confirmed by Western blot.

A total of 10µl cell-free reaction containing halo- and His-tag fusion proteins was incubated in PBST (0.1% tween) at 4°C overnight. Ten microliter HaloLink beads (Cat. #G931, Promega) were blocked in BSA at 4°C for overnight. After washes with PBS, the beads were mixed with AR-NSD2-HMGa and TM mixture and incubated at RT for 1hr. Halolink beads were then washed with PBST for 4 times and eluted in SDS loading buffer. Proteins were separated on SDS gel and blotted with anti-His Ab (CST: Cat#2365S).

Colony formation assays

For the colony formation assay, approximately 10,000 cells/well in six-well plates (n=3) were seeded and treated with the required drugs/compounds or vehicle for 12-14 days. Media was replenished every 3-4 days. Colonies were fixed and stained using 0.5% (w/v) crystal violet (Sigma, C0775) in 20% (v/v) methanol for 30 min, washed with distilled deionized water, and air-dried. After scanning the plate, the stained wells were destained with 500 µL 10% acetic acid, and the absorbance was determined at 590 nm using a spectrophotometer (Synergy HT, BioTek Instruments, Vermont-USA).

Cellular protein fractionation assays

Chromatin-bound proteins were extracted following a protocol previously described[17]. In brief, 10 million cells were collected, washed with DPBS, and resuspended in 250 µl Buffer A (10 mM HEPES pH 7.9, 10 mM KCl, 1.5 mM $MgCl_2$, 0.34 M sucrose, 10% glycerol, 1 mM DTT)

supplemented with 0.1% TritonX-100. After incubation on ice for 10 min, the nuclear pellet was collected by centrifugation at 1300 x g for 5 min at 4°C, washed in Buffer A, and resuspended in Buffer B (3 mM EDTA, 0.2 mM EGTA, 1 mM DTT) with the same centrifugation settings, and incubated on ice for 30 min. The chromatin pellet was collected by centrifugation at 1700 x g for 5 min at 4°C, washed and resuspended in Buffer B with 150 mM NaCl, and incubated on ice for 20 min. After centrifugation at 1700 x g for 5 min to remove proteins soluble in 150 mM salt concentrations, the pellet was then incubated in Buffer B with 300 mM NaCl on ice for 20 min and centrifuged again at 1700 x g to obtain the final chromatin pellet. The chromatin pellet was dissolved in a sample buffer, sonicated for 15 seconds, and boiled at 95°C for 10 min. Immunoblot analysis was conducted on samples as described above. All buffers were supplemented with Pierce protease inhibitor and Halt protease & phosphatase inhibitors.

RNA isolation and quantitative real-time PCR

Standard protocol from the miRNeasy Mini kit (Qiagen) was used to extract total RNA with the inclusion of on-column genomic DNA digestion step using the RNase-free DNase Kit (Qaigen). RNA concentration was estimated 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) following manufacturer's instructions. 20ng of cDNA was used for each polymerase chain reaction (PCR) using the FAST SYBR Green Universal Master Mix (ThermoFisher Scientific), and every sample was quantified in triplicates. Gene expression was normalized and 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 provided in Supplementary Table 2.

siRNA/ASO-mediated gene knockdown

Mammalian cells were seeded in a 6-well plate format at the density ranging from 100,000–250,000 cells per well. 12 hours post seeding, cells were transfected with 25nM of gene-targeting ON-TARGETplus SMARTpool siRNAs (or ASOs) or non-targeting pool siRNAs (or ASOs) as negative control (Dharmacon) using the RNAiMAX reagent (Life Technologies; Cat#: 13778075)

on two consecutive days, following manufacturer's instructions. 72 hours after transfection, total RNA and protein were extracted to confirm efficient (>80%) knockdown of the target genes. For the siRNA-treated VCaP DMSO/EPZ-6438 RNA-seq experiment (Fig. 3k), cells were pre-treated with control siRNA (siNC) or siRNA targeting NSD1, NSD2, or NSD1/2 (siNSD1, siNSD2) for 30 days, followed by 72hours of EPZ-6438 treatment. Catalog numbers and guide RNA sequences of siRNA SMARTpools (Dharmacon) are provided in Supplementary Table 3.

CRISPR-Cas9-mediated gene knockout

For gene knockouts, cells were seeded in a 6-well plate at a density of 200,000–300,000 cells per well and transduced with viral particles with lentiCRISPR-V2 plasmids coding either non-targeting (sgNC) or sgRNAs targeting NSD1 and NSD2. 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). sgRNA sequences are provided in Supplementary Table 4.

Proliferation assays

For siRNA growth assays, cells were directly plated in a 96-well plate at the density of 2,500–8,000 cells per well and transfected with gene-specific or non-targeting siRNAs, as described above, on day 0 and day 1. Every treatment was carried out in six independent replicate wells. CellTiter-Glo reagent (Promega) was used to assess cell viability at multiple time points after transfection, following the manufacturer's protocol. Data were normalized to readings from siNC treatment on day 1 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 96-well plate at 2500 cells per well, with five-six replicates per group.

Matrigel invasion assay

LNCaP CRISPR clones were grown in 10% CSS-supplemented medium for 48 h for androgen starvation. Matrigel-coated invasion chambers were used and additionally coated with polyethylene terephthalate membrane to allow for fluorescent quantification of the invaded cells (Biocoat: 24-well format, no. 354166). On the upper layer of the chamber, fifty thousand starved

cells were resuspended in serum-free medium and were added to each invasion chamber while twenty percent FBS-supplemented medium was added to the bottom wells to serve as a chemoattractant. 12 hours later, medium from the bottom well was aspirated and replaced with 1x HBSS (Gibco) containing 2 μg/ml Calcein-green AM dye (Thermo Fisher Scientific; C3100MP) and incubated for 30 min at 37 °C. Invasion chambers were then placed in a fluorescent plate reader (Tecan-Infinite M1000 PRO), and fluorescent signals from the invaded cells at the bottom were averaged across 16 distinct regions per chamber to determine the extent of invasion. For rescue experiments, stable lines overexpressing the NSD2 isoforms were generated. Briefly, to LNCaP NSD2 KO lines, GFP or NSD2-Long isoform containing viruses were added. These lines were then used to perform the invasion assay as described above.

RNA-seq and analysis

RiboErase RNA-seq libraries were prepared using 200–1,000 ng of total RNA. Ribosomal RNA was removed by enzymatic digestion of the specific probe-bound duplex rRNA (KAPA RNA Hyper+RiboErase HMR, Roche) and then fragmented to around 200-300bp with heat in the fragmentation buffer. Following this, double-stranded cDNA was generated, and end-repair and ligation was performed using New England Biolabs (NEB) adapters. Final library preparation was performed by amplification with the 2x KAPA HiFi HotStart mix and NEB dual barcode following the manufacturer's protocol. 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.

RNA data was first processed using kallisto (version 0.46.1)[54]. Then analysis was performed in R, first read counts were normalized and filtered (counts >10) using EdgeR[55] (edgeR_3.39.6), and differential expression was performed using Limma-Voom (limma_3.53.10)[56]. Gene Set Enrichment Analysis (GSEA) was performed using fgsea (fgsea_1.24.0)[57] and comparisons were made to several signatures, including an experimentally derived AR signature, the human hallmark MsigDB signatures (/www.gsea-msigdb.org), and the hallmark androgen response signature (HALLMARK_ANDROGEN_RESPONSE.v7.5.1.gmt). In addition, R packages tidyverse, gtable, gplots, ggplot2, and EnhancedVolcano (EnhancedVolcano_1.15.0) were also used for generating summary figures (R version 4.2.1[58–60]).

ChIP-seq and data analysis

Chromatin immunoprecipitation (ChIP) experiments were carried out using the Ideal ChIP-seq Kit for Transcription Factors or Histones (Diagenode) as per the manufacturer's protocol. Chromatin from 2×10^6 cells (for transcription factors) and $1x10^6$ cells (for histones) was used for each ChIP reaction with 4 or 2 μg of the target protein antibody, respectively. 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, *Escherichia coli* DNA polymerase I large fragment (Klenow polymerase), and T4 polynucleotide kinase (New England BioLabs (NEB)). A single adenine 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).

ChIP-seq analysis was carried out by first assessing reads and performing trimming using Trimmomatic version 0.39 (settings TruSeq3-PE-2.fa:2:30:10, minlen 50)[61]. Paired end reads were aligned to hg38 (GRCh38) human genome reference using bwa ("bwa mem" command with options -5SP -T0, version 0.7.17-r1198-dirty) [62]. Alignments were then filtered using both samtools[63] (v1.1,quality score cutoff of 20) and picard [64] MarkDuplicates (v(2.26.0-1-gbaf4d27-SNAPSHOT), removed duplicates). Peak calling was performed using MACS2 (v2.2.7.1) [65] using narrowpeak setting for narrow peaks and a second set for broad peaks (e.g., H3K27ac, --broad -B --cutoff-analysis --broad-cutoff 0.05 --max-gap 500). Finally, bedtools (v2.27.1)[66] was used to remove blacklisted regions of the genome from the peak list (Encode's exclusion list

ENCFF356LFX.bed). UCSC's tool wigtoBigwig (v2.8) was used for conversion to bigwig formats[67].

Overlap analysis of ChIP-seq data

Peak lists from MACS were compared between samples using R package ChIPpeakAnno [68–70]. Peaks within 500bp of each other were reduced to single peaks. Overlaps were calculated using settings maxgap=-1L, minoverlap=0L, ignore.strand=TRUE, connectedPeaks=c('keepAll', 'min', 'merge'). Comparisons of enrichment sites to the known gene database (TxDb.Hsapiens.UCSC.hg38.knownGene) were performed using R package ChIPseeker. A distance of +/- 1 kb was used to assess relative distance from gene regions.

HOMER motif calling

De novo and known motif enrichment analysis was performed using HOMER (version v.4.10)[42,71]. Custom motif matrices were generated manually, then assigned score thresholds using HOMER's utility seq2profile, allowing for two mismatches. This setting was chosen after iteratively comparing performance with the pre-existing FOXA1:AR motif. Further customization was achieved by checking for presence of motif elements with different spacings, ranging from 0-8 'N's added between elements, and flipping the order of elements in each of these: FOXA1-ARE, ARE-FOXA1, FOXA-N-ARE, ARE-N-FOXA1, FOXA1-NN-ARE, ARE-NN-FOXA1, etc.

Custom motifs were then further validated using XSTREME (v5.5.5)[72] from the MEME Suite[71] to check for additional configurations and variations in padding between motif elements.

Enrichment heatmaps

The software Deeptools (v3.5.1) was used to generate enrichment plots and read density heatmaps. A reference point parameter of +/- 2.5kb for histone signals and +/- 1.5kb for AR/FOX signals was used. Other settings included using 'skipzeros', 'averagetype mean,' and 'plotype se'. The Encode blacklist ENCFF356LFX was used [73].

Motif and signal plots

Sushi (Sushi_1.32.0) package in R was used to layer signal tracks. The plotBedgraph(), plotGenes(), plotBed() functions were used with output from ChIP-seq alignments and output from HOMER motif enrichment analysis[74].

Super-enhancer analysis

Super-enhancer regions were identified with findPeaks function from HOMER (version v.4.10) [42] 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 identify super-enhancer clusters. The input files to findPeaks were tag directories generated from alignment files in SAM format with makeTagDirectory function from HOMER. Super-enhancer scores were plotted using the normalized tag count values between the datasets.

Single-cell data analysis

Three public scRNA-seq datasets from primary prostate cancer were downloaded from GEO or a website provided by the author (GSE193337, GSE185344, www.prostatecellatlas.org)[75]. Using cell annotation from the Tuong et al. dataset as reference, luminal cells were annotated for the other two datasets with the label transfer method of Seurat[29]. Pseudo-bulk expression profiles[30] were generated by summing counts from all cells annotated as luminal cells for each patient (tumor and normal samples separately). Normalization was achieved by computing normalization factors with the trimmed mean of M-values method[31] and applying the cpm function from edgeR (v 3.36.0)[32]. Boxplots of *NSD2* and *PCA3* expression were generated with ggpubr[33] and paired Wilcoxon test was used to test the significance of the difference between benign and tumor (only patients with paired benign and tumor samples were included).

Immunohistochemistry (IHC) and immunofluorescence (IF)

Immunohistochemistry (IHC) was performed on 4-μm-thick formalin-fixed, paraffin-embedded (FFPE) tissue sections using anti-NSD2 mouse monoclonal primary antibody (catalog no. ab75359, Abcam), anti-AR rabbit monoclonal primary antibody (catalog no. 760-4605, Roche-Ventana), and anti-CK8 rabbit monoclonal primary antibody (catalog no. ab53280, Abcam).

Singleplex IHC was carried out on the Ventana ULTRA automated slide staining system (Roche-Ventana Medical Systems) using the OmniView Universal diaminobenzidine (DAB) detection kit (catalog no. 760-500, Roche-Ventana) and Hematoxylin II (catalog no. 790-2208, Roche-Ventana) for counterstain. Staining was evaluated under 100× and 200× magnification using a bright-field microscope.

Assessment of drug synergism

To determine the 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.

Statistics and Reproducibility

All immunoblot experiments were repeated at least two to three times. For immunofluorescent staining experiments, number of biological replicates used in each case are noted in the figure legend. While representative images are shown in some panels, e.g. Figure 3b and Extended Data Figure 5e, quantitation from all independent replicates is included. No statistical methods were used to predetermine sample sizes for any experiments. For all analyses, data distribution was assumed to be normal, but this was not formally tested. All immunofluorescence data quantification was performed in a double-blinded manner by in-house pathologists. For in vivo animal experiments,

Data Availability

All data are available in the manuscript or the supplementary information. Raw next-generation sequencing data, including ChIP-seq and RNA-seq, generated in this study are deposited in the Gene Expression Omnibus (GEO) repository (accession number: GSE242737) at National Center for Biotechnology Information. ChIP-seq data from normal, primary prostate cancer (PCa), and metastatic castration-resistant prostate cancer were pulled from GEO repositories GSE130408 and

GSE70079. Three public scRNA-seq datasets from primary PCa were downloaded from GEO (GSE193337 and GSE185344) or a web-portal provided by the authors.

Code Availability

All custom codes used for data analyses are freely available from the following public repositories: Github: https://github.com/mctp/NSD2_req_subunit[76] and Zenodo: https://doi.org/10.5281/zenodo.12979564)[77].

Chapter acknowledgments

The contents of Chapter 2 were published online as part of this article: NSD2 is a requisite subunit of the AR/FOXA1 neo-enhanceosome in promoting prostate tumorigenesis. *Nature Genetics*, 2024. doi: 10.1038/s41588-024-01893-6. 1. Parolia A, Eyunni S, Verma BK, Young E, Liu L, George J, Aras S, Das CK, Mannan R, Rasool RU, Luo J, Carson S, Velasquez EM, Liu Y, Xiao L, Gajjala PR, Jaber M, He T, Qiao Y, Pang M, Zhang Y, Alhusayan M, Apel I, Cao X, Tavana O, Hou C, Wang Z, Ding K, Chinnaiyan AM, Asangani I.

Author Contributions

A.P., A.M.C., and I.A.A. conceived the study and the experiments; A.P. and S.E. designed and carried out the multi-omics and functional experiments with assistance from J.L., S.E.C., Y.L., L.X., X.W., T.H., Y.Q., P.G., M.J., R.R., S.M., M.P., E.M., J.C.T., M.L., F.S., and X.C.; B.K.V. designed and carried out the fragment-based coimmunoprecipitation and functional experiments with assistance from S.A. and R.R; B.K.V. and C.K.D. conducted *in vivo* experiments. A.P. and E.Y. carried out all the bioinformatics analyses with assistance from J.G., S.V., and M.A.; Y.Z. analyzed the single-cell RNA-seq datasets; O.T. provided guidance; L.L., C.H., Z.W., and K.D. were involved in the synthesis and structural validation of the LLC0150 compound; A.P. wrote the manuscript and developed the figures with feedback from A.M.C. and I.A.A.; S.E. wrote the Methods section with help from B.K.V., E.Y., Z.W., R.M., and Y.Z. Next-generation sequencing (NGS) related to data shown in Fig. 1c, 3k, 3m, and 4d as well as related supplementary figures was performed at the University of Michigan, while NGS related to data shown in Fig. 1g, 1i, 2a-e, 2i,j, and 4e,f as well as related supplementary figures was carried out at the University of

Pennsylvania. All authors discussed the results, provided feedback, and reviewed the final manuscript.

Chapter 3

Divergent FOXA1 Mutations Drive Prostate Tumorigenesis and Therapy-Resistant Cellular Plasticity[2]

Abstract

FOXA1 is frequently altered in hormone-driven malignancies, with activating mutations found in 10-40% of primary prostate cancers based on ethnicity. Yet, the tumorigenic potential and pathobiology of FOXA1 alterations remain unexplored *in vivo*. Here, we characterize prostate tissues from knock-in mouse models engineered to harbor *FOXA1* mutant transgenes, representing distinct classes of alterations found in human cancers. Our findings reveal that FOXA1 Class 1 mutations (i.e., Wing 2 alterations) in a p53-null background, drive high-grade, invasive luminal prostate adenocarcinoma with full penetrance by 40 weeks. These lesions retain luminal characteristics, gain NSD2 expression, and regress upon androgen withdrawal—mirroring key features of human prostate adenocarcinomas. Mechanistically, Class 1 mutants aberrantly activate mTORC1/2 signaling and reprogram androgen receptor (AR) activity by expanding the enhancer circuitry at non-canonical, chimeric FOXA1:AR-half elements enriched in human tumors. In contrast, FOXA1 Class 2 mutations (i.e., C-terminal truncations) do not drive cellular transformation. Instead, these mutations induce intra-luminal plasticity by partially reprogramming differentiated cells into a progenitor, stem-like state within androgen-replete normal tissues—a phenomenon otherwise induced upon castration. Class 2 mutants lead to a 10 to 20-fold expansion of AR+/CK8+ luminal epithelia that gain expression of stemness markers such as *Trop2*, *Ck4*, and *Psca*. Mechanistically, while maintaining AR activity, Class 2 mutants de-compact over 40,000 new *cis*-regulatory elements bound by KLF5 and AP-1 transcription factors to activate luminal progenitor and castration-induced gene programs. Accordingly, Class 2-mutant-expressing prostate epithelial cells better survive after castration and remain proliferative despite androgen withdrawal. Furthermore, castration-resistant prostate cancers from human

[2]Contents of this chapter were previously published as part of the following manuscript: **Eyunni S**, et al. Divergent FOXA1 mutations drive prostate tumorigenesis and therapy-resistant cellular plasticity. *Science*. 2025. doi: 10.1126/science.adv2367.

patients that harbor Class 2 mutations show highest KLF5 expression as well as cancer stemness signatures. Overall, our findings position FOXA1 as a multifaceted oncogene in AR-dependent prostate cancers, wherein divergent selection of distinct mutational patterns either drive cancer formation or therapy-resistant disease progression.

Introduction

Prostate cancer (PCa) is characterized by transcriptional dysregulation, with over 70% of the primary disease driven by somatic alterations affecting transcription factors (TFs) and their regulatory chromatin and epigenetic cofactors (*1*–*3*). Activating ETS family gene fusions (involving ERG, ETV1, ETV4, and FLI1) with highly active promoters are pathognomonic driver alterations in primary PCa—recurrent in over 60% of cases in European White men (*4*). Genomic studies have also uncovered frequent alterations in the pioneer transcription factor FOXA1 in 10–35% of PCa in this population (*2*, *3*, *5*–*7*). Strikingly, in the Asian population, FOXA1 mutations are detected in over 40% of primary PCa cases, surpassing ETS fusions to become the predominant driver alteration in this patient ethnicity (*8*). These genetic alterations reshape the transcriptional landscape of prostate epithelial cells, activating oncogenic gene programs that drive tumorigenesis and metastatic progression.

Almost all primary prostate tumors express the lineage-specific androgen receptor (AR) protein. Upon binding to its ligand (i.e., androgens), AR homodimerizes and enters the nucleus to activate transcription of genes that are crucial for terminal differentiation and secretory function of the prostate luminal epithelial cells (*9*, *10*). AR activity, however, is extensively reprogrammed during prostate tumorigenesis to acquire an oncogenic gene program (*11*, *12*), becoming essential for hyperproliferation and survival of tumor cells (*13*). Thus, targeted therapies blocking either the production of androgen (i.e., surgical or medical castration) or directly inhibiting AR (e.g. enzalutamide, etc.) comprise mainstay regimens after surgical resection or radiation of the localized disease (*14*, *15*). Androgen deprivation therapies (ADTs) induce only temporary remission of the tumor and, in all treated patients, the disease relapses by further acquiring mutations in AR pathway genes or co-factors that restore AR activity, including copy amplification or hotspot mutations of AR itself, and is clinically referred to as castration-resistant prostate cancer (CRPC) (*16*, *17*).

As a pioneer factor, FOXA1 binds to nucleosomal DNA at distal enhancers and triggers chromatin decompaction to enable transcriptional activity of lineage-specific transcription factors (*18*, *19*), including AR in the prostate luminal epithelia (*20*, *21*). Our group recently characterized FOXA1 alterations into three distinct structural classes, each showing disparate clinical incidence and molecular gain-of-functions (*5*) (**Figure. 3-1A**). Class 1 alterations are detected in localized, primary tumors (i.e., are initiating events) that primarily comprise missense and in-frame indel mutations in the Wing 2 region of the DNA-binding forkhead domain, imparting FOXA1 higher nuclear mobility and transactivation potential for oncogenic AR signaling (*5*). In contrast, FOXA1 Class 2 mutations are clonally detected only in the metastatic disease (i.e., are acquired events) and comprise frameshift mutations in the C-terminal half of the protein, imparting dominance in engaging chromatin and de-repressing metastatic WNT signaling (*5*). Class 3 alterations comprise structural variations within the *FOXA1* gene locus that amplify a conserved enhancer to increase the wild-type FOXA1 activity. Similar FOXA1 alterations are also detected in other hormone-driven malignancies, including breast, salivary gland, and bladder tumors (*22–26*).

Despite such high recurrence of FOXA1 alterations, their pathobiology remains unexplored in mouse models. Mouse models harboring other recurrent truncal genetic alterations such as *ETS* fusions, *SPOP* mutations, *NKX3.1* deletion, *MYC* amplification, and *CHD1* deletion have uncovered important mechanisms underlying PCa initiation (*27–35*). For instance, modeling the characteristic chromosome 8p loss and chromosome 8q gain observed in human PCa (*2*, *3*), several groups have demonstrated that NKX3.1 inactivation or MYC overexpression in the mouse prostate epithelia triggers malignant transformation (*31*, *33*). Notably, more high-grade invasive tumors are detected in compound mouse models where, based on co-alterations found in the human disease, PTEN or p53 is simultaneously inactivated (*27*, *28*, *31*, *36*, *37*). These *in vivo* findings establish these genetic alterations as collaborative drivers in the development and/or progression of PCa.

In this study, we present genetically engineered mice with targeted knock-in of human cancer-associated FOXA1 mutant transgenes, that provide crucial mechanistic insights into divergent mechanisms of FOXA1 Class 1 and Class 2 alterations in PCa. We found that FOXA1 Class 1 mutants, in conjunction with *Trp53* loss, co-activate oncogenic AR and mTORC1 gene programs to drive high-grade, prostate adenocarcinomas, which invade into the surrounding stroma upon progression. These tumors retain luminal features and resemble the primary human disease,

exhibiting the loss of NKX3.1, gain of MYC, and abnormal induction of NSD2 (a known co-activator of cancer-specific AR signaling (*12*). Furthermore, Class 1-driven tumors are androgen-sensitive and regress upon castration. In contrast, FOXA1 Class 2 mutations are unable to drive prostatic transformation. Instead, these DNA-binding-dominant FOXA1 mutants reprogram differentiated luminal epithelial cells into a progenitor, stem-like state through *de novo* genesis of AP-1/KLF5 neo-enhancer circuitry in androgen-replete normal prostate tissues—a molecular phenomenon herein described as induced intra-luminal plasticity. Class 2 mutant-expressing prostate epithelium exhibits enhanced survival upon castration, suggesting FOXA1 Class 2 mutations serve as promoter events in the course of disease progression. Overall, our findings causally implicate FOXA1 mutations in AR-dependent prostate tumorigenesis and progression, while also explaining the divergent evolution of distinct classes of FOXA1 alterations in human PCa.

Results

FOXA1 Class 1 mutation with *Trp53* deletion drives high-grade invasive adenocarcinoma

We first engineered a novel transgenic mouse model via targeted knock-in of the human *FOXA1 R265-71del* (Class 1) mutant transgene into the first intron of the mouse *Rosa26* gene locus. The knock-in sequence consisted of a synthetic CAG promoter driving constitutive expression of the LoxP-flanked puromycin-resistance (puroR) selection marker, followed by the FOXA1-mutant coding sequence fused to a V5 epitope tag (**Figure. 3-1B, Supplementary Figure. A3-1A**). We crossed the FOXA1 transgene harboring female mice with the well-characterized Pb-Cre4 males (*38*) to obtain *FOXA1;Pb-Cre* genotypes, where *FOXA1+/+;Pb-Cre+* males were denoted as "case" while littermate *FOXA1+/+;Pb-Cre-* males were used as controls (**Figure. 3-1B**). As expected, we saw specific and robust nuclear staining of V5 in prostate epithelial cells in case relative to the control prostate tissue (**Figure. 3-1B**), which looked histologically normal at 10-12 weeks of age (**Supplementary Figure. A3-1B**). However, histopathological evaluation of older animals revealed FOXA1 Class 1 mutants to drive hyperplasia in the epithelial lining, with intraepithelial neoplasia (PIN) spanning 20-25% of the tissue by 60 weeks of age. We also detected focal high-grade, Ki67-positive intra-ductal carcinomas (IDCs) in 25% of the cases. Similar hyper-proliferative lesions were not detected in the control tissues (**Supplementary Figure. A3-1C-E**).

Thus, FOXA1 Class 1 mutants are able to initiate transformation of the prostate epithelia, albeit with a long latency and incomplete penetrance.

Next, assessment of genomic co-alterations from metastatic tumors uncovered FOXA1 Class 1 mutations to often co-occur with biallelic loss of *TP53* while being mutually exclusive to *PTEN* deletions, when compared to other truncal alterations (**Supplementary Figure. A3-1F**). Thus, we generated compound two-gene models by crossing FOXA1 Class 1 mice with transgenic mice harboring floxed, inactivating alleles of the *Trp53* tumor suppressor gene (**Figure. 3-1B, Supplementary Figure. A3-1G**). Strikingly, in prostate tissues from the *FOXA1 R265-71del+/+;Trp53f/f;Pb-Cre+* (i.e., case) mice, we detected widespread hyper-stratified epithelial lesions with enlarged nuclei in the anterior prostate starting around the 40-week timepoint (**Figure. 3-1C,D**). Littermate control mice lacking Cre recombinase largely showed a normal prostate histology denoted by a monolayer epithelium. Notably, with age, the cancerous lesions expanded to cover a larger tissue area as well as exhibited characteristics of high-grade disease (**Figure. 3-1C,E**), as assessed using our in-house grading system (see Methods and **Supplementary Figure. A3-1H**). Ki67 staining revealed 30-40% of the cells within cancerous lesions to be actively proliferating (**Figure. 3-1F, Supplementary Figure. A3-1I**). At 80 weeks, case animals had a significant enlargement of the urogenital tract (**Supplementary Figure. A3-1J**), and the prostate tissues showed expansive, luminous glandular structures that were almost entirely encapsulated by high-grade intraductal adenocarcinoma (**Supplementary Figure. A3-1K**). Importantly, these lesions strongly expressed the FOXA1 R265-71del mutant (detected using the V5 tag) along with luminal markers like AR and CK8 (**Figure. 3-1G**), akin to primary prostatic tumors detected in humans.

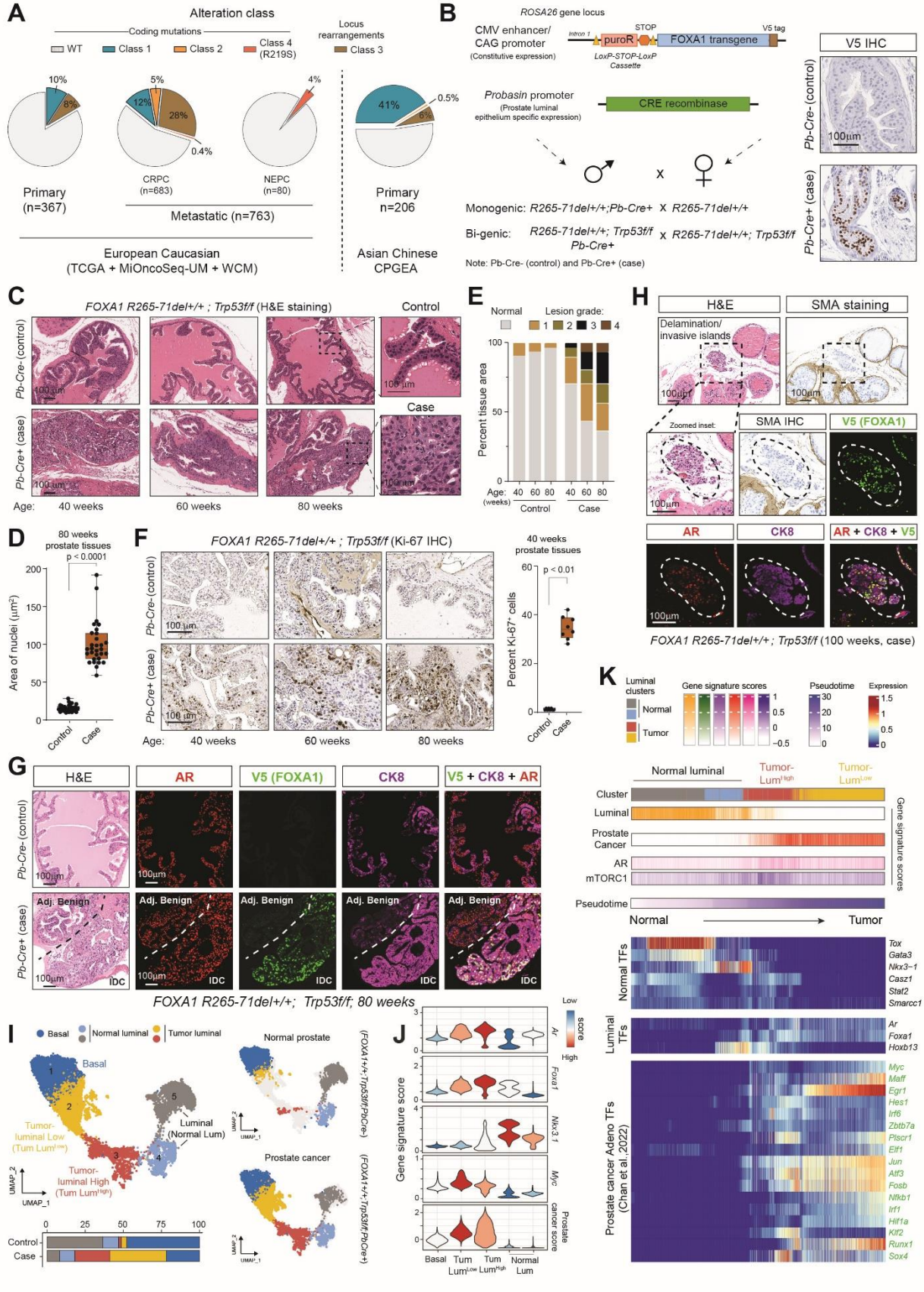

A
Alteration class
Coding mutations
WT
Class 1
Class 2
Class 4 (R219S)
Locus rearrangements
Class 3
10%
8%
5%
12%
28%
0.4%
4%
41%
0.5%
6%
Primary (n=367)
CRPC (n=683)
NEPC (n=80)
Metastatic (n=763)
Primary n=206
European Caucasian (TCGA + MiOncoSeq-UM + WCM)
Asian Chinese CPGEA
B
ROSA26 gene locus
CMV enhancer/ CAG promoter
(Constitutive expression)
Intron 1
STOP
puroR
FOXA1 transgene
V5 tag
LoxP-STOP-LoxP Cassette
Probasin promoter
(Prostate luminal epithelium specific expression)
CRE recombinase
Monogenic: R265-71del+/+;Pb-Cre+ X R265-71del+/+
Bi-genic: R265-71del+/+; Trp53f/f Pb-Cre+ X R265-71del+/+; Trp53f/f
Note: Pb-Cre- (control) and Pb-Cre+ (case)
V5 IHC
Pb-Cre- (control)
Pb-Cre+ (case)
100µm
C
FOXA1 R265-71del+/+ ; Trp53f/f (H&E staining)
Control
Case
Pb-Cre- (control)
Pb-Cre+ (case)
100 µm
Age: 40 weeks 60 weeks 80 weeks
E
Normal
Lesion grade: 1 2 3 4
Percent tissue area
Age: (weeks) 40 60 80 40 60 80
Control Case
H
H&E
SMA staining
Delamination/ invasive islands
Zoomed inset:
SMA IHC
V5 (FOXA1)
AR
CK8
AR + CK8 + V5
100µm
FOXA1 R265-71del+/+; Trp53f/f (100 weeks, case)
D
80 weeks prostate tissues
p < 0.0001
Area of nuclei (µm²)
Control
Case
F
FOXA1 R265-71del+/+ ; Trp53f/f (Ki-67 IHC)
Pb-Cre- (control)
Pb-Cre+ (case)
100 µm
Age: 40 weeks 60 weeks 80 weeks
40 weeks prostate tissues
p < 0.01
Percent Ki-67+ cells
K
Luminal clusters
Normal
Tumor
Gene signature scores
Pseudotime
Expression
G
H&E
AR
V5 (FOXA1)
CK8
V5 + CK8 + AR
Pb-Cre- (control)
Pb-Cre+ (case)
Adj. Benign
IDC
100µm
FOXA1 R265-71del+/+; Trp53f/f; 80 weeks
Normal luminal
Tumor-Lum^High
Tumor-Lum^Low
Cluster
Luminal
Prostate Cancer
AR
mTORC1
Pseudotime
Normal
Tumor
Gene signature scores
Normal TFs
Tox
Gata3
Nkx3–1
Casz1
Stat2
Smarcc1
Luminal TFs
Ar
Foxa1
Hoxb13
Prostate cancer Adeno TFs (Chan et al. 2022)
Myc
Maff
Egr1
Hes1
Irf6
Zbtb7a
Plscr1
Elf1
Jun
Atf3
Fosb
Nfkb1
Irf1
Hif1a
Klf2
Runx1
Sox4
I
Basal
Normal luminal
Tumor luminal
Basal
Tumor-luminal Low (Tum Lum^Low)
Luminal (Normal Lum)
Tumor-luminal High (Tum Lum^High)
UMAP_1
UMAP_2
Control
Case
Normal prostate
Prostate cancer
(FOXA1+/+;Trp53f/f;PbCre+)
J
Low
High
score
Ar
Foxa1
Nkx3.1
Myc
Prostate cancer score
Gene signature score
Basal
Tum Lum^Low
Tum Lum^High
Normal Lum

Figure 3-1: FOXA1 Class 1 mutants drive high-grade invasive adenocarcinoma in the background of p53 loss. A) Frequency of FOXA1 alterations in the European (primary and metastatic) and Asian (primary) prostate cancer patient cohorts. FOXA1 Class 1 mutations are truncal alterations seen in the primary and metastatic disease (10-12%) and comprise in-frame indels and missense mutations in the wing2 region of the protein. Class 2 mutations are acquired in the metastatic disease and comprise truncating frameshift alterations in the C-terminal region of the protein. Class 3 mutations are structural alterations (duplications and translocations) within the *FOXA1* syntenic loci found in the primary and metastatic disease (8-28%). FOXA1 Class 4 mutations comprise the missense R219S alteration that is primary enriched in NEPC tumors (4%). Indel, insertion, and deletion; CRPC, castration-resistant prostate cancer; NEPC, neuroendocrine prostate cancer; U-M, University of Michigan; WCM, Weill Cornell Medicine; CPGEA: Chinese Prostate Cancer Genome and Epigenome Atlas. B) Left: Schematic of FOXA1 Class 1 (*R265-71del*) knock-in transgenic mouse models and mating strategies for monogenic (*R265-71del+/+*) and bigenic (*R265-71del+/+; Trp53f/f*) mouse lines. Probasin-Cre (Pb-Cre) ensures prostate-specific expression of FOXA1. Control and case animals are denoted as *Pb-Cre-* and *Pb-Cre+*, respectively. Right: V5-epitope tag immunohistochemistry (IHC) in control (*Pb-Cre-*) and case (*Pb-Cre+*) *R265-71del+/+* animals. Scale=100µm. C) Representative Hematoxylin and Eosin (H&E) stained cross-sections of *FOXA1 R265-71del+/+;Trp53f/f* control and case prostate tissues (anterior lobe) from 40 to 80 week-old mice. Black-dotted boxes denote zoomed insets. Scale=100µm. D) Area of nuclei in 80-week-old control and case *FOXA1 R265-71del+/+;Trp53f/f* prostate tissues (two-tailed t-test). E) Histopathological grading of tissues from panel c. Grade 1: Hyperplasia, Grade 2: Low/high-grade prostate intraepithelial neoplasia (PINs), Grade 3: Florid high-grade PINs and/or Atypical Intraductal Proliferation (AIP) and/or Intraductal carcinoma (IDC), Grade 4: Invasive adenocarcinoma. F) Left: Ki67 IHC in control and case *FOXA1 R265-71del+/+;Trp53f/f* prostate tissues at 40, 60, and 80 weeks. (Scale=100µm). Right: Statistical representation of quantification (percentage of KI67-positive cells) from 40-week-old tissues from the left panel (two-tailed t-test). Box plot: center line, median; box, interquartile range (Q1–Q3); whiskers, minimum to maximum values; all individual data points shown. G) Representative H&E and multiplex immunofluorescence of V5, AR, and CK8 in *FOXA1 R265-71del+/+;Trp53f/f* 80-week-old prostate tissues. Adjacent benign and intraductal carcinoma (IDC) tumor regions are marked. Scale=100µm. H) Representative H&E, alpha-smooth muscle actin (SMA) IHC and multiplex immunofluorescence of noted proteins in *FOXA1 R265-71del+/+;Trp53f/f* 100-week-old tissues. Invasive islands denoting breach of basement membrane (lack of SMA) are highlighted in a black, dotted box. Scale=100µm. I) Left: Uniform Manifold Approximation and Projection (UMAP) plots from single-cell RNA-seq data of *FOXA1 R265-71del+/+;Trp53f/f* 100-week-old anterior and dorsal prostate lobes. Tumor cell clusters are highlighted in yellow and red. Right: Split UMAP plots with only case or control tissue-derived cells colored across clusters. J) Violin plots showing expression of *Ar, Foxa1, Nkx3.1, Myc* and a prostate cancer score (defined in this paper; see Methods) in distinct cell populations described in panel I. K) Heatmap of select transcription factor (TF) gene expression along the pseudotime (Slingshot, see Methods) from normal luminal cells to tumor cells derived from *FOXA1 R265-71del+/+;Trp53f/f* control and case prostate tissues. The top heatmaps show aggregate z-scores of luminal, AR, mTORC1, and prostate cancer signature genes.

Strikingly, starting around 60 weeks of age, *FOXA1 R265-71del+/+;Trp53f/f;Pb-Cre+* mice also showed signs of invasion of the adenocarcinoma into the surrounding stromal regions (denoted as grade 4 lesions; **Figure. 3-1E and Supplementary Figure. A3-2A**). These invasive lesions were identified by breach of the encapsulating SMA-positive fibromuscular layer and detection of delaminated stromal islands of epithelial tumor cells co-expressing CK8, AR, and the FOXA1 transgene (**Figure. 3-1H, Supplementary Figure. A3-2B**). In a single 100-week-old case animal,

we also detected a large tumor mass with adenocarcinoma histology and strong expression of transgenic FOXA1 mutant protein (**Supplementary Figure. A3-2C,D**). In summary, FOXA1 Class 1 mutants in absence of p53 drive the formation of luminal adenocarcinoma by 40 weeks of age, which invade into the surrounding stromal regions upon disease progression.

Next, to mechanistically characterize the *FOXA1 R265-71del+/+;Trp53f/f* prostatic tumors, we performed single-cell RNA-sequencing of anterior and dorsal prostate lobes harvested from 100 week-old case and control animals (n = 2 per genotype; **Supplementary Figure. A3-3A**). We subset epithelial cells where, as expected, we found a specific loss of the puroR transcript in Cre-expressing cells derived from the case animals (**Supplementary Figure. A3-3B,C**). Consistent with the histological data, we found the *FOXA1 R265-71del+/+;Trp53f/f;Pb-Cre+* prostate tissues to have a 2-fold higher fraction of epithelial cells in the DNA replication (i.e., S) phase of the cell cycle, indicative of hyper-proliferation (**Supplementary Figure. A3-3D**). Louvain clustering identified five distinct epithelial cell clusters, with clusters 2 and 3 comprising cells predominantly coming from the cancerous tissues (**Figure. 3-1I**). Remarkably, these cancer-specific epithelial clusters showed highest expression of both *Ar* and mouse *Foxa1* genes (**Figure. 3-1J**). Simultaneously, these cells showed a distinct molecular signature indicative of malignant transformation, characterized by a significant downregulation of the prostate-specific tumor suppressor gene *Nkx3.1* and aberrant upregulation of the oncogene *Myc* (**Figure. 3-1J**). We were able to confirm a marked decrease in NKX3.1 protein expression within cancerous lesions relative to the adjacent normal prostate epithelia (**Supplementary Figure. A3-3E**). Thus, we labeled the tumor-specific epithelial clusters as luminal high (cluster 3) and low (cluster 2), reflecting expression levels of Ar and Foxa1 luminal markers (**Figure. 3-1I,J**). Through differential expression analyses between tumor versus normal luminal clusters, we also defined a 50-gene signature of FOXA1 Class 1-driven mouse prostate adenocarcinoma, which was specifically activated in the tumor luminal cells (**Supplementary Figure. A3-3F**). Cross-species analyses revealed reciprocal enrichment of FOXA1 Class 1-associated gene signatures between human and mouse prostate tumors, with human Class 1 mutant PCa genes significantly upregulated in murine models and vice versa (**Supplementary Figure. A3-3G,H**). Pathway analyses demonstrated conserved activation of hyperproliferative programs—including E2F/G2M cell cycle progression, mitotic spindle formation, MYC signaling—as well as activation of luminal androgen response

and PI3K-mTORC1 pathways in both human and mouse Class 1-mutant tumors (**Supplementary Figure. A3-3I**). Next, to characterize sequential changes in the transcriptional landscape upon FOXA1 Class 1 mutation triggered transformation of the normal prostate epithelia, we performed Slingshot pseudotime trajectory analysis (**Supplementary Figure. A3-3J**). This analysis confirmed the loss of *Nkx3.1* expression upon emergence of PCa with a parallel increase in *Ar* and *Foxa1* levels (**Figure. 3-1K**). Most strikingly, we found an aberrant gain of a host of AR+ adenocarcinoma-associated transcription factors that were recently identified using the *Pten* deletion-triggered mouse PCa models (*39*) (**Figure. 3-1K**).

Altogether, through detailed histopathological and single-cell transcriptomic characterization, we show that overexpression of FOXA1 Class 1 mutants along with p53 inactivation in the prostate epithelia triggers the formation of AR+ adenocarcinoma that strongly recapitulates luminal features of the primary human disease. This also constitutes the first experimental evidence causally linking FOXA1 Class 1 mutations to prostate tumorigenesis, thereby substantiating these early genetic alterations as bona fide drivers of human PCa.

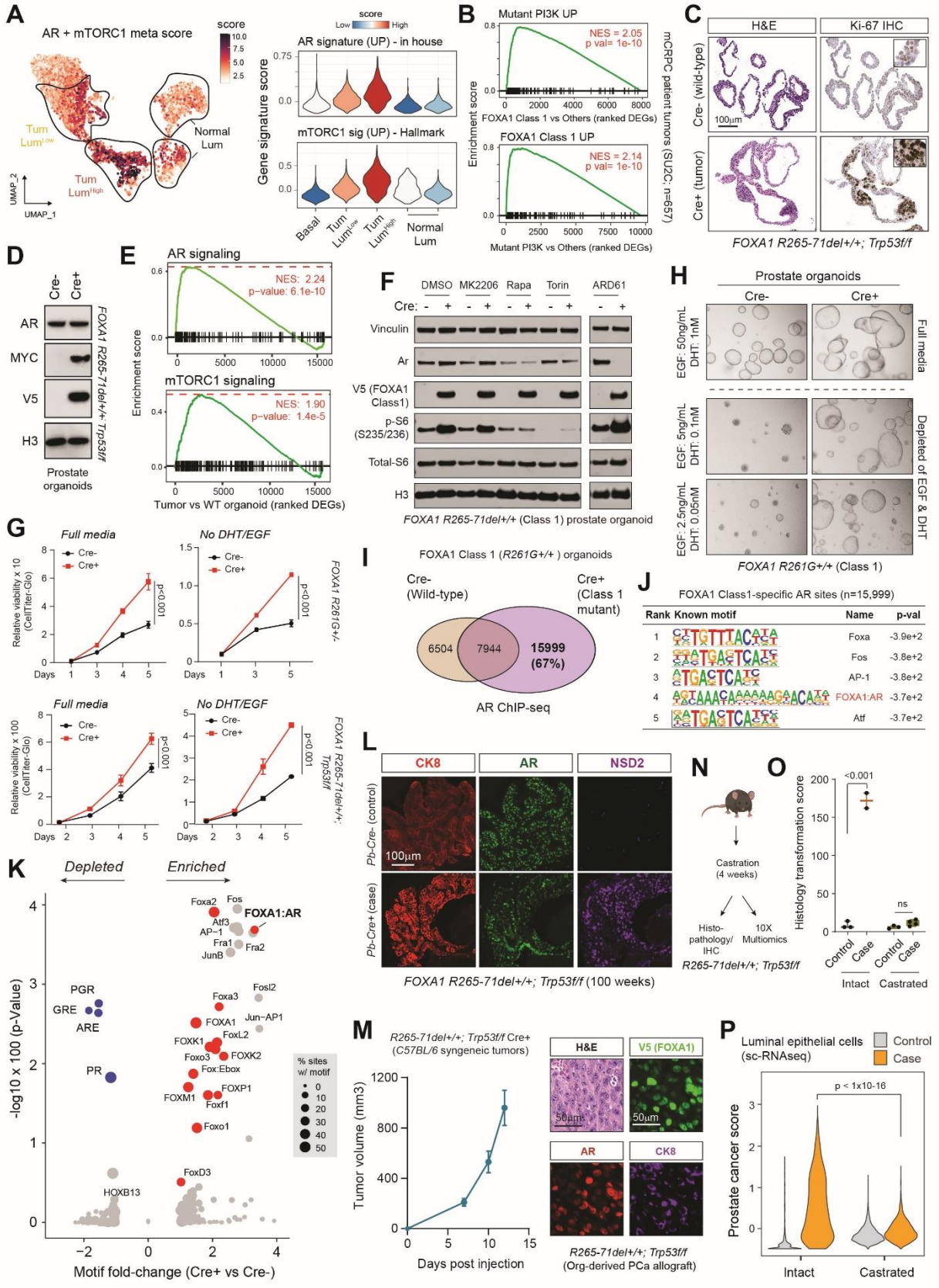

A
AR + mTORC1 meta score
score
Tum Lum^Low
Tum Lum^High
Normal Lum
UMAP_2
UMAP_1
AR signature (UP) - in house
mTORC1 sig (UP) - Hallmark
Gene signature score
Basal
Tum Lum^Low
Tum Lum^High
Normal Lum
B
Mutant PI3K UP
NES = 2.05
p val= 1e-10
FOXA1 Class 1 vs Others (ranked DEGs)
FOXA1 Class 1 UP
NES = 2.14
p val= 1e-10
Mutant PI3K vs Others (ranked DEGs)
Enrichment score
mCRPC patient tumors (SU2C; n=657)
C
H&E
Ki-67 IHC
Cre- (wild-type)
Cre+ (tumor)
100μm
FOXA1 R265-71del+/+; Trp53f/f
D
Cre-
Cre+
AR
MYC
V5
H3
FOXA1 R265-71del+/+ Trp53f/f
Prostate organoids
E
AR signaling
NES: 2.24
p-value: 6.1e-10
mTORC1 signaling
NES: 1.90
p-value: 1.4e-5
Enrichment score
Tumor vs WT organoid (ranked DEGs)
F
DMSO
MK2206
Rapa
Torin
ARD61
Cre: - + - + - + - + - +
Vinculin
Ar
V5 (FOXA1 Class1)
p-S6 (S235/236)
Total-S6
H3
FOXA1 R265-71del+/+ (Class 1) prostate organoid
H
Prostate organoids
Cre-
Cre+
EGF: 50ng/mL DHT: 1nM
EGF: 5ng/mL DHT: 0.1nM
EGF: 2.5ng/mL DHT: 0.05nM
Full media
Depleted of EGF & DHT
FOXA1 R261G+/+ (Class 1)
G
Full media
No DHT/EGF
Cre-
Cre+
Relative viability x 10 (CellTiter-Glo)
Relative viability x 100 (CellTiter-Glo)
Days
p<0.001
FOXA1 R261G+/+
FOXA1 R265-71del+/+; Trp53f/f
I
FOXA1 Class 1 (R261G+/+) organoids
Cre- (Wild-type)
Cre+ (Class 1 mutant)
6504
7944
15999 (67%)
AR ChIP-seq
J
FOXA1 Class1-specific AR sites (n=15,999)
Rank
Known motif
Name
p-val
1 Foxa -3.9e+2
2 Fos -3.8e+2
3 AP-1 -3.8e+2
4 FOXA1:AR -3.7e+2
5 Atf -3.7e+2
K
Depleted
Enriched
-log10 x 100 (p-Value)
Motif fold-change (Cre+ vs Cre-)
FOXA1:AR
Foxa2
Fos
Atf3
AP-1
Fra1
Fra2
JunB
PGR
GRE
ARE
PR
Fosl2
Foxa3
FOXA1
Jun-AP1
FoxL2
FOXK1
FOXK2
Foxo3
Fox:Ebox
FOXP1
FOXM1
Foxf1
Foxo1
FoxD3
HOXB13
% sites w/ motif
L
CK8
AR
NSD2
Pb-Cre- (control)
Pb-Cre+ (case)
100μm
FOXA1 R265-71del+/+; Trp53f/f (100 weeks)
N
Castration (4 weeks)
Histo-pathology/ IHC
10X Multiomics
R265-71del+/+; Trp53f/f
O
Histology transformation score
<0.001
ns
Control
Case
Intact
Castrated
M
R265-71del+/+; Trp53f/f Cre+ (C57BL/6 syngeneic tumors)
Tumor volume (mm3)
Days post injection
H&E
V5 (FOXA1)
AR
CK8
50μm
R265-71del+/+; Trp53f/f (Org-derived PCa allograft)
P
Luminal epithelial cells (sc-RNAseq)
Control
Case
p < 1x10-16
Prostate cancer score
Intact
Castrated

Figure 3-2: FOXA1 Class 1 mutants co-activate AR and mTORC1 oncogenic signaling. A) Left: UMAP plots from single-cell RNA-seq data of *R265-71del+/+;Trp53f/f* tissues showing the meta-score of AR and mTORC1 gene signatures. Right: Violin plots of AR and mTORC1 gene signature scores in the tumor and normal clusters. B) Gene set enrichment analysis (GSEA) plots for PI3K and FOXA1 Class 1 activated genes from FOXA1 Class 1 or PI3K mutant patient tumor tissues. DEGS, differentially expressed genes (GSEA enrichment test). C) Representative H&E and Ki67 IHC in Class 1 *R265-71del+/+;Trp53f/f* isogenic organoids. *Cre-* organoids are wild-type/normal organoids, while *Cre+* organoids are tumor organoids. Scale=100µm. D) Immunoblot of labeled proteins in the Cre- and Cre+ tumor organoids. Total H3 is used as a loading control. E) GSEA plots of AR and mTORC1 up-regulated genes using the fold change from *FOXA1 R265-71del+/+;Trp53f/f* Cre+ vs Cre- tumor organoids. DEGS, differentially expressed genes. (n=2 biological replicates, GSEA enrichment test). F) Immunoblot of labeled proteins in *R265-71del+/+* Cre- and Cre+ organoids treated with MK2206 (AKT inhibitor), rapamycin (mTORC1 inhibitor), torin (mTORC1/2 inhibitor), or ARD61 (AR degrader) for 24 hours. G) Growth curves (cell titer glow) of Class 1 single gene (*R261G+/-*) and two-gene (*R265-71del+/+; Trp53f/f*) Cre- and Cre+ organoid lines grown in full media or media lacking DHT and EGF (n=4 biological replicates, two-sided t-test, mean with SD is shown). H) Representative images of Class 1 (*R261G+/+*) Cre- and Cre+ organoids grown in complete media or media depleted of EGF and DHT. I) Venn diagram showing overlaps of AR ChIP-seq in Cre- and Cre+ Class 1 *R261G+/+* organoids. J) Top five known HOMER motifs (ranked by p-value) enriched within Cre+, i.e., Class 1 specific AR binding sites. (HOMER, hypergeometric test). K) Fold change and significance of HOMER motifs enriched within Cre+ AR sites over Cre- AR binding sites (HOMER, hypergeometric test). L) Representative multiplex-IF images of CK8, AR, and NSD2 in anterior lobes of FOXA1 Class 1 *R265-71del+/+;Trp53f/f* tumors. Scale = 100µm. M) Left: Tumor volumes of syngeneic *R265-71del+/+;Trp53f/f* Cre+ tumor organoids grafted in the subcutaneous flanks of *C57BL/6* mice. Right: Representative H&E and multiplex-IF of labeled proteins in tumor allografts obtained from the left panel. Scale =50µm. N) Schematic of castration experiment performed in 60-week-old *FOXA1 R265-71del+/+;Trp53f/f* mice. O) Transformation score (defined in this paper, see methods) in 60-week-old intact and castrated *FOXA1 R265-71del+/+;Trp53f/f* prostate tissues (two-sided t-test). Box plot: center line, median; box, interquartile range (Q1–Q3); whiskers, minimum to maximum values; all individual data points shown. P) Prostate cancer score in single-cell RNA-seq from 60-week-old intact and castrated *FOXA1 R265-71del+/+;Trp53f/f* prostate tissues (two-sided Wilcoxon test).

FOXA1 Class 1 mutants co-activate AR and mTORC1/2 tumorigenic pathways

We next set out to delineate the molecular mechanism through which FOXA1 Class 1 mutants drive prostate adenocarcinoma. First, we derived pure, pseudo-bulk transcriptional profiles of normal and cancerous luminal epithelial cells by combining gene marker-verified cells from case and control prostate tissues (see Methods). As expected, gene signature enrichment analysis of genes differentially expressed between cancer and normal luminal epithelia showed a significant activation of hyper-proliferative and other cancer-associated hallmark pathways (**Supplementary Figure. A3-4A**). Serving as an intrinsic control, the same cancer pathways were not differentially expressed in the basal epithelia between case and control tissues (**Supplementary Figure. A3-4A**). More strikingly, we found a significant and parallel up-regulation of both AR and mTORC1/2 signaling in case cells, as evidenced by the higher AR+mTORC1 meta-scores in the tumor luminal

clusters (**Figure. 3-2A**, see Methods). To further evaluate this transcriptional association in human clinical samples, we compared bulk transcriptomes in Caucasian primary (TCGA, n=503) and CRPC tumors (SU2C cohort, n=657) (*3*) harboring FOXA1 Class 1 mutations or PI3K pathway alterations, the latter mostly comprising PTEN deletions. Our analysis revealed a significant concordance in differentially expressed genes between the two genomic subtypes (**Supplementary Figure. A3-4B-D**), showing a reciprocal gene signature enrichment pattern. The PI3K alteration activated gene signature was significantly up-regulated in Class 1-mutant tumors, while conversely, the FOXA1 Class 1 gene program was overexpressed in PI3K-altered cases (**Figure. 3-2B**). Similarly, integrative multi-omics analyses in Asian primary PCa (CPGEA cohort; n=206), where FOXA1 Class 1 is a pre-dominant driver alteration in over 41% of the cases (**Figure. 3-1A**), revealed a strong activation of AR signaling in Class 1-mutant tumors (**Supplementary Figure. A3-4E**). We also saw a significant concordance in differentially expressed genes between the Class 1-mutant and PI3K-altered PCa subtypes (**Supplementary Figure. A3-4F,G**). Altogether, these findings suggest that concomitant activation of AR and PI3K-mTORC1/2 signaling is a salient feature of FOXA1 Class 1-driven prostate carcinogenesis that is conserved across human patient ethnicities as well as our mouse models.

For further experimental interrogation, we established prostate epithelial organoids from the *FOXA1 R265-71del+/+;Trp53f/f* transgenic mice. We transduced these organoids with either an empty-vector control (annotated as Cre-) or Cre recombinase-coding (Cre+) adenovirus to establish isogenic lines (**Supplementary Figure. A3-4H,** left), with Cre recombinase activity triggering expression of the V5-tagged FOXA1 protein (**Supplementary Figure. A3-4H,** right). Compared to the Cre- wild-type lines, *FOXA1 R265-71del+/+;Trp53f/f;Cre+* organoids showed hyper-stratification of the epithelial lining and a significantly higher proportion of Ki67-positive cells (**Figure. 3-2C**), while retaining their cystic, luminal morphology. Transcriptomic profiling revealed activation of several hallmark cancer pathways, which were concordantly activated in mouse PCa tissues (**Supplementary Figure. A3-4I,J**). This notably included the increased expression of MYC and its downstream oncogenic gene programs associated with malignant transformation (**Figure. 3-2D, Supplementary Figure. A3-4I**). Even in the organoid models, we found a significant co-activation of AR and mTORC1 signaling (**Figure. 3-2E**), credentialing them for use in further mechanistic studies.

To define putative FOXA1 targets, we first performed FOXA1 ChIP-seq in our isogenic wild-type and Class 1-mutant mouse prostate organoids and identified genes that show FOXA1 binding within their putative *cis*-regulatory elements using the GREAT algorithm (*40*) (**Supplementary Figure. A3-5A**). These predicted targets were then intersected with genes that are significantly up-regulated in either FOXA1 Class 1-mutant *R261G+/+* prostate organoids or the *FOXA1 R265-71del+/+;Trp53f/f;Cre+* tumor line (**Supplementary Figure. A3-5B**), revealing 141 likely direct targets of Class 1 mutants of FOXA1 (**Supplementary Table 5**). Enrichr pathway analysis (*41*) of this gene set showed a significant enrichment of many cancer-associated MSigDB pathways, including epithelial-mesenchymal transitions (EMT), hypoxia, and glycolysis. Most notably, putative FOXA1 targets were also enriched for genes comprising the mTORC1 signature (**Supplementary Figure. A-5C**), which is consistent with our prior findings of PI3K-mTORC1 activation being a key mechanism for FOXA1 Class 1-driven prostate tumorigenesis.

Next, to discern specific contributions of Class 1 mutant overexpression and p53 inactivation in driving tumorigenesis, we employed the *FOXA1 R261G* and *R265-71del* single-gene organoids. Here, overexpression of Class 1 mutants alone was sufficient to parallelly activate both AR and mTORC1 pathways (**Supplementary Figure. A3-5D,E**), along with the E2F/G2M proliferative, glycolysis, and other oncogenic gene programs (**Supplementary Figure. A3-5F**). In organoid models, however, CRISPR-mediated inactivation of p53 alone failed to significantly modulate these oncogenic pathways (**Supplementary Figure. A3-5F**). Consistent with our *in vivo* findings, histological assessment of organoid morphology also showed profound hyper-stratification of the normal epithelial lining only in the compound *FOXA1 R261G+/-;Trp53-KO* models (**Supplementary Figure. A3-5G**). p53-null organoids, however, more closely resembled the histology of isogenic wild-type (i.e., Cre-) models. The *FOXA1 R261G+/-;Trp53-KO* organoids also had the highest percent of Ki67 positive cells and showed the fastest growth when compared to the normal or FOXA1 R261G-expressing isogenic lines (**Supplementary Figure. A3-5H-J**). Altogether, these findings credential FOXA1 Class 1 mutations to serve as initiators of malignant transcriptional programs, with subsequent loss of *Trp53* acting as a potent promoting alteration in driving PCa formation and progression.

Expression of FOXA1 Class 1 mutants in organoids also led to a strong and dose-dependent increase in pS473-AKT and pS6 (S235/236), which are direct substrates of mTORC1/2 kinases(*42*,

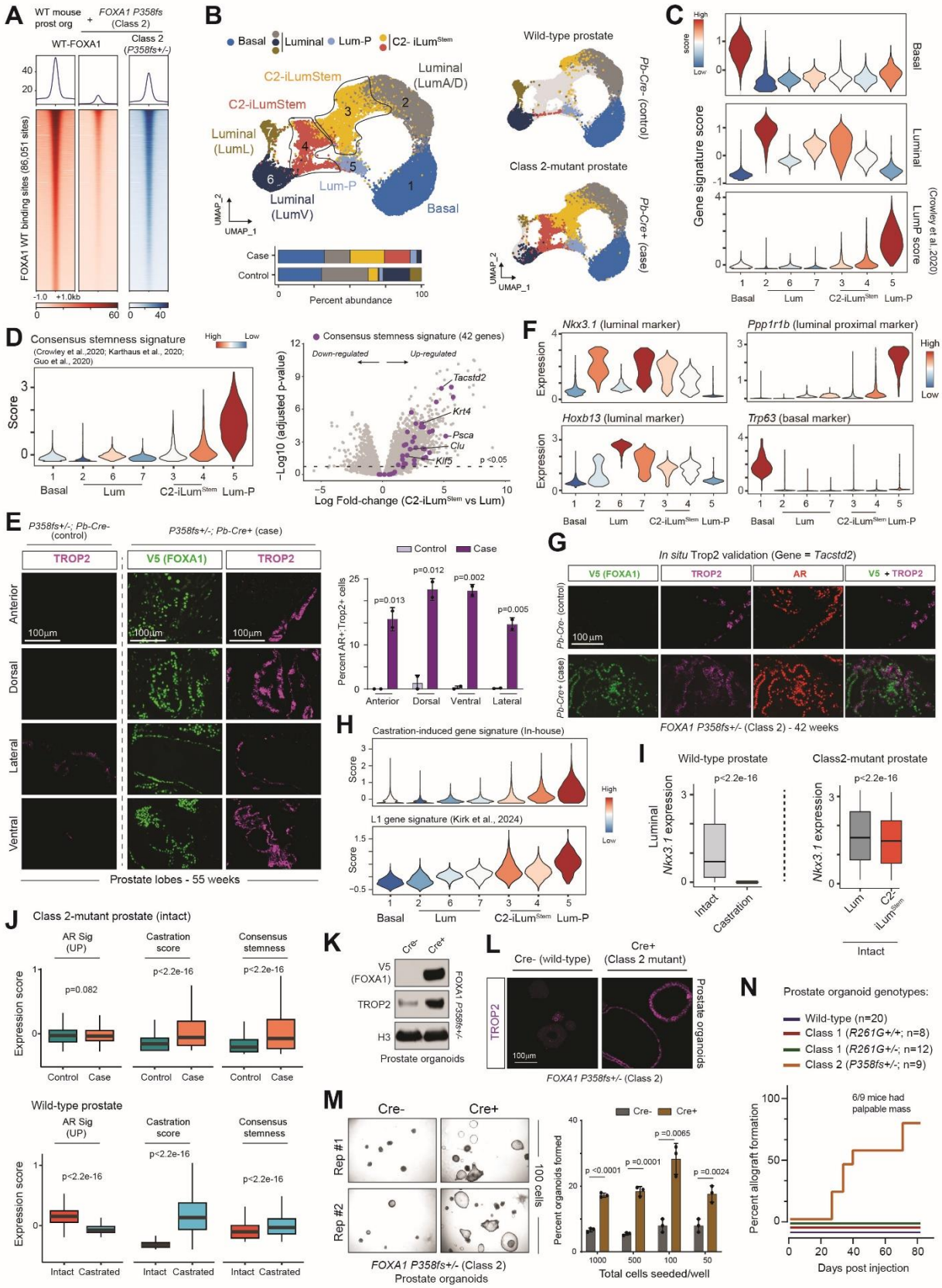
43) (**Figure. 3-2F, Supplementary Figure. A3-5K**). This was corroborated in human clinical specimens, where primary PCa from both Caucasian (*2*) and Asian (*44*) men showed significantly elevated levels of phosphorylated RPS6 at serine 235/236 and 240/244 residues in FOXA1 Class 1-mutant tumors compared to FOXA1 wild-type cases, with no change in total RPS6 levels (**Supplementary Figure. A3-5L,M**). Intriguingly, we found treatment with mTORC1 and/or mTORC2 inhibitors (i.e., rapamycin and torin) completely blocked Class 1-induced accumulation of p-S6, while treatment with an AKT inhibitor (MK2206) had little to no effect (**Figure. 3-2F**). We also found that AR degradation in Class 1 organoids had no effect on the induction of p-S6 (**Figure. 3-2F**). Overexpression of wild-type FOXA1 (mimicking Class 3 amplifications) also triggered a strong induction of p-S6 (**Supplementary Figure. A3-5K**). In growth assays, we found single-gene and compound models of Class 1-expressing prostate organoids to grow better than their wild-type counterparts in complete media. This enhanced proliferation of Class 1-mutant organoids was even more profound in growth media lacking stimulatory ligands of AR and PI3K/mTORC1 pathways, namely DHT and EGF (**Figure. 3-2G,H**). This suggests that FOXA1 Class 1 mutants function as hypermorphs and activate mTORC1/2 signaling independently of AKT as well as AR activities.

Next, we asked how FOXA1 Class 1 mutants modulate AR signaling. We first performed AR ChIP-seq in Class 1 mutant-expressing or wild-type (i.e., Cre-) control prostate organoids. Here, in the presence of the Class 1 mutant, we detected an extensive reprogramming and expansion of the AR cistrome, with almost 16,000 new *cis*-regulatory elements gained within distal intergenic and intronic regions (**Figure. 3-2I, Supplementary Figure. A3-6A**). HOMER motif analysis of Class 1-specific AR sites uncovered an enrichment of FOXA1 and chimeric FOXA1:AR-half motifs (**Figure. 3-2J,K**). In contrast, full palindromic androgen response elements (ARE) and related nuclear receptor motifs were enriched in the wild-type-specific, normal AR cistrome (**Figure. 3-2K**, **Supplementary Figure. A3-6B**). A recent study from our group identified chimeric AR half-sites as a distinctive feature of tumor-specific AR enhancer circuitries, which is activated by NSD2 as a neo-coactivator of AR (*12*). In line with these findings, we found that NSD2 expression was induced in Class 1-positive PCa lesions, with no detectable expression in control tissues (**Figure. 3-2l, Supplementary Figure. A3-6D**). NSD2 expression was further elevated within invasive tumor edges and delaminated stromal islands (**Supplementary Figure.**

A3-6E). In derivative tumor organoids, proximity ligation assay revealed strong nuclear interaction between NSD2 and AR proteins, which was completely lost in isogenic lines with CRISPR-mediated knock-out (KO) of NSD2 (**Supplementary Figure. A3-6F-I**). Loss of NSD2 in the tumor organoids attenuated growth and significantly dampened AR signaling (**Supplementary Figure. A3-6J,K**), with chromatin binding of AR disrupted at over 85% of its distal and intergenic binding sites (**Supplementary Figure. A3-6L,M**). Next, HOMER motif analysis of NSD2-dependent AR sites uncovered FOXA1:AR-half chimeric motif as the topmost significantly enriched motif, while the NSD2-independent sites housed the 15 bp canonical, palindromic androgen response element (ARE; **Supplementary Figure. A3-6N,O**). Consistently, quantitation of NSD2-catalyzed H3K36me2 signal from ChIP-seq showed a significant depletion at nucleosomes flanking the NSD2-dependent AR binding sites in the NSD2-KO models (**Supplementary Figure. A3-6P**). These findings are consistent with our prior observations in human PCa (*12*), and suggest that FOXA1 Class 1 mutants pioneer neo-enhancer elements of AR that harbor non-palindromic, chimeric AR-half motifs and, in concert with NSD2, activate a distinct luminal gene program of prostate tumorigenesis

Further validating their malignant phenotype, *FOXA1 R265-71del+/+;Trp53f/f;Pb-Cre+* tumor organoids, but not the isogenic *Pb-Cre-* controls, formed large tumors upon subcutaneous implantation into NSG mice (**Supplementary Figure. A3-7A**). The tumor organoids also established syngeneic allografts in the C57BL/6 host mice retaining their adenocarcinoma histology as well as expression of transgenic FOXA1 and AKT and CK8 luminal markers (**Figure. 3-2M)**. Human PCa is treated with ADT as the first line of targeted therapies (*45*, *46*). Thus, we assessed if FOXA1 Class 1-driven tumors were sensitive to androgen withdrawal by performing surgical castration on 50-65 week-old *FOXA1 R265-71del+/+;Trp53f/f* case and control animals (n=3-4 per genotype), followed by histological grading and single-cell RNA sequencing of residual tissues four weeks later (**Figure. 3-2N**). In contrast to widespread tumors detected in the hormonally-intact case animals, castrated animals had no atypical lesions that stained for Ki67 (**Figure. 3-2O, Supplementary Figure. A3-7B,C**). More compelling, epithelial cells from case tissues showed a significant loss of the FOXA1 Class 1-driven PCa gene signature upon castration relative to the intact condition (**Figure. 3-2P**). Altogether, our findings show that FOXA1 Class 1 mutations promoted by p53 inactivation trigger the formation of high-grade and invasive prostate

adenocarcinoma that retain luminal features as well as regress upon castration, recapitulating cardinal characteristics of the primary human disease.


A
WT mouse prost org
FOXA1 P358fs (Class 2)
WT-FOXA1
Class 2 (P358fs+/-)
FOXA1 WT binding sites (86,051 sites)
B
Basal
Luminal
Lum-P
C2- iLumStem
C2-iLumStem
Luminal (LumA/D)
Luminal (LumL)
Lum-P
Luminal (LumV)
Basal
Case
Control
Percent abundance
Wild-type prostate
Class 2-mutant prostate
Pb-Cre- (control)
Pb-Cre+ (case)
C
Gene signature score
Basal
Luminal
LumP score
(Crowley et al.,2020)
Basal
Lum
C2-iLumStem
Lum-P
D
Consensus stemness signature
(Crowley et al.,2020; Karthaus et al., 2020; Guo et al., 2020)
Score
Consensus stemness signature (42 genes)
Down-regulated
Up-regulated
Tacstd2
Krt4
Psca
Clu
Krt5
p <0.05
−Log10 (adjusted p-value)
Log Fold-change (C2-iLumStem vs Lum)
F
Nkx3.1 (luminal marker)
Ppp1r1b (luminal proximal marker)
Hoxb13 (luminal marker)
Trp63 (basal marker)
Expression
E
P358fs+/-; Pb-Cre- (control)
P358fs+/-; Pb-Cre+ (case)
TROP2
V5 (FOXA1)
Anterior
Dorsal
Lateral
Ventral
100µm
Control
Case
Percent AR+:Trop2+ cells
p=0.013
p=0.012
p=0.002
p=0.005
Prostate lobes - 55 weeks
G
In situ Trop2 validation (Gene = Tacstd2)
V5 (FOXA1)
TROP2
AR
V5 + TROP2
Pb-Cre- (control)
Pb-Cre+ (case)
FOXA1 P358fs+/- (Class 2) - 42 weeks
H
Castration-induced gene signature (In-house)
L1 gene signature (Kirk et al., 2024)
Score
I
Wild-type prostate
Class2-mutant prostate
p<2.2e-16
Luminal Nkx3.1 expression
Intact
Castration
Lum
C2-iLumStem
J
Class 2-mutant prostate (intact)
AR Sig (UP)
Castration score
Consensus stemness
Expression score
p=0.082
p<2.2e-16
Control
Case
Wild-type prostate
Intact
Castrated
K
Cre-
Cre+
V5 (FOXA1)
TROP2
H3
FOXA1 P358fs+/-
Prostate organoids
L
Cre- (wild-type)
Cre+ (Class 2 mutant)
TROP2
Prostate organoids
100µm
FOXA1 P358fs+/- (Class 2)
M
Cre-
Cre+
Rep #1
Rep #2
100 cells
FOXA1 P358fs+/- (Class 2)
Prostate organoids
Percent organoids formed
p <0.0001
p =0.0001
p =0.0065
p =0.0024
Total cells seeded/well
N
Prostate organoid genotypes:
Wild-type (n=20)
Class 1 (R261G+/+; n=8)
Class 1 (R261G+/-; n=12)
Class 2 (P358fs+/-; n=9)
6/9 mice had palpable mass
Percent allograft formation
Days post injection

Fiure 3-3: FOXA1 Class 2 mutants trigger castration-associated intra-luminal plasticity. A) ChIP-seq read-density heatmaps of FOXA1 wild-type and Class 2 mutant at 86,000 sites in Class 2 *P358fs+/-* Cre- and Cre+ organoids. B) Left: UMAP plots of young and old Class 2 *P358fs+/-* mouse prostate tissues. Two distinct case-specific populations are highlighted in yellow and red. Right: Split UMAP plots with only case or control tissue-derived cells colored across clusters. C) Violin plots of noted signatures in the distinct clusters emanating from the Class 2 case and control prostate tissues. D) Left: UMAP plots of the consensus stemness signature (see Methods) in the Class 2 case and control clusters. Right: Volcano plot of up and down-regulated genes from the case vs control Class 2 single-cell RNA-seq data. Consensus stemness genes are highlighted as purple dots. E) Left: Representative multiplex-IF stained sections of V5 and TROP2 in distinct lobes of 55-week-old Class 2 mutant tissues (*Pb-Cre-* represents control tissues, and *Pb-Cre+* represents case tissues). Scale =100µm. Right: Percentage of AR+/TROP2+ cells in distinct lobes of control and case Class 2 mutant tissues (two-sided t-test). F) Violin plots showing mRNA expression (imputed) of noted genes in the distinct Class 2 case and control-specific clusters. G) Representative multiplex-IF stained sections of V5, TROP2, and AR in 42-week-old Class 2 mutant case and control tissues. Scale=100µm. H) Violin plots of the castration-induced gene signature (defined in-house, see methods) and an L1 gene signature (defined by Kirk et al., 2024) in the Class 2 case and control clusters. I) Boxplots of *Nkx3.1* mRNA expression in intact and castrated wild-type prostate tissues (left) and intact FOXA1 Class 2 control and case tissues (right) (two-sided Wilcoxon test). J) Boxplots of distinct gene signature scores in intact FOXA1 Class 2 control and case tissues (top) and castrated or intact wild-type prostate tissues (bottom) (pseudo-bulk analyses from single-cell data, Wilcoxon test). K) Immunoblot of V5 and TROP2 in isogenic Class 2 Cre- (control) and Cre+ (Class 2 mutant expressing) mouse prostate organoids. Total H3 is used as a loading control. L) TROP2 immunofluorescence in Cre- and Cre+ Class 2 mutant mouse prostate organoids. Scale = 100µm. M) Left: Representative images from the limited dilution assay (100 cells/well) in Class 2 Cre- and Cre+ organoids. Right: Percentage of organoids formed in the limited dilution assay at varying cell numbers (n= 3 biological replicates, two-sided t-test). N) Reverse Kaplan-Meier plot of subcutaneous organoid grafting of wild-type, Class 1 or 2 monogenic mouse prostate organoids in *CB17/SCID* mice.

FOXA1 Class 2 mutations trigger castration-associated intra-luminal plasticity

We engineered a knock-in mouse model carrying the FOXA1 P358fs mutant transgene in the *Rosa26* gene locus and, using a similar mating strategy with the Pb-Cre4 mice, conditionally overexpressed the FOXA1 Class 2 mutant in the prostate epithelia (**Supplementary Figure. A3-7D**). While the transgenic protein was robustly expressed, in contrast to Class 1 mutants, overexpression of FOXA1 P358fs did not trigger the formation of precancerous PINs or hyperproliferative IDCs, even in animals as old as 145 weeks (**Supplementary Figure. A3-7E-H**). However, given the neomorphic characteristics and dominant DNA binding of Class 2 mutants (*5*), we investigated their transcriptional effects. Using isogenic prostate organoids, we first found that the transgenic FOXA1 P358fs mutant exhibited cistromic dominance by entirely displacing the endogenous wild-type FOXA1 protein from over 65% of its binding sites (i.e., 44,500 DNA elements; **Figure. 3-3A** and **Supplementary Figure. A3-7I-K**). Notably, even at the residual sites, wild-type FOXA1 binding was significantly attenuated in Class 2-mutant-expressing prostate

organoids (**Supplementary Figure. A3-7L,M**). Next, we profiled wild-type (control) and Class 2-expressing (case) prostate tissues using the single-cell multi-omics (gene expression + ATAC) assay. Louvain clustering of epithelial cells identified seven distinct clusters, with clusters 3 and 4 predominantly comprising cells coming from the *FOXA1 P358fs+/-;Pb-Cre+* (case) prostate tissues (**Supplementary Figure. A3-8A, Figure. 3-3B**). Notably, these cellular clusters constituted over 35-40% of the Class 2-mutant-expressing epithelia and strongly expressed gene markers of secretory luminal (Lum) cells defined in mouse cellular atlases (**Figure. 3-3C**). Notably, while remaining negative for basal genes, Class 2-enriched cells in clusters 3 and 4 gained expression of genes that are exclusively expressed in the luminal proximal (LumP) cells (**Figure. 3-3C**), which is a rare subpopulation of luminal cells located near the urethra (*47*). LumP cells have been shown to possess progenitor/stem-like properties and demonstrate enhanced regenerative capacity in prostate reconstitution assays (*47*). Based on this expression profile, we annotated cells in clusters 3 and 4 as FOXA1 Class 2-induced luminal stem-like cells (abbreviated as C2-iLum^{Stem}).

To further corroborate the hybrid secretory–proximal luminal character of C2-iLum^{Stem} cells, we defined a consensus stemness signature by intersecting independent gene markers of urethra-proximal, luminal progenitors (i.e., LumP, LumC, and L2) reported in three mouse prostate cell atlases (*47–49*) (**Supplementary Figure. A3-8B**). Comparing pure, pseudo-bulk transcriptomic profiles between C2-iLum^{Stem} and Lum populations from case and control tissues, revealed a striking up-regulation of most of the consensus signature genes (**Figure. 3-3D**), including well-characterized markers like *Tacstd2*, *Krt4*, and *Psca* (**Supplementary Figure. A3-8C**). Strikingly, immunofluorescent staining of TROP2 (the protein product of *Tacstd2*) in Class 2-mutant-expressing prostate tissues corroborated a marked, 15-20-fold expansion of TROP2+ epithelial cells, which also stained for V5 (i.e., FOXA1 P358fs) expression (**Figure. 3-3E and Supplementary Figure. A3-8D**). TROP2+/V5+ epithelial cells were uniformly distributed within distal and proximal regions in all prostatic lobes (**Figure. 3-3E**) and showed evident expansion even in prostatic glands harvested from 6-week-old mice (**Supplementary Figure. A3-8E**). Given their uniform spatial distribution, we next questioned if C2-iLum^{Stem} cells originated from the differentiated luminal compartment. Consistent with this notion, unlike the proximally located LumP cells, C2-iLum^{Stem} cells strongly expressed differentiated luminal markers like *Nkx3.1* and

Hoxb13 (**Figure. 3-3F**), and strongly co-stained for AR, TROP2, and FOXA1-V5 proteins (**Figure. 3-3G**). C2-iLum^{Stem} cells were also negative for LumP and basal cell markers *Ppp1r1b* and *Trp63*, respectively (**Figure. 3-3F**). Furthermore, we detected a significant increase in luminal progenitor scores within differentiated luminal cell types identified in distinct mouse prostate atlases in Class 2-mutant-expressing prostate tissues relative to cells from wild-type controls (**Supplementary Figure. A3-8F,G**). Altogether, we found overexpression of the DNA-binding dominant FOXA1 P358fs Class 2 mutant to reprogram transcriptomes of differentiated luminal epithelial cells to activate progenitor/stemness gene programs—a phenomenon we herein refer to as induced intra-luminal plasticity.

Previous studies have documented extensive transcriptional reprogramming in mouse prostate epithelial cells in response to androgen deprivation (*50–52*). More recently, seminal studies mapping the mouse cellular prostate atlas identified distinct sub-populations of luminal epithelial cells that persist after castration, all of which commonly express stemness genes, including TROP2 (*48*, *49*, *53*). Consistent with these reports, we found a marked induction of TROP2 expression in the wild-type prostate epithelium after castration (**Supplementary Figure. A3-9A**). Through single-cell transcriptional profiling of residual prostate tissues two weeks post-castration, we identified an adaptive gene program activated in the AR+ epithelial cells. Notably, these castration-induced genes were also activated in the C2-iLum^{Stem} cell population of *FOXA1 P358fs+/-;Pb-Cre+* animals even without androgen withdrawal (**Figure. 3-3H, top; Supplementary Figure. A3-9B**). A recent study also defined a distinct subpopulation of luminal cells, called L1, that emerge 7 days post-castration and repopulate the epithelium upon androgen restoration (*53*). Strikingly, we found C2-iLum^{Stem} cells to highly express the L1 gene signature without androgen withdrawal (**Figure. 3-3H, bottom**). Consistently, we detected a sustained expression of *Nkx3.1*—an AR up-regulated gene—in the C2-iLum^{Stem} cells, which was undetectable in the AR+ epithelial cells after castration (**Figure. 3-3I**). This was further corroborated by analyzing gene signature scores in pseudo-bulk transcriptional profiles, which revealed that overexpression of FOXA1 P358fs mutant in hormone-replete prostate tissues maintained comparable AR chromatin binding, co-factor enrichment, and transcriptional activity, while triggering a significant activation of genes associated with luminal stemness and adaption to castration (**Figure. 3-3J, top, Supplementary Figure. A3-9C,D**). Conversely, castration induced

the expression of analogous stemness genes in the prostate epithelium while simultaneously suppressing the AR pathway (**Figure. 3-3J, bottom**). Our findings indicate that FOXA1 Class 2 mutants maintain their chromatin pioneering functions related to AR signaling while acquiring a neomorphic capacity to activate stemness and castration-associated genes in the hormone-replete prostate epithelium.

Cre-mediated overexpression of the FOXA1 P358fs mutant in prostate organoids also induced the expression of TROP2, which was evident even through *in situ* staining of TROP2 in intact organoids (**Figure. 3-3K,L, Supplementary Figure. A3-9E**). This was also accompanied by a concomitant enrichment of the gene signature of luminal progenitor cells in Class 2-mutant organoids (**Supplementary Figure. A3-9F**). To phenotypically characterize their stemness character, we performed limited dilution assays. Here, Class 2-expressing prostate cells showed a 2-3-fold higher organoid forming capacity relative to isogenic wild-type (Cre-) cells (**Figure. 3-3M, Supplementary Figure. A3-9G,H**). More notably, prostate organoids expressing the *FOXA1 P358fs* Class 2-mutant were the only single-gene model to form palpable allografts when transplanted into mouse dorsal flanks, in contrast to both wild-type as well as Class 1 mutant-expressing organoids (**Figure. 3-3N, Supplementary Figure. A3-9I**). In summary, while incapable of driving prostatic epithelial transformation, we found FOXA1 Class 2 mutants to induce significant transcriptional reprogramming that leads to activation of luminal progenitor as well as castration-adaptive gene programs in hormone-intact mice. Notably, this is consistent with their clonal detection in CRPC tumors and suggests that Class 2 mutants “prime” the prostate epithelia to better withstand androgen deprivation therapies.

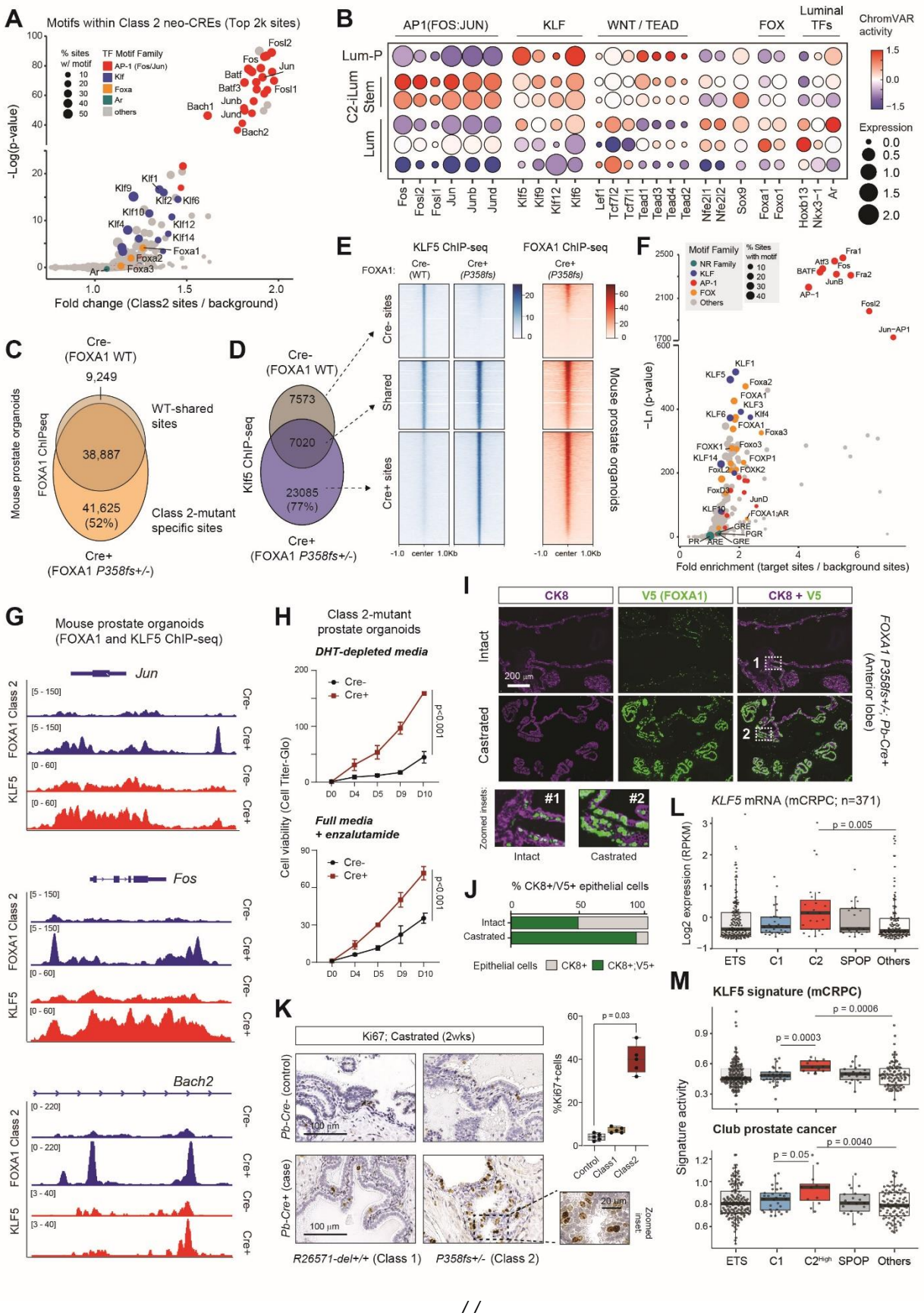

A
Motifs within Class 2 neo-CREs (Top 2k sites)
-Log(p-value)
Fold change (Class2 sites / background)
% sites w/ motif
TF Motif Family
AP-1 (Fos/Jun)
Klf
Foxa
Ar
others
B
AP1(FOS:JUN)
KLF
WNT / TEAD
FOX
Luminal TFs
Lum-P
C2-iLum
Stem
Lum
ChromVAR activity
Expression
C
Cre- (FOXA1 WT)
9,249
WT-shared sites
38,887
41,625 (52%)
Class 2-mutant specific sites
Cre+ (FOXA1 P358fs+/-)
Mouse prostate organoids
FOXA1 ChIPseq
D
Cre- (FOXA1 WT)
7573
7020
23085 (77%)
Cre+ (FOXA1 P358fs+/-)
Klf5 ChIPseq
E
KLF5 ChIP-seq
FOXA1 ChIP-seq
FOXA1:
Cre- (WT)
Cre+ (P358fs)
Cre- sites
Shared
Cre+ sites
Mouse prostate organoids
center 1.0Kb
F
Motif Family
NR Family
KLF
AP-1
FOX
Others
% Sites with motif
−Ln (p-value)
Fold enrichment (target sites / background sites)
G
Mouse prostate organoids (FOXA1 and KLF5 ChIP-seq)
Jun
Fos
Bach2
FOXA1 Class 2
KLF5
H
Class 2-mutant prostate organoids
DHT-depleted media
Full media + enzalutamide
Cell viability (Cell Titer-Glo)
p<0.001
I
CK8
V5 (FOXA1)
CK8 + V5
Intact
Castrated
FOXA1 P358fs+/-; Pb-Cre+ (Anterior lobe)
Zoomed insets:
J
% CK8+/V5+ epithelial cells
Epithelial cells
CK8+
CK8+;V5+
K
Ki67; Castrated (2wks)
Pb-Cre- (control)
Pb-Cre+ (case)
%Ki67+cells
p = 0.03
R26571-del+/+ (Class 1)
P358fs+/- (Class 2)
L
KLF5 mRNA (mCRPC; n=371)
Log2 expression (RPKM)
p = 0.005
ETS
C1
C2
SPOP
Others
M
KLF5 signature (mCRPC)
p = 0.0006
p = 0.0003
Club prostate cancer
p = 0.0040
p = 0.05
Signature activity
C2High

Figure 3-4: FOXA1 Class 2 mutants activate KLF5/AP-1 neo-enhancer circuitry to drive therapy-resistance. A) Fold change and significance of HOMER motifs enriched within Class 2 neo *cis*-regulatory elements (CREs) from single-cell multiome ATAC data in Class 2 control and case prostate tissues (HOMER, hypergeometric test). B) Bubble plots highlighting expression and chromVAR motif activity of noted transcription factors from 10X single-cell multi-omics of Class 2 mutant control and case tissues. The size of the dot indicates expression and heatmap color intensity indicates ChromVar activity. C) Venn diagram showing overlap of FOXA1 wild-type and Class 2 mutant cistromes in *P358fs*+/- Cre- and Cre+ mouse prostate organoids. D) Venn diagram showing overlap of the KLF5 cistrome in Class 2 Cre- and Cre+ mouse prostate organoids. E) ChIP-seq read-density heatmaps of KLF5 and FOXA1 Class 2 mutant at the Class 2-specific, wild-type, or shared sites in the FOXA1 *P358fs*+/- Cre- and Cre+ mouse prostate organoids. F) Fold change and significance of HOMER motifs enriched at KLF5 and Class 2 mutant co-bound sites (HOMER, hypergeometric test). G) ChIP-seq read-density tracks of the FOXA1 Class 2 mutant and KLF5 within the *Jun, Fos,* and *Bach2* locus in the Class 2 Cre- and Cre+ mouse prostate organoids. H) Growth curves (Cell Titer-Glow) of Cre- and Cre+ Class 2 mutant organoids in DHT-depleted and enzalutamide-treated media conditions (n=4 biological replicates, two-sided t-test). I) Representative multiplex-IF images of CK8 and V5 in the intact and castrated Class 2 mutant (*Pb-Cre*+) tissues. Scale =100µm. Zoomed insets of select regions are shown. J) Percentage of epithelial cells (CK8+, grey) that are V5+ (green) in intact and castrated Class 2 tissues from panel (I). K) Left: Representative Ki67 IHC stained sections of Class 1 (*R265-71del*+/+) and Class 2 (*P358fs*+/-) control and case castrated tissues. Right: Percentage of Ki67+ cells in control, Class 1, and Class 2 tissues (two-sided t-test). Box plot: center line, median; box, interquartile range (Q1–Q3); whiskers, minimum to maximum values; all individual data points shown. L) Boxplots showing *KLF5* mRNA expression in metastatic CRPC patient tissues (SU2C, n=371 patients) across distinct genomic driver groups. SU2C, Stand Up To Cancer (Wilcoxon rank-sum test). M) Boxplots showing the KLF5 gene signature (CRPC, top) and Club prostate cancer signature (bottom) in CRPC patient tissues (SU2C, n=371 patients) across distinct genomic driver groups (Wilcoxon rank-sum test).

Class 2 mutants activate KLF5/AP-1 neo-enhancers to drive castration resistance

To interrogate the molecular mechanism of FOXA1 Class 2-mutant induced stemness, we compared chromatin accessibility profiles between C2-iLum^{Stem} and differentiated luminal cells using single cell ATACseq data. Here, we detected a significant increase in de-compacted, open chromatin in the presence of the Class 2 mutant, with over 8,800 *cis*-regulatory sites located within distal non-coding regions gaining accessibility (**Supplementary Figure. A3-10A**). Intriguingly, motif analyses of the Class 2-pioneered *cis*-regulatory sites revealed a significant enrichment of motifs belonging to the AP-1 (FOS/JUN), KLF, and FOXA transcription factor families (**Figure. 3-4A**), with a surprising absence of AR and other nuclear receptor motifs. To further elucidate differential activation of transcriptional networks across distinct epithelial cell populations detected within Class 2-mutant prostate tissues, we performed ChromVAR analysis on the corresponding chromatin accessibility (ATAC) profiles from our single-cell multi-omics data. This revealed the C2-iLum^{Stem} cells to show specific activation of AP-1 transcriptional complexes (**Figure. 3-4B**). KLF factors instead were predicted to have higher activity in both the innate LumP

and C2-iLum^{Stem} clusters (**Figure. 3-4B**). Among these, KLF5 was particularly strongly expressed and significantly up-regulated in the C2-iLum^{Stem} cells relative to the differentiated luminal epithelium (**Supplementary Figure. A3-10B**). We also detected a significant activation of WNT (consistent with our earlier findings (*5*)) and TEAD transcriptional regulators (recently described to interact with AP-1 in a subset of CRPC tumors (*54*)) in the C2-iLum^{Stem} epithelial cells (**Figure. 3-4B**).

Next, we employed our isogenic organoid models to experimentally interrogate the crosstalk between FOXA1 P358fs mutant, AP-1, and KLF5 transcriptional complexes. As a neomorphic gain-in-function, while the truncated FOXA1 Class 2 mutant replaced the wild-type protein on the chromatin, it additionally acquired *de novo* binding at over 40,000 new *cis*-regulatory sites within distal non-coding regions (**Figure. 3-4C, Supplementary Figure. A3-10C,D**). As anticipated, HOMER motif analysis uncovered AR and FOX recognition elements to be most significantly enriched within the common cistrome (**Supplementary Figure. A3-10E**). In contrast, Class 2-mutant-specific binding sites lacked the AR motif and instead showed enrichment of motifs recognized by the AP-1 (FOS/JUN) family of transcription factors (**Supplementary Figure. A3-10E**). As seen in mouse prostate tissues, overexpression of the FOXA1 P358fs mutant in prostate organoids also led to an up-regulation of KLF5 (**Supplementary Figure. A3-10F**), with the P358fs mutant occupying *cis*-regulatory sites within the Klf5 gene locus (**Supplementary Figure. A3-10G**). We also found the endogenous FOXA1 P358fs mutant in LAPC4—a human-derived CRPC cell line—to bind within the *Klf5* gene locus (**Supplementary Figure. A3-10G**).

KLF5 was recently implicated in driving castration resistance in human PCa(*55*). Thus, we next profiled KLF5 chromatin occupancy (ChIP-seq) in our Class 2-mutant organoids. We detected a striking expansion of the KLF5 cistrome, gaining over 23,000 new binding sites in Class 2-mutant-expressing organoids (**Figure. 3-4D**), which were notably co-occupied by the FOXA1 P358fs mutant (**Figure. 3-4E**). HOMER motif analysis of the FOXA1 Class 2 and KLF5 shared enhancers uncovered a significant enrichment of their own as well as the AP-1 family transcription factor motifs (**Figure. 3-4F**). We also found FOXA1 P358fs to bind within gene loci of several AP-1 complex transcription factors, including *Jun*, *Fos*, and *Maff*, which was accompanied by a concurrent gain in KLF5 binding at these sites and up-regulation of their transcripts (**Figure. 3-4G, Supplementary Figure. A3-10H,I**). In summary, our findings indicate that FOXA1 Class 2

mutants *de novo* pioneer a novel enhancer circuitry that is bound and activated by KLF5 and AP-1 complexes, thereby driving transcriptional reprogramming and luminal plasticity in prostate epithelial cells.

In phenotypic assays, we found Class 2-mutant prostate organoids to grow significantly better in androgen-depleted media as well as upon treatment with the AR antagonist enzalutamide (**Figure. 3-4H**), suggesting that Class 2 mutant-expressing prostate epithelial cells can better withstand inhibition of AR signaling. To assess this *in vivo*, we castrated the *FOXA1 P358fs+/-;Pb-Cre+* mice and quantified the survival of transgene-expressing luminal epithelial cells. Here, we focused on the anterior prostate lobes leveraging the well-described incomplete Cre recombinase activity that leads to transgene expression in roughly half of the prostate epithelium (*38*). Accordingly, in intact animals, we detected about 50% of the CK8+ luminal epithelial cells to co-express the V5-tagged FOXA1 P358fs transgenic protein (**Figure. 3-4I,J**). The CK8+;V5+ cells, however, comprised almost 95% of the prostate epithelium two weeks post-castration (**Figure. 3-4I,J**), suggesting that the FOXA1 P358fs-expressing cells are able to survive and persist upon androgen removal. Consistently, Ki67 staining of the residual prostate tissues revealed that the Class 2-mutant-expressing epithelia had significantly higher proliferative activity in the absence of androgen when compared to wild-type (Pb-Cre-) controls as well as Class 1-mutant-expressing prostate tissues (**Figure. 3-4K**). This was also reflected in significantly higher weights of residual prostate glands harvested from the case versus control animals (**Supplementary Figure. A3-11A**).

When credentialing these findings in human patient tumors, we found the highest expression of KLF5 in CRPC specimens that harbored clonal FOXA1 Class 2 mutation (**Figure. 3-4L**). In Class 2-mutant tumors, KLF5 expression was also positively correlated to two independent gene signatures of KLF5 (**Supplementary Figure. A3-11B;** published in (*55*)). Assessment of KLF5 activity based on both signatures also uncovered strongest activation in Class 2-mutant CRPC tumors versus other genomic subtypes (**Figure. 3-4M, Supplementary Figure. A3-11C**). Strikingly, in Class 2-mutant tumors, we also detected a strong activation of gene signature of PCa-associated club cells (**Figure. 3-4M**), which were identified via single-cell methodologies in metastatic PCa specimens (*56*) and are associated with treatment resistance to AR-targeted therapies (*57*). Furthermore, we observed a strong positive correlation between *KLF5* transcripts

and the PCa club signature in CRPC, which was not the case in FOXA1 Class 1 mutant tumors (**Supplementary Figure. A3-11D**). Altogether, our findings suggest that FOXA1 Class 2 mutants reshape the *cis*-regulatory chromatin landscape to activate KLF5 and AP-1 enhancer-driven gene programs that confer resistance to androgen deprivation therapies, positioning the acquired FOXA1 Class 2 alterations as promoters of metastatic and therapy-resistant disease progression.

Discussion

Prostate cancer is predominantly driven by transcription factor oncogenes (*2–6*). FOXA1 is a pioneer transcription factor that is frequently altered in PCa. In the Chinese population, surpassing the frequency of ETS fusions, FOXA1 mutations become the predominant driver alteration detected in over 40% of primary prostatic tumors (*8*). Yet, FOXA1 pathobiology remains unexplored in PCa mouse models. In this study, we developed novel knock-in transgenic mouse models to elucidate the functional consequences of FOXA1 Class 1 and Class 2 mutations in PCa initiation and progression. Through longitudinal histopathological analysis and multi-omic characterization, we show that FOXA1 Class 1 mutations serve as initiating events in prostate tumorigenesis. These mutations, in conjunction with p53 inactivation, give rise to AR-dependent adenocarcinomas that significantly regress upon androgen deprivation. In contrast, FOXA1 Class 2 mutations emerge as key drivers of castration resistance by specifying a luminal progenitor cell state, thus enhancing survival in androgen-depleted environments. Our findings provide experimental evidence causally implicating FOXA1 alterations in AR-dependent prostate carcinogenesis and progression, while also explaining the divergent evolution of distinct FOXA1 mutation classes observed in human PCa.

Since the clinical demonstration of PCa regression following androgen blockade, inhibition of the androgen/AR axis has remained the cornerstone of treatment for locally advanced and metastatic PCa (*14, 58*). However, replicating this acute androgen dependence of primary PCa has been particularly challenging in mouse models. The widely-used PTEN inactivation-driven PCa mouse models typically result in prostatic tumors exhibiting castration-resistant phenotypes (*27*), thus representing a more advanced disease. Distinctively, the FOXA1 Class 1-driven prostate adenocarcinomas faithfully recapitulate key features of the primary human PCa. These tumors exhibited key characteristics including NKX3.1 loss, MYC up-regulation, and sustained AR

expression. Consistent with prior findings from our group (*12*), we also detected an abnormal induction of NSD2 in Class 1 tumors along with extensive chromatin reprogramming of AR to activate its oncogenic gene programs. Most importantly, akin to primary human PCa, these tumors showed sensitivity to androgen deprivation and regressed upon castration. While further characterization is necessary, FOXA1 Class 1-driven murine PCa and derivative tumor organoids offer valuable systems for investigating the pathobiology that underlies PCa initiation and development of castration resistance.

Mechanistically, we found Class 1 mutants to simultaneously activate the AR and PI3K/mTORC1 oncogenic pathways. Similar transcriptional changes were observed in the recently developed SPOP-driven PCa murine model (*32*). Also, in localized human PCa, cases harboring either the SPOP or FOXA1 Class 1 mutations were reported to have elevated AR activity (*5, 32*). These insights are particularly significant given the well-established antagonistic relationship between AR and PI3K pathways in PCa (*59*), and lower prevalence of PI3K pathway alterations in the organ-confined, localized disease (*2, 3, 60*). In this context, and based on our findings, we posit that FOXA1 Class 1 mutations amplify AR activity to counteract the inherent antagonism of AR signaling resulting from non-canonical activation of oncogenic mTORC1 signaling in these tumors. Furthermore, our findings offer a plausible explanation for lower frequency of PTEN alterations even upon metastatic progression of primary tumors initiated by FOXA1 Class 1 mutations in human patients.

In contrast to Class 1 mutations, C-terminal truncating FOXA1 Class 2 mutations failed to initiate PCa formation in our mouse models. This finding is consistent with clonal emergence of Class 2 mutations only in CRPC human tumors (*5*). FOXA1 Class 2 mutants instead triggered profound reprogramming of the chromatin regulatory landscape (i.e., the regulome) by commissioning a KLF5/AP-1 neo-cistrome in prostate luminal epithelial cells that specifies a TROP2+/CK4+ progenitor cell state, activating castration-induced gene programs even without exposure to androgen deprivation therapies. This confirms the C-terminal truncated Class 2 mutants to function as neomorphs and serve as promoters of PCa progression. Several studies have reported that TROP2 is upregulated in approximately 90% of CRPC tumors (*61–63*). It would be clinically valuable to determine whether Class 2-mutant CRPCs exhibit higher or more uniform expression of TROP2, making them sensitive to treatment with TROP2-targeted antibody-drug conjugates,

such as sacituzumab govitecan that was FDA approved for breast cancer treatment (*64*). Earlier, we had shown FOXA1 Class 2 mutants promote metastasis in human models by de-repressing the WNT pathway (*5*). Thus, we propose a coalescent evolutionary model wherein FOXA1 Class 2 mutations emerge as primary prostatic tumors subsequently evolve towards metastatic (via WNT de-repression) or castration-resistant (via inducing intra-luminal plasticity) phenotypes. However, further investigation is necessary to determine if FOXA1 Class 2 mutations would arise spontaneously during natural disease progression (i.e., in *de novo* metastatic tumors) or are induced by the selective pressure of androgen deprivation therapies.

Intriguingly, motif analyses of the Class 2 neo-cistrome uncovered AP-1 transcription factors as key collaborators in mediating the castration-resistant phenotype. This is consistent with emerging literature implicating the AP-1 complex in driving therapy-emergent PCa subtypes (*54*), including AR-negative neuroendocrine disease (*65*). We also observed extensive remodeling of the *cis*-regulome in FOXA1 Class 2-mutant prostate epithelial cells, wherein the acquired neo-enhancers did not contain motifs of canonical FOXA1 partnering TFs, like AR and NFI. These neo-sites were instead enriched for the AP-1 family motifs. A comprehensive mouse prostate cellular atlas by Tang and colleagues recently identified a novel castration-induced luminal cell population, designated L1, which persists following castration and repopulates the epithelial lining upon androgen restoration (*53*). This study suggested this luminal plasticity to be regulated by the loss of AR and gain of AP-1 *cis*-regulatory elements. Here, while maintaining AR activity, we found FOXA1 Class 2 mutants to parallelly co-opt AP-1 and KLF5 cis-regulatory enhancers in inducing luminal plasticity in androgen-replete mouse prostate epithelium. Beyond KLF5, our analyses also revealed increased activity of TEAD family factors in the Class 2 mutant-pioneered cis-regulome. Recent studies using human patient specimens have reported TEAD proteins to collaborate with AP-1 in establishing a distinct chromatin profile in stem cell-like CRPC tumors (designated as CRPC-SCL) that show shorter remission upon treatment with AR signaling inhibitors (*54*). These insights collectively position AP-1 transcription factors as key drivers of lineage plasticity—both intra-luminal as well as neuroendocrine—that enables PCa progression and resistance to androgen/AR-targeted therapies.

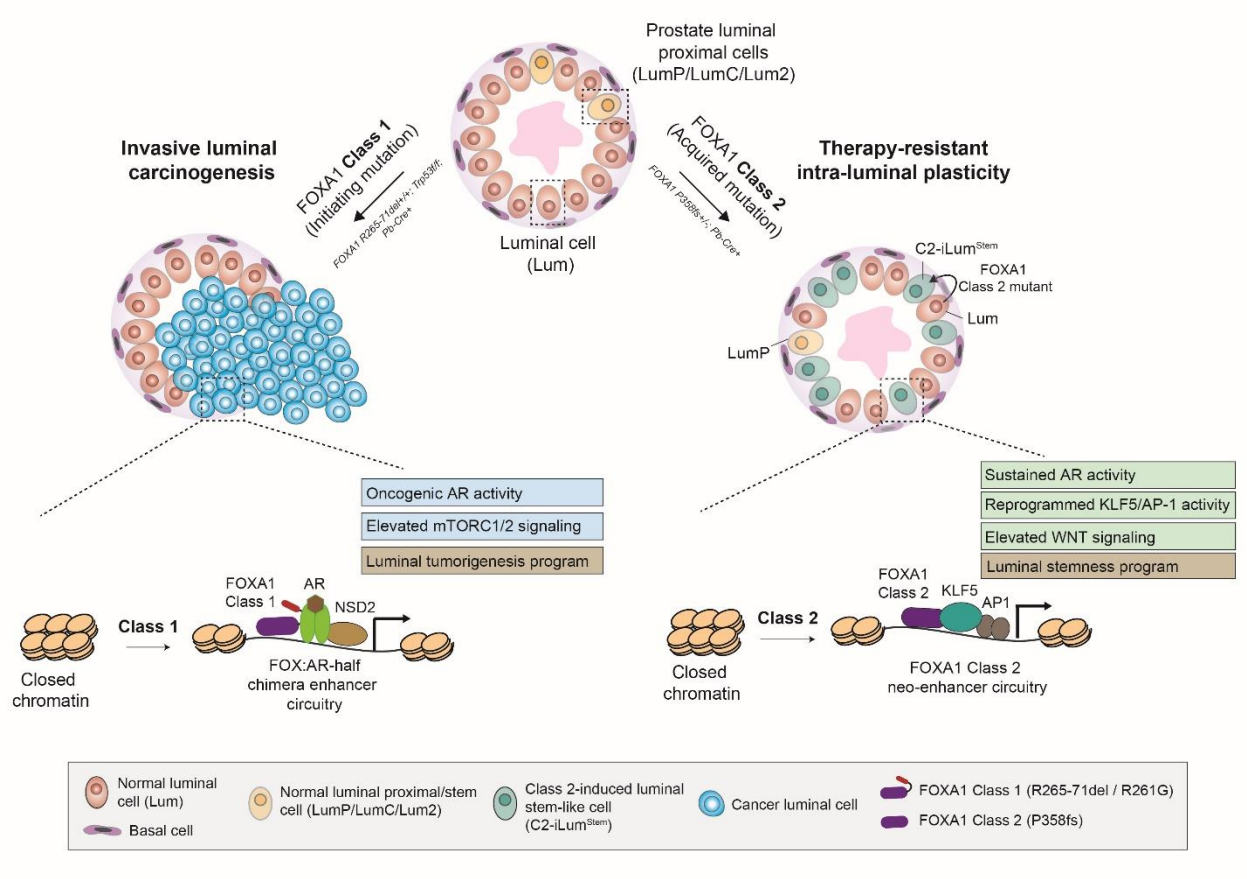


Figure 3-5: Schema depicting distinct oncogenic mechanisms of FOXA1 Class 1 and Class 2 mutations in driving prostate tumorigenesis or therapy-resistant progression. Normal prostate epithelia comprise three main cell types: terminally differentiated, secretory luminal (Lum) cells, basal cells, and luminal proximal stem-like cells (annotated as LumP, LumC, or Lum2 in published cellular atlases). Overexpression of FOXA1 Class 1 mutants in a p53-deficient prostate epithelium (left) drives formation of androgen-sensitive, invasive prostate adenocarcinoma. This malignant transformation is enabled by the aberrant induction of NSD2, cistromic redistribution of the androgen receptor (AR) at chimeric FOXA1:AR-half enhancer elements, and co-activation of the mTORC1/2 oncogenic pathway. In contrast, overexpression of FOXA1 Class 2 mutants (right) reprograms AR+/CK8+ differentiated luminal epithelial cells to express luminal stemness markers, thereby partially acquiring progenitor-like characteristics. Class 2 mutants induce this intra-luminal plasticity by commissioning a *neo*-cistrome that is occupied and activated by the AP-1 and KLF5 transcriptional complexes. Class 2-induced luminal stem-like cells (annotated C2-iLum^{Stem}) retain AR activity and resist cell death upon castration, suggesting their role in enabling PCa progression following androgen deprivation therapy.

In summary, our study provides the first experimental evidence that distinct classes of FOXA1 alterations drive disparate PCa phenotypes (**Figure. 3-5**), while explaining their divergent evolution in human disease. The truncal FOXA1 Class 1 mutations initiate AR-dependent and androgen-sensitive prostate adenocarcinoma by co-activation of oncogenic AR and mTORC1 pathways, while acquired FOXA1 Class 2 mutations reprogram epithelial cells into a castration-

resistant, luminal progenitor state by activating AP-1/KLF5 neo-enhancer-driven gene programs. Further validating the original classification schema of FOXA1 alterations found in human cancers (*5*, *6*), these findings underscore the functional versatility of FOXA1 as a principal PCa oncogene. Depending on the mutation type, FOXA1 either drives PCa formation or metastatic, therapy-resistant progression. Additionally, the FOXA1 transgenic mouse models developed in this study provide a valuable bioresource for studying the pathobiology of breast, salivary gland, and bladder tumors harboring oncogenic FOXA1 alterations.

References

1. C. S. Grasso, Y.-M. Wu, D. R. Robinson, X. Cao, S. M. Dhanasekaran, A. P. Khan, M. J. Quist, X. Jing, R. J. Lonigro, J. C. Brenner, I. A. Asangani, B. Ateeq, S. Y. Chun, J. Siddiqui, L. Sam, M. Anstett, R. Mehra, J. R. Prensner, N. Palanisamy, G. A. Ryslik, F. Vandin, B. J. Raphael, L. P. Kunju, D. R. Rhodes, K. J. Pienta, A. M. Chinnaiyan, S. A. Tomlins, The mutational landscape of lethal castration-resistant prostate cancer. *Nature* 487, 239–243 (2012).

2. Cancer Genome Atlas Research Network, The Molecular Taxonomy of Primary Prostate Cancer. *Cell* 163, 1011–1025 (2015).

3. D. Robinson, E. M. Van Allen, Y.-M. Wu, N. Schultz, R. J. Lonigro, J.-M. Mosquera, B. Montgomery, M.-E. Taplin, C. C. Pritchard, G. Attard, H. Beltran, W. Abida, R. K. Bradley, J. Vinson, X. Cao, P. Vats, L. P. Kunju, M. Hussain, F. Y. Feng, S. A. Tomlins, K. A. Cooney, D. C. Smith, C. Brennan, J. Siddiqui, R. Mehra, Y. Chen, D. E. Rathkopf, M. J. Morris, S. B. Solomon, J. C. Durack, V. E. Reuter, A. Gopalan, J. Gao, M. Loda, R. T. Lis, M. Bowden, S. P. Balk, G. Gaviola, C. Sougnez, M. Gupta, E. Y. Yu, E. A. Mostaghel, H. H. Cheng, H. Mulcahy, L. D. True, S. R. Plymate, H. Dvinge, R. Ferraldeschi, P. Flohr, S. Miranda, Z. Zafeiriou, N. Tunariu, J. Mateo, R. Perez-Lopez, F. Demichelis, B. D. Robinson, M. Schiffman, D. M. Nanus, S. T. Tagawa, A. Sigaras, K. W. Eng, O. Elemento, A. Sboner, E. I. Heath, H. I. Scher, K. J. Pienta, P. Kantoff, J. S. de Bono, M. A. Rubin, P. S. Nelson, L. A. Garraway, C. L. Sawyers, A. M. Chinnaiyan, Integrative clinical genomics of advanced prostate cancer. *Cell* 161, 1215–1228 (2015).

4. S. A. Tomlins, D. R. Rhodes, S. Perner, S. M. Dhanasekaran, R. Mehra, X.-W. Sun, S. Varambally, X. Cao, J. Tchinda, R. Kuefer, C. Lee, J. E. Montie, R. B. Shah, K. J. Pienta, M. A. Rubin, A. M. Chinnaiyan, Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* 310, 644–648 (2005).

5. A. Parolia, M. Cieslik, S.-C. Chu, L. Xiao, T. Ouchi, Y. Zhang, X. Wang, P. Vats, X. Cao, S. Pitchiaya, F. Su, R. Wang, F. Y. Feng, Y.-M. Wu, R. J. Lonigro, D. R. Robinson, A. M. Chinnaiyan, Distinct structural classes of activating FOXA1 alterations in advanced prostate cancer. *Nature* 571, 413–418 (2019).

6. E. J. Adams, W. R. Karthaus, E. Hoover, D. Liu, A. Gruet, Z. Zhang, H. Cho, R. DiLoreto, S. Chhangawala, Y. Liu, P. A. Watson, E. Davicioni, A. Sboner, C. E. Barbieri, R. Bose, C. S. Leslie, C. L. Sawyers, FOXA1 mutations alter pioneering activity, differentiation and prostate cancer phenotypes. *Nature* 571, 408–412 (2019).

7. C. E. Barbieri, S. C. Baca, M. S. Lawrence, F. Demichelis, M. Blattner, J.-P. Theurillat, T. A. White, P. Stojanov, E. Van Allen, N. Stransky, E. Nickerson, S.-S. Chae, G. Boysen, D. Auclair, R. C. Onofrio, K. Park, N. Kitabayashi, T. Y. MacDonald, K. Sheikh, T. Vuong, C. Guiducci, K. Cibulskis, A. Sivachenko, S. L. Carter, G. Saksena, D. Voet, W. M. Hussain, A. H. Ramos, W. Winckler, M. C. Redman, K. Ardlie, A. K. Tewari, J. M. Mosquera, N. Rupp, P. J. Wild, H. Moch, C. Morrissey, P. S. Nelson, P. W. Kantoff, S. B. Gabriel, T. R. Golub, M. Meyerson, E. S. Lander, G. Getz, M. A. Rubin, L. A. Garraway, Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nat. Genet.* 44, 685–689 (2012).

8. J. Li, C. Xu, H. J. Lee, S. Ren, X. Zi, Z. Zhang, H. Wang, Y. Yu, C. Yang, X. Gao, J. Hou, L. Wang, B. Yang, Q. Yang, H. Ye, T. Zhou, X. Lu, Y. Wang, M. Qu, Q. Yang, W. Zhang, N. M. Shah, E. C. Pehrsson, S. Wang, Z. Wang, J. Jiang, Y. Zhu, R. Chen, H. Chen, F. Zhu, B. Lian, X. Li, Y. Zhang, C. Wang, Y. Wang, G. Xiao, J. Jiang, Y. Yang, C. Liang, J. Hou, C. Han, M. Chen, N. Jiang, D. Zhang, S. Wu, J. Yang, T. Wang, Y. Chen, J. Cai, W. Yang, J. Xu, S. Wang, X. Gao, T. Wang, Y. Sun, A genomic and epigenomic atlas of prostate cancer in Asian populations. *Nature* 580, 93–99 (2020).

9. C. A. Heinlein, C. Chang, Androgen receptor in prostate cancer. *Endocr. Rev.* 25, 276–308 (2004).

10. E. Mulder, A. A. Vrij, A. O. Brinkmann, DNA and ribonucleotide binding characteristics of two forms of the androgen receptor from rat prostates. *Biochem. Biophys. Res. Commun.* 114, 1147–1153 (1983).

11. M. M. Pomerantz, F. Li, D. Y. Takeda, R. Lenci, A. Chonkar, M. Chabot, P. Cejas, F. Vazquez, J. Cook, R. A. Shivdasani, M. Bowden, R. Lis, W. C. Hahn, P. W. Kantoff, M. Brown, M. Loda, H. W. Long, M. L. Freedman, The androgen receptor cistrome is extensively reprogrammed in human prostate tumorigenesis. *Nat. Genet.* 47, 1346–1351 (2015).

12. A. Parolia, S. Eyunni, B. K. Verma, E. Young, Y. Liu, L. Liu, J. George, S. Aras, C. K. Das, R. Mannan, R. Ur Rasool, E. Mitchell-Velasquez, S. Mahapatra, J. Luo, S. E. Carson, L. Xiao, P. R. Gajjala, S. Venkatesh, M. Jaber, X. Wang, T. He, Y. Qiao, M. Pang, Y. Zhang, J. C.-Y. Tien, M. Louw, M. Alhusayan, X. Cao, F. Su, O. Tavana, C. Hou, Z. Wang, K. Ding, A. M. Chinnaiyan, I. A. Asangani, NSD2 is a requisite subunit of the AR/FOXA1 neo-enhanceosome in promoting prostate tumorigenesis. *Nat. Genet.*, 1–12 (2024).

13. I. G. Mills, Maintaining and reprogramming genomic androgen receptor activity in prostate cancer. *Nat. Rev. Cancer* 14, 187–198 (2014).

14. C. Tran, S. Ouk, N. J. Clegg, Y. Chen, P. A. Watson, V. Arora, J. Wongvipat, P. M. Smith-Jones, D. Yoo, A. Kwon, T. Wasielewska, D. Welsbie, C. D. Chen, C. S. Higano, T. M. Beer, D. T. Hung, H. I. Scher, M. E. Jung, C. L. Sawyers, Development of a second-generation antiandrogen for treatment of advanced prostate cancer. *Science* 324, 787–790 (2009).

15. X. Yuan, C. Cai, S. Chen, S. Chen, Z. Yu, S. P. Balk, Androgen receptor functions in castration-resistant prostate cancer and mechanisms of resistance to new agents targeting the androgen axis. *Oncogene* 33, 2815–2825 (2014).

16. T. Visakorpi, E. Hyytinen, P. Koivisto, M. Tanner, R. Keinänen, C. Palmberg, A. Palotie, T. Tammela, J. Isola, O. P. Kallioniemi, In vivo amplification of the androgen receptor gene and progression of human prostate cancer. *Nat. Genet.* 9, 401–406 (1995).

17. Y. Li, S. C. Chan, L. J. Brand, T. H. Hwang, K. A. T. Silverstein, S. M. Dehm, Androgen receptor splice variants mediate enzalutamide resistance in castration-resistant prostate cancer cell lines. *Cancer Res.* 73, 483–489 (2013).

18. M. Iwafuchi-Doi, G. Donahue, A. Kakumanu, J. A. Watts, S. Mahony, B. F. Pugh, D. Lee, K. H. Kaestner, K. S. Zaret, The Pioneer Transcription Factor FoxA Maintains an Accessible Nucleosome Configuration at Enhancers for Tissue-Specific Gene Activation. *Mol. Cell* 62, 79–91 (2016).

19. J. S. Carroll, X. S. Liu, A. S. Brodsky, W. Li, C. A. Meyer, A. J. Szary, J. Eeckhoute, W. Shao, E. V. Hestermann, T. R. Geistlinger, E. A. Fox, P. A. Silver, M. Brown, Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell* 122, 33–43 (2005).

20. M. Lupien, J. Eeckhoute, C. A. Meyer, Q. Wang, Y. Zhang, W. Li, J. S. Carroll, X. S. Liu, M. Brown, FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell* 132, 958–970 (2008).

21. J. L. L. Robinson, T. E. Hickey, A. Y. Warren, S. L. Vowler, T. Carroll, A. D. Lamb, N. Papoutsoglou, D. E. Neal, W. D. Tilley, J. S. Carroll, Elevated levels of FOXA1 facilitate androgen receptor chromatin binding resulting in a CRPC-like phenotype. *Oncogene* 33, 5666 (2013).

22. M. G. Dalin, A. Desrichard, N. Katabi, V. Makarov, L. A. Walsh, K.-W. Lee, Q. Wang, J. Armenia, L. West, S. Dogan, L. Wang, D. Ramaswami, A. L. Ho, I. Ganly, D. B. Solit, M. F. Berger, N. D. Schultz, J. S. Reis-Filho, T. A. Chan, L. G. T. Morris, Comprehensive molecular characterization of salivary duct carcinoma reveals actionable targets and similarity to apocrine breast cancer. *Clin. Cancer Res.* 22, 4623–4633 (2016).

23. Cancer Genome Atlas Research Network, Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature* 507, 315–322 (2014).

24. X. Fu, R. Pereira, C. De Angelis, J. Veeraraghavan, S. Nanda, L. Qin, M. L. Cataldo, V. Sethunath, S. Mehravaran, C. Gutierrez, G. C. Chamness, Q. Feng, B. W. O'Malley, P. Selenica, B. Weigelt, J. S. Reis-Filho, O. Cohen, N. Wagle, A. Nardone, R. Jeselsohn, M. Brown, M. F. Rimawi, C. K. Osborne, R. Schiff, FOXA1 upregulation promotes enhancer and transcriptional reprogramming in endocrine-resistant breast cancer. *Proc. Natl. Acad. Sci. U. S. A.*, doi: 10.1073/pnas.1911584116 (2019).

25. G. Ciriello, M. L. Gatza, A. H. Beck, M. D. Wilkerson, S. K. Rhie, A. Pastore, H. Zhang, M. McLellan, C. Yau, C. Kandoth, R. Bowlby, H. Shen, S. Hayat, R. Fieldhouse, S. C. Lester, G. M. K. Tse, R. E. Factor, L. C. Collins, K. H. Allison, Y.-Y. Chen, K. Jensen, N. B. Johnson, S. Oesterreich, G. B. Mills, A. D. Cherniack, G. Robertson, C. Benz, C. Sander, P. W. Laird, K. A. Hoadley, T. A. King, TCGA Research Network, C. M. Perou, Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. *Cell* 163, 506–519 (2015).

26. E. Rheinbay, P. Parasuraman, J. Grimsby, G. Tiao, J. M. Engreitz, J. Kim, M. S. Lawrence, A. Taylor-Weiner, S. Rodriguez-Cuevas, M. Rosenberg, J. Hess, C. Stewart, Y. E. Maruvka, P. Stojanov, M. L. Cortes, S. Seepo, C. Cibulskis, A. Tracy, T. J. Pugh, J. Lee, Z. Zheng, L. W. Ellisen, A. J. Iafrate, J. S. Boehm, S. B. Gabriel, M. Meyerson, T. R. Golub, J. Baselga, A. Hidalgo-Miranda, T. Shioda, A. Bernards, E. S. Lander, G. Getz, Recurrent and functional regulatory mutations in breast cancer. *Nature* 547, 55–60 (2017).

27. S. Wang, J. Gao, Q. Lei, N. Rozengurt, C. Pritchard, J. Jiao, G. V. Thomas, G. Li, P. Roy-Burman, P. S. Nelson, X. Liu, H. Wu, Prostate-specific deletion of the murine Pten tumor suppressor gene leads to metastatic prostate cancer. *Cancer Cell* 4, 209–221 (2003).

28. B. S. Carver, J. Tran, A. Gopalan, Z. Chen, S. Shaikh, A. Carracedo, A. Alimonti, C. Nardella, S. Varmeh, P. T. Scardino, C. Cordon-Cardo, W. Gerald, P. P. Pandolfi,

Aberrant ERG expression cooperates with loss of PTEN to promote cancer progression in the prostate. *Nat. Genet.* 41, 619–624 (2009).

29. O. Klezovitch, M. Risk, I. Coleman, J. M. Lucas, M. Null, L. D. True, P. S. Nelson, V. Vasioukhin, A causal role for ERG in neoplastic transformation of prostate epithelium. *Proc. Natl. Acad. Sci. U. S. A.* 105, 2105–2110 (2008).

30. S. A. Tomlins, B. Laxman, S. Varambally, X. Cao, J. Yu, B. E. Helgeson, Q. Cao, J. R. Prensner, M. A. Rubin, R. B. Shah, R. Mehra, A. M. Chinnaiyan, Role of the TMPRSS2-ERG gene fusion in prostate cancer. *Neoplasia* 10, 177–188 (2008).

31. M. J. Kim, R. D. Cardiff, N. Desai, W. A. Banach-Petrosky, R. Parsons, M. M. Shen, C. Abate-Shen, Cooperativity of Nkx3.1 and Pten loss of function in a mouse model of prostate carcinogenesis. *Proc. Natl. Acad. Sci. U. S. A.* 99, 2884–2889 (2002).

32. M. Blattner, D. Liu, B. D. Robinson, D. Huang, A. Poliakov, D. Gao, S. Nataraj, L. D. Deonarine, M. A. Augello, V. Sailer, L. Ponnala, M. Ittmann, A. M. Chinnaiyan, A. Sboner, Y. Chen, M. A. Rubin, C. E. Barbieri, SPOP Mutation Drives Prostate Tumorigenesis In Vivo through Coordinate Regulation of PI3K/mTOR and AR Signaling. *Cancer Cell* 31, 436–451 (2017).

33. K. Ellwood-Yen, T. G. Graeber, J. Wongvipat, M. L. Iruela-Arispe, J. Zhang, R. Matusik, G. V. Thomas, C. L. Sawyers, Myc-driven murine prostate cancer shares molecular features with human prostate tumors. *Cancer Cell* 4, 223–238 (2003).

34. D. Zhao, L. Cai, X. Lu, X. Liang, J. Li, P. Chen, M. Ittmann, X. Shang, S. Jiang, H. Li, C. Meng, I. Flores, J. H. Song, J. W. Horner, Z. Lan, C.-J. Wu, J. Li, Q. Chang, K.-C. Chen, G. Wang, P. Deng, D. J. Spring, Y. A. Wang, R. A. DePinho, Chromatin regulator CHD1 remodels the immunosuppressive tumor microenvironment in PTEN-deficient prostate cancer. *Cancer Discov.* 10, 1374–1387 (2020).

35. D. Zhao, X. Lu, G. Wang, Z. Lan, W. Liao, J. Li, X. Liang, J. R. Chen, S. Shah, X. Shang, M. Tang, P. Deng, P. Dey, D. Chakravarti, P. Chen, D. J. Spring, N. M. Navone, P.

Troncoso, J. Zhang, Y. A. Wang, R. A. DePinho, Synthetic essentiality of chromatin remodelling factor CHD1 in PTEN-deficient cancer. *Nature* 542, 484–488 (2017).

36. Z. Chen, L. C. Trotman, D. Shaffer, H.-K. Lin, Z. A. Dotan, M. Niki, J. A. Koutcher, H. I. Scher, T. Ludwig, W. Gerald, C. Cordon-Cardo, P. P. Pandolfi, Crucial role of p53-dependent cellular senescence in suppression of Pten-deficient tumorigenesis. *Nature* 436, 725–730 (2005).

37. D. Ding, A. M. Blee, J. Zhang, Y. Pan, N. A. Becker, L. J. Maher 3rd, R. Jimenez, L. Wang, H. Huang, Gain-of-function mutant p53 together with ERG proto-oncogene drive prostate cancer by beta-catenin activation and pyrimidine synthesis. *Nat. Commun.* 14, 4671 (2023).

38. X. Wu, J. Wu, J. Huang, W. C. Powell, J. Zhang, R. J. Matusik, F. O. Sangiorgi, R. E. Maxson, H. M. Sucov, P. Roy-Burman, Generation of a prostate epithelial cell-specific Cre transgenic mouse model for tissue-specific gene ablation. *Mech. Dev.* 101, 61–69 (2001).

39. J. M. Chan, S. Zaidi, J. R. Love, J. L. Zhao, M. Setty, K. M. Wadosky, A. Gopalan, Z.-N. Choo, S. Persad, J. Choi, J. LaClair, K. E. Lawrence, O. Chaudhary, T. Xu, I. Masilionis, I. Linkov, S. Wang, C. Lee, A. Barlas, M. J. Morris, L. Mazutis, R. Chaligne, Y. Chen, D. W. Goodrich, W. R. Karthaus, D. Pe'er, C. L. Sawyers, Lineage plasticity in prostate cancer depends on JAK/STAT inflammatory signaling. *Science* 377, 1180–1191 (2022).

40. C. Y. McLean, D. Bristor, M. Hiller, S. L. Clarke, B. T. Schaar, C. B. Lowe, A. M. Wenger, G. Bejerano, GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* 28, 495–501 (2010).

41. E. Y. Chen, C. M. Tan, Y. Kou, Q. Duan, Z. Wang, G. V. Meirelles, N. R. Clark, A. Ma'ayan, Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* 14, 128 (2013).

42. P. P. Roux, D. Shahbazian, H. Vu, M. K. Holz, M. S. Cohen, J. Taunton, N. Sonenberg, J. Blenis, RAS/ERK signaling promotes site-specific ribosomal protein S6

phosphorylation via RSK and stimulates cap-dependent translation. *J. Biol. Chem.* 282, 14056–14064 (2007).

43. J. A. Hutchinson, N. P. Shanware, H. Chang, R. S. Tibbetts, Regulation of ribosomal protein S6 phosphorylation by casein kinase 1 and protein phosphatase 1. *J. Biol. Chem.* 286, 8688–8696 (2011).

44. B. Dong, J.-Y. Xu, Y. Huang, J. Guo, Q. Dong, Y. Wang, N. Li, Q. Liu, M. Zhang, Q. Pan, H. Wang, J. Jiang, B. Chen, D. Shen, Y. Ma, L. Zhai, J. Zhang, J. Li, W. Xue, M. Tan, J. Qin, Integrative proteogenomic profiling of high-risk prostate cancer samples from Chinese patients indicates metabolic vulnerabilities and diagnostic biomarkers. *Nat. Cancer* 5, 1427–1447 (2024).

45. G. L. Lu-Yao, P. C. Albertsen, D. F. Moore, W. Shih, Y. Lin, R. S. DiPaola, S.-L. Yao, Survival following primary androgen deprivation therapy among men with localized prostate cancer. *JAMA* 300, 173–181 (2008).

46. Hormone Therapy for Prostate Cancer. https://www.cancer.org/cancer/types/prostate-cancer/treating/hormone-therapy.html.

47. L. Crowley, F. Cambuli, L. Aparicio, M. Shibata, B. D. Robinson, S. Xuan, W. Li, H. Hibshoosh, M. Loda, R. Rabadan, M. M. Shen, A single-cell atlas of the mouse and human prostate reveals heterogeneity and conservation of epithelial progenitors. *Elife* 9 (2020).

48. W. R. Karthaus, M. Hofree, D. Choi, E. L. Linton, M. Turkekul, A. Bejnood, B. Carver, A. Gopalan, W. Abida, V. Laudone, M. Biton, O. Chaudhary, T. Xu, I. Masilionis, K. Manova, L. Mazutis, D. Pe'er, A. Regev, C. L. Sawyers, Regenerative potential of prostate luminal cells revealed by single-cell analysis. *Science* 368, 497–505 (2020).

49. W. Guo, L. Li, J. He, Z. Liu, M. Han, F. Li, X. Xia, X. Zhang, Y. Zhu, Y. Wei, Y. Li, R. Aji, H. Dai, H. Wei, C. Li, Y. Chen, L. Chen, D. Gao, Single-cell transcriptomics identifies a distinct luminal progenitor cell type in distal prostate invagination tips. *Nat. Genet.* 52, 908–918 (2020).

50. X.-D. Wang, B.-E. Wang, R. Soriano, J. Zha, Z. Zhang, Z. Modrusan, G. R. Cunha, W.-Q. Gao, Expression profiling of the mouse prostate after castration and hormone replacement: implication of H-cadherin in prostate tumorigenesis. *Differentiation* 75, 219–234 (2007).

51. Y. Sun, B.-E. Wang, K. G. Leong, P. Yue, L. Li, S. Jhunjhunwala, D. Chen, K. Seo, Z. Modrusan, W.-Q. Gao, J. Settleman, L. Johnson, Androgen deprivation causes epithelial-mesenchymal transition in the prostate: implications for androgen-deprivation therapy. *Cancer Res.* 72, 527–536 (2012).

52. S. Verma, E. Shankar, F. N. C. Kalayci, A. Mukunda, M. Alassfar, V. Singh, E. R. Chan, G. T. MacLennan, S. Gupta, Androgen deprivation induces transcriptional reprogramming in prostate cancer cells to develop stem cell-like characteristics. *Int. J. Mol. Sci.* 21, 9568 (2020).

53. J. S. Kirk, J. Wang, A. Tracz, M. Long, S. Rosario, Y. Ji, R. Kumar, X. Liu, P. K. Singh, I. Puzanov, G. Chatta, Q. Cheng, J. Huang, J. L. Wrana, J. Lovell, H. Yu, S. Liu, M. M. Shen, T. Liu, D. G. Tang, Integrated single-cell analysis defines the epigenetic basis of castration-resistant prostate luminal cells, *bioRxiv* (2023)p. 2023.03.03.530998.

54. F. Tang, D. Xu, S. Wang, C. K. Wong, A. Martinez-Fundichely, C. J. Lee, S. Cohen, J. Park, C. E. Hill, K. Eng, R. Bareja, T. Han, E. M. Liu, A. Palladino, W. Di, D. Gao, W. Abida, S. Beg, L. Puca, M. Meneses, E. de Stanchina, M. F. Berger, A. Gopalan, L. E. Dow, J. M. Mosquera, H. Beltran, C. N. Sternberg, P. Chi, H. I. Scher, A. Sboner, Y. Chen, E. Khurana, Chromatin profiles classify castration-resistant prostate cancers suggesting therapeutic targets. *Science* 376, eabe1505 (2022).

55. M. Che, A. Chaturvedi, S. A. Munro, S. P. Pitzen, A. Ling, W. Zhang, J. Mentzer, S.-Y. Ku, L. Puca, Y. Zhu, A. M. Bergman, T. M. Severson, C. Forster, Y. Liu, J. Hildebrand, M. Daniel, T.-Y. Wang, L. A. Selth, T. Hickey, A. Zoubeidi, M. Gleave, R. Bareja, A. Sboner, W. Tilley, J. S. Carroll, W. Tan, M. Kohli, R. Yang, A. C. Hsieh, P. Murugan, W. Zwart, H. Beltran, R. S. Huang, S. M. Dehm, Opposing transcriptional programs of

KLF5 and AR emerge during therapy for advanced prostate cancer. *Nat. Commun.* 12, 1–15 (2021).

56. H. Song, H. N. W. Weinstein, P. Allegakoen, M. H. Wadsworth 2nd, J. Xie, H. Yang, E. A. Castro, K. L. Lu, B. A. Stohr, F. Y. Feng, P. R. Carroll, B. Wang, M. R. Cooperberg, A. K. Shalek, F. W. Huang, Single-cell analysis of human primary prostate cancer reveals the heterogeneity of tumor-associated epithelial cell states. *Nat. Commun.* 13, 141 (2022).

57. A. Kiviaho, S. K. Eerola, H. M. L. Kallio, M. K. Andersen, M. Hoikka, A. M. Tiihonen, I. Salonen, X. Spotbeen, A. Giesen, C. T. A. Parker, S. Taavitsainen, O. Hantula, M. Marttinen, I. Hermelo, M. Ismail, E. Midtbust, M. Wess, W. Devlies, A. Sharma, S. Krossa, T. Häkkinen, E. Afyounian, K. Vandereyken, S. Kint, J. Kesseli, T. Tolonen, T. L. J. Tammela, T. Viset, Ø. Størkersen, G. F. Giskeødegård, M. B. Rye, T. Murtola, A. Erickson, L. Latonen, G. S. Bova, I. G. Mills, S. Joniau, J. V. Swinnen, T. Voet, T. Mirtti, G. Attard, F. Claessens, T. Visakorpi, K. J. Rautajoki, M.-B. Tessem, A. Urbanucci, M. Nykter, Single cell and spatial transcriptomics highlight the interaction of club-like cells with immunosuppressive myeloid cells in prostate cancer. *Nat. Commun.* 15, 9949 (2024).

58. Y. N. S. Wong, R. Ferraldeschi, G. Attard, J. de Bono, Evolution of androgen receptor targeted therapy for advanced prostate cancer. *Nat. Rev. Clin. Oncol.* 11, 365–376 (2014).

59. D. J. Mulholland, L. M. Tran, Y. Li, H. Cai, A. Morim, S. Wang, S. Plaisier, I. P. Garraway, J. Huang, T. G. Graeber, H. Wu, Cell autonomous role of PTEN in regulating castration-resistant prostate cancer growth. *Cancer Cell* 19, 792–804 (2011).

60. W. Abida, J. Cyrta, G. Heller, D. Prandi, J. Armenia, I. Coleman, M. Cieslik, M. Benelli, D. Robinson, E. M. Van Allen, A. Sboner, T. Fedrizzi, J. M. Mosquera, B. D. Robinson, N. De Sarkar, L. P. Kunju, S. Tomlins, Y. M. Wu, D. Nava Rodrigues, M. Loda, A. Gopalan, V. E. Reuter, C. C. Pritchard, J. Mateo, D. Bianchini, S. Miranda, S. Carreira, P. Rescigno, J. Filipenko, J. Vinson, R. B. Montgomery, H. Beltran, E. I. Heath, H. I. Scher, P. W. Kantoff, M.-E. Taplin, N. Schultz, J. S. deBono, F. Demichelis, P. S. Nelson, M. A. Rubin, A. M. Chinnaiyan, C. L. Sawyers, Genomic correlates of clinical outcome

in advanced prostate cancer. *Proc. Natl. Acad. Sci. U. S. A.*, doi: 10.1073/pnas.1902651116 (2019).

61. A. Ajkunic, E. Sayar, M. P. Roudier, R. A. Patel, I. M. Coleman, N. De Sarkar, B. Hanratty, M. Adil, J. Zhao, S. Zaidi, L. D. True, J. M. Sperger, H. H. Cheng, E. Y. Yu, R. B. Montgomery, J. E. Hawley, G. Ha, T. Persse, P. Galipeau, J. K. Lee, S. A. Harmon, E. Corey, J. M. Lang, C. L. Sawyers, C. Morrissey, M. T. Schweizer, R. Gulati, P. S. Nelson, M. C. Haffner, Assessment of TROP2, CEACAM5 and DLL3 in metastatic prostate cancer: Expression landscape and molecular correlates. *NPJ Precis. Oncol.* 8, 104 (2024).

62. E.-C. Hsu, M. A. Rice, A. Bermudez, F. J. G. Marques, M. Aslan, S. Liu, A. Ghoochani, C. A. Zhang, Y.-S. Chen, A. Zlitni, S. Kumar, R. Nolley, F. Habte, M. Shen, K. Koul, D. M. Peehl, A. Zoubeidi, S. S. Gambhir, C. A. Kunder, S. J. Pitteri, J. D. Brooks, T. Stoyanova, Trop2 is a driver of metastatic prostate cancer with neuroendocrine phenotype via PARP1. *Proc. Natl. Acad. Sci. U. S. A.* 117, 2032–2042 (2020).

63. J. M. Sperger, K. T. Helzer, C. N. Stahlfeld, D. Jiang, A. Singh, K. R. Kaufmann, D. J. Niles, E. Heninger, N. R. Rydzewski, L. Wang, L. Wang, R. Yang, Y. Ren, J. W. Engle, P. Huang, C. E. Kyriakopoulos, S. F. Slovin, H. R. Soule, S. G. Zhao, M. Kohli, S. T. Tagawa, W. Cai, S. M. Dehm, J. M. Lang, Expression and therapeutic targeting of TROP-2 in treatment-resistant prostate cancer. *Clin. Cancer Res.* 29, 2324–2335 (2023).

64. For adult patients. https://www.trodelvy.com/.

65. Z. Wang, S. L. Townley, S. Zhang, M. Liu, M. Li, M. Labaf, S. Patalano, K. Venkataramani, K. R. Siegfried, J. A. Macoska, D. Han, S. Gao, G. P. Risbridger, R. A. Taylor, M. G. Lawrence, H. H. He, L. A. Selth, C. Cai, FOXA2 rewires AP-1 for transcriptional reprogramming and lineage plasticity in prostate cancer. *Nat. Commun.* 15, 4914 (2024).

66. C. Cheng, J. Hu, R. Mannan, T. He, R. Bhattacharyya, B. Magnuson, J. P. Wisniewski, S. Peters, S. A. Karim, D. J. MacLean, H. Karabürk, L. Zhang, N. J. Rossiter, Y. Zheng, L.

Xiao, C. Li, D. Awad, S. Mahapatra, Y. Bao, Y. Zhang, X. Cao, Z. Wang, R. Mehra, P. Morlacchi, V. Sahai, M. Pasca di Magliano, Y. M. Shah, L. S. Weisman, J. P. Morton, K. Ding, Y. Qiao, C. A. Lyssiotis, A. M. Chinnaiyan, Targeting PIKfyve-driven lipid metabolism in pancreatic cancer. *Nature*, 1–9 (2025).

67. J. Drost, W. R. Karthaus, D. Gao, E. Driehuis, C. L. Sawyers, Y. Chen, H. Clevers, Organoid culture systems for prostate epithelial and cancer tissue. *Nat. Protoc.* **11**, 347–358 (2016).

68. A. S. Goldstein, J. M. Drake, D. L. Burnes, D. S. Finley, H. Zhang, R. E. Reiter, J. Huang, O. N. Witte, Purification and direct transformation of epithelial progenitor cells from primary human prostate. *Nat. Protoc.* **6**, 656–667 (2011).

69. J. C.-Y. Tien, J. Luo, Y. Chang, Y. Zhang, Y. Cheng, X. Wang, J. Yang, R. Mannan, S. Mahapatra, P. Shah, X.-M. Wang, A. J. Todd, S. Eyunni, C. Cheng, R. J. Rebernick, L. Xiao, Y. Bao, J. Neiswender, R. Brough, S. J. Pettitt, X. Cao, S. J. Miner, L. Zhou, Y.-M. Wu, E. Labanca, Y. Wang, A. Parolia, M. Cieslik, D. R. Robinson, Z. Wang, F. Y. Feng, J. Chou, C. J. Lord, K. Ding, A. M. Chinnaiyan, CDK12 loss drives prostate cancer progression, transcription-replication conflicts, and synthetic lethality with paralog CDK13. *Cell Rep. Med.* **5**, 101758 (2024).

70. N. L. Bray, H. Pimentel, P. Melsted, L. Pachter, Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).

71. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

72. M. E. Ritchie, B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, G. K. Smyth, limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

73. Z. Gu, R. Eils, M. Schlesner, Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).

74. ggplot2. https://ggplot2.tidyverse.org/.

75. A. S. Castanza, J. M. Recla, D. Eby, H. Thorvaldsdóttir, C. J. Bult, J. P. Mesirov, Extending support for mouse data in the Molecular Signatures Database (MSigDB). *Nat. Methods* **20**, 1619–1620 (2023).

76. B. Chen, T. M. Lynn-Nguyen, P. Jadhav, B. S. Halligan, N. J. Rossiter, R. M. Guerra, S. Koshkin, I. Koo, P. Morlacchi, D. A. Hanna, J. Lin, R. Banerjee, D. J. Pagliarini, A. D. Patterson, S. Mosalaganti, J. Z. Sexton, T. Calì, C. A. Lyssiotis, Y. M. Shah, BRD4-mediated epigenetic regulation of endoplasmic reticulum-mitochondria contact sites is governed by the mitochondrial complex III, *bioRxivorg* (2024). https://doi.org/10.1101/2024.02.02.578646.

77. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

78. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

79. P. Danecek, J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A. Whitwham, T. Keane, S. A. McCarthy, R. M. Davies, H. Li, Twelve years of SAMtools and BCFtools. *Gigascience* **10** (2021).

80. Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li, X. S. Liu, Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).

81. B. J. Raney, G. P. Barber, A. Benet-Pagès, J. Casper, H. Clawson, M. S. Cline, M. Diekhans, C. Fischer, J. Navarro Gonzalez, G. Hickey, A. S. Hinrichs, R. M. Kuhn, B. T. Lee, C. M. Lee, P. Le Mercier, K. H. Miga, L. R. Nassar, P. Nejad, B. Paten, G. Perez, D. Schmelter, M. L. Speir, B. D. Wick, A. S. Zweig, D. Haussler, W. J. Kent, M. Haeussler, The UCSC Genome Browser database: 2024 update. *Nucleic Acids Res.* **52**, D1082–D1088 (2024).

82. L. J. Zhu, C. Gazin, N. D. Lawson, H. Pagès, S. M. Lin, D. S. Lapointe, M. R. Green, ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics* **11**, 237 (2010).

83. G. Yu, L.-G. Wang, Q.-Y. He, ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382–2383 (2015).

84. F. Ramírez, F. Dündar, S. Diehl, B. A. Grüning, T. Manke, deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187–91 (2014).

85. S. Heinz, C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh, C. K. Glass, Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).

86. D. H. Phanstiel, A. P. Boyle, C. L. Araya, M. P. Snyder, Sushi.R: flexible, quantitative and integrative genomic visualizations for publication-quality multi-panel figures. *Bioinformatics* **30**, 2808–2810 (2014).

87.Chromium Single Cell 3' Reagent Kits User Guide (v3.1 Chemistry Dual Index) - Official 10x Genomics Support, *10x Genomics*. https://www.10xgenomics.com/support/single-cell-gene-expression/documentation/steps/library-prep/chromium-single-cell-3-reagent-kits-user-guide-v-3-1-chemistry-dual-index.

88. Chromium Single Cell, *10x Genomics*. https://www.10xgenomics.com/platforms/chromium.

89. M. D. Young, S. Behjati, SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *Gigascience* **9** (2020).

90. Y. Hao, S. Hao, E. Andersen-Nissen, W. M. Mauck 3rd, S. Zheng, A. Butler, M. J. Lee, A. J. Wilk, C. Darby, M. Zager, P. Hoffman, M. Stoeckius, E. Papalexi, E. P. Mimitou, J. Jain, A. Srivastava, T. Stuart, L. M. Fleming, B. Yeung, A. J. Rogers, J. M. McElrath,

C. A. Blish, R. Gottardo, P. Smibert, R. Satija, Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).

91. P.-L. Germain, A. Lun, C. Garcia Meixide, W. Macnair, M. D. Robinson, Doublet identification in single-cell sequencing data using scDblFinder. *F1000Res.* **10**, 979 (2021).

92. A. Sergushichev, An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation, *bioRxiv* (2016)p. 060012.

93. A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, J. P. Mesirov, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550 (2005).

94. D. van Dijk, R. Sharma, J. Nainys, K. Yim, P. Kathail, A. J. Carr, C. Burdziak, K. R. Moon, C. L. Chaffer, D. Pattabiraman, B. Bierie, L. Mazutis, G. Wolf, S. Krishnaswamy, D. Pe'er, Recovering gene interactions from single-cell data using data diffusion. *Cell* **174**, 716–729.e27 (2018).

95. K. Street, D. Risso, R. B. Fletcher, D. Das, J. Ngai, N. Yosef, E. Purdom, S. Dudoit, Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477 (2018).

96. S. C. Zheng, G. Stein-O'Brien, J. J. Augustin, J. Slosberg, G. A. Carosso, B. Winer, G. Shin, H. T. Bjornsson, L. A. Goff, K. D. Hansen, Universal prediction of cell-cycle position using transfer learning. *Genome Biol* **23** (2022).

97. T. Stuart, A. Srivastava, S. Madad, C. A. Lareau, R. Satija, Single-cell chromatin state analysis with Signac. *Nat. Methods* **18**, 1333–1341 (2021).

98. A. N. Schep, B. Wu, J. D. Buenrostro, W. J. Greenleaf, chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* **14**, 975–978 (2017).

99. O. Fornes, J. A. Castro-Mondragon, A. Khan, R. van der Lee, X. Zhang, P. A. Richmond, B. P. Modi, S. Correard, M. Gheorghe, D. Baranašić, W. Santana-Garcia, G. Tan, J. Chèneby, B. Ballester, F. Parcy, A. Sandelin, B. Lenhard, W. W. Wasserman, A. Mathelier, JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **48**, D87–D92 (2020).

100. S. Durinck, P. T. Spellman, E. Birney, W. Huber, Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184–1191 (2009).

101. A. Zehir, R. Benayed, R. H. Shah, A. Syed, S. Middha, H. R. Kim, P. Srinivasan, J. Gao, D. Chakravarty, S. M. Devlin, M. D. Hellmann, D. A. Barron, A. M. Schram, M. Hameed, S. Dogan, D. S. Ross, J. F. Hechtman, D. F. DeLair, J. Yao, D. L. Mandelker, D. T. Cheng, R. Chandramohan, A. S. Mohanty, R. N. Ptashkin, G. Jayakumaran, M. Prasad, M. H. Syed, A. B. Rema, Z. Y. Liu, K. Nafa, L. Borsu, J. Sadowska, J. Casanova, R. Bacares, I. J. Kiecka, A. Razumova, J. B. Son, L. Stewart, T. Baldi, K. A. Mullaney, H. Al-Ahmadie, E. Vakiani, A. A. Abeshouse, A. V. Penson, P. Jonsson, N. Camacho, M. T. Chang, H. H. Won, B. E. Gross, R. Kundra, Z. J. Heins, H.-W. Chen, S. Phillips, H. Zhang, J. Wang, A. Ochoa, J. Wills, M. Eubank, S. B. Thomas, S. M. Gardos, D. N. Reales, J. Galle, R. Durany, R. Cambria, W. Abida, A. Cercek, D. R. Feldman, M. M. Gounder, A. A. Hakimi, J. J. Harding, G. Iyer, Y. Y. Janjigian, E. J. Jordan, C. M. Kelly, M. A. Lowery, L. G. T. Morris, A. M. Omuro, N. Raj, P. Razavi, A. N. Shoushtari, N. Shukla, T. E. Soumerai, A. M. Varghese, R. Yaeger, J. Coleman, B. Bochner, G. J. Riely, L. B. Saltz, H. I. Scher, P. J. Sabbatini, M. E. Robson, D. S. Klimstra, B. S. Taylor, J. Baselga, N. Schultz, D. M. Hyman, M. E. Arcila, D. B. Solit, M. Ladanyi, M. F. Berger, Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat. Med.* **23**, 703–713 (2017).

102. M. J. Goldman, B. Craft, M. Hastie, K. Repečka, F. McDade, A. Kamath, A. Banerjee, Y. Luo, D. Rogers, A. N. Brooks, J. Zhu, D. Haussler, Visualizing and interpreting cancer genomics data via the Xena platform. *Nat. Biotechnol.* **38**, 675–678 (2020).

103. A. Parolia, M. Cieslik, S.-C. Chu, L. Xiao, T. Ouchi, Y. Zhang, X. Wang, P. Vats, X. Cao, S. Pitchiaya, Others, Distinct structural classes of activating FOXA1 alterations in advanced prostate cancer. *Nature* **571**, 413–418 (2019).

104. S. Hänzelmann, R. Castelo, J. Guinney, GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**, 7 (2013).

105. A. Liberzon, C. Birger, H. Thorvaldsdóttir, M. Ghandi, J. P. Mesirov, P. Tamayo, The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417–425 (2015).

106. A. Parolia, M. Cieslik, S.-C. Chu, L. Xiao, T. Ouchi, Y. Zhang, X. Wang, P. Vats, X. Cao, S. Pitchiaya, Others, Distinct structural classes of activating FOXA1 alterations in advanced prostate cancer. *Nature* **571**, 413–418 (2019).

107. G. Yu, L.-G. Wang, Y. Han, Q.-Y. He, clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).

108. E. Cerami, J. Gao, U. Dogrusoz, B. E. Gross, S. O. Sumer, B. A. Aksoy, A. Jacobsen, C. J. Byrne, M. L. Heuer, E. Larsson, Y. Antipin, B. Reva, A. P. Goldberg, C. Sander, N. Schultz, The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).

109. H. Wickham, Springer; New York: 2009. *Ggplot2: elegant graphics for data analysis*.

110. K. Alboukadel, ggpubr:"ggplot2" based publication ready plots. . *R-project. org/package= ggpubr (accessed 12 August …*.

111. J. Qin, B. Dong, J.-Y. Xu, Y. Huang, J. Guo, Q. Dong, Y. Wang, N. Li, Q. Liu, M. Zhang, Q. Pan, H. Wang, J. Jiang, B. Chen, D. Shen, Y. Mao, L.-H. Zhai, J. Li, W. Xue, M. Tan, J. Zhang, Integrative proteomic analysis reveals metabolic vulnerabilities and diagnostic biomarkers in high-risk prostate cancer, figshare (2024); https://doi.org/10.6084/M9.FIGSHARE.25706982.

Materials and Methods

Ethical statement

All experiments detailed in this paper complied with the Institutional Review Board and the Institutional Animal Care and Use Committee (IACUC) at the University of Michigan.

Animal procurement

Animal studies were approved by the Institutional Animal Care and Use Committee at the University of Michigan. Animal use and care were in strict compliance with institutional guidelines, and all experiments were performed in compliance with relevant regulatory standards laid down by the university. Transgenic FOXA1 mouse lines were obtained from Baylor College of Medicine and were generated in a *C57BL/6* background. *Probasin-Cre* (Stock#026662) mice and *Trp53f/f* (Stock#008462) mice were obtained from Jackson Laboratory (Bar Harbor, ME). For the generation of the syngeneic models, male *C57BL/6J* (Stock# 000664) mice (6-8-weeks old) and NOD Cg-Prkdc<scid> ll2rg<tm1Wjl>SzJ (NSG) mice were obtained from the Jackson Laboratory (Stock#005557). For the subcutaneous grafting assay, male *CB17* SCID mice were obtained from Charles River (Stock#236). Mice were housed in a pathogen-free animal barrier facility under a pathogen-free, 12 h light/12 h dark cycle, 18–23 °C temperatures, and 40–60% humidity.

Generation of FOXA1 transgenic mice

FOXA1 mutant transgenic models were generated as described earlier (*41*). Briefly, plasmids containing CMV enhancer-driven human FOXA1 mutations (Class 1, 2, and 3) were injected into the pronuclei of *C57BL/6* mouse fertilized eggs, which were then implanted into pseudo-pregnant females. While these models were generated using the human FOXA1 coding sequence, the mouse and human FOXA1 sequences are highly conserved, with a 96.6% similarity at the amino acid level, with the Wing 2 and forkhead functional domains being identical. These experiments were performed as a service in Dr. Jianming Xu's laboratory at Baylor College of Medicine, Houston, Texas. PCR screening was used to identify potential founder animals. Sperms from all transgenic lines have been cryopreserved at the University of Michigan and can be revived and shared with the scientific community upon request.

Genotyping of tail DNA from FOXA1 mutant mice was performed using Agarose Gel Electrophoresis and PCR. The following sequences of primers were used for PCR analysis:

Genotyping_Left_F: 5′-GAGTTCTCTGCTGCCTCCTGG-3′
Genotyping_Left_R: 5′-GAGCGGCCAGCTTATCGATAC-3′
Genotyping_Right_F: 5′-GAAAGAACCAGCTGGGGCTC-3′
Genotyping_Right_R: 5′-CGAAAATCTGTGGGAAGTCTTG-3′
Span_Forward: 5′-GAGTTCTCTGCTGCCTCCTGG-3′
Span_Reverse: 5′-CGAAAATCTGTGGGAAGTCTTG-3′

The Genotyping Left (GT-L) primers generated a product size of 225 bp, and the Genotyping Right (GT-R) primers generated a product size of 385 bp, while the span primers generated a product size of 213 bp. PCR conditions used were 35 cycles (15 s at 94 °C, 30 s at 58 °C and 30 s at 72 °C) using high fidelity Platinum™ Taq DNA Polymerase (Thermo Fisher Scientific, Cat#11304011), 10 µM of the primers and 2µl of the tail DNA. For the castration experiments, sham or castration surgeries were performed in compliance with the University of Michigan IACUC guidelines.

Allograft model

Organoids were trypsinized, filtered, and resuspended in Matrigel for subcutaneous injections. Briefly, 3 x 10^6 cells were injected into the dorsal flanks of NSG mice. Tumor growth was monitored weekly. Once *R265-71del+/+; Trp53f/f* tumors reached 1000 mm^3, the resected tumors were cut into small size pieces and implanted into the subcutaneous space of both flanks of *C57BL/6* mice to generate the syngeneic model.

Antibodies

For immunoblotting, the following antibodies were used: AR (Millipore: 06-680); V5 tag (Thermo Fisher Scientific: R960-25); FOXA1 (Millipore: 05-1466); c-MYC (Cell Signaling Technologies: 5605S); vinculin (Sigma Aldrich: V9131); H3 (Cell Signaling Technologies: 3638S); KLF5 (Proteintech: 21017-1-AP); phospho-AKT (Ser473) (D9E) (Cell Signaling Technologies: 4060S); phospho-S6 ribosomal protein (Ser235/236) (Cell Signaling Technologies: 2211S).

For immunohistochemistry, the following antibodies were used: Ki67 (BD Pharmingen™: 550609); V5 tag (Thermo Fisher Scientific: R960-25); AR (Millipore: 06-680); FOXA1 (Thermo Fisher Scientific: PA5-27157); TROP2 (Bio-Techne R&D Systems: BAF1122); alpha-SMA (Ventana Medical Systems, 760-2833).

For immunofluorescence, the following antibodies were used: CK8 (Abcam: ab53280); AR (Ventana: 760-4605, Abcam: ab108341); V5 tag (Thermo Fisher Scientific: R960-25); TROP2 (Bio-Techne R&D Systems: BAF1122); NSD2 (Abcam: ab75359); NKX3.1 (Athena Enzyme Systems/Athena Environmental Sciences 0317).

For ChIP-seq, the following antibodies were used: AR (Millipore: 06-680); KLF5 (Proteintech: 21017-1-AP); FOXA1 (Thermo Fisher Scientific: PA5-27157); FOXA1 (Cell Signaling Technology: 53528).

Immunoblotting

Cell lysates were prepared and immunoblotting was performed as described earlier (*46*). Briefly, cells were lysed in RIPA buffer (ThermoFisher Scientific: 89900) containing Protease and Phosphatase Inhibitor Cocktail (ThermoFisher Scientific: 78440) and denatured for 10 minutes by heating at 70 °C. Next, Pierce 660nM protein assay reagent was used to measure the protein concentration (Thermo Fisher Scientific: 22660). 10-30 µg of total protein was loaded per lane and separated in either NuPAGE 3 - 8%, Tris-Acetate Protein Gel (ThermoFisher Scientific) or NuPAGE 4 - 12%, Bis-Tris Protein Gel (ThermoFisher Scientific), and transferred in semi-dry transfer system (Trans-blot Turbo System; BioRad) at 25 volts for one hour. Membrane blocking was performed for one hour in 5% non-fat dry milk and incubated overnight at 4°C with primary antibodies. Secondary antibodies with horseradish peroxidase (HRP; BioRad) conjugation were used at 1:20,000 diluted in TBST, and the membrane was incubated for one hour at room temperature. Blots were finally developed using enhanced chemiluminescence (ECL Prime, Thermo Fisher Scientific) and imaged on an Odyssey Fx Imager (LiCOR Biosciences) following the manufacturer's protocol.

The following antibodies were used for immunoblotting: AR (Millipore: 06-680); V5 tag (Thermo Fisher Scientific: R960-25); FOXA1 (Millipore: 05-1466, Thermo Fisher Scientific: PA5-27157); c-MYC (Cell Signaling Technologies: 5605S); vinculin (Sigma Aldrich: V9131); H3 (Cell Signaling Technologies: 3638S); KLF5 (Proteintech: 21017-1-AP); PTEN (D4.3) (Cell Signaling Technology: 9188S); AKT (Cell Signaling Technology: 9272S); phospho-AKT (Ser473) (D9E) (Cell Signaling Technologies: 4060S); S6 ribosomal protein (5G10) (Cell Signaling Technologies: 2217S); phospho-S6 ribosomal protein (Ser240/244) (Cell Signaling Technologies: 2215S), TROP2 (Thermo Fisher Scientific: PA5-86376).

Generation of mouse prostate organoids

To generate mouse prostate organoids, mice were euthanized, and the distinct prostate lobes were micro-dissected. Unless specified otherwise, all four prostate lobes (anterior, dorsal, ventral, and lateral) were pooled and minced using a disposable scalpel into small pieces of approximately 1.0 mm3 in a 10-cm culture dish under aseptic conditions. Subsequently, the minced tissues were transferred into 15 mL tubes containing prewarmed (37 °C) advanced DMEM F-12 +/+/+ media with 5 mg/mL of collagenase Type II (Gibco, 17101015) and incubated on a rotator for 1 hour at 37°C. Next, the tissues were digested with TrypLE (Thermo Fisher Scientific, 12605010) for 15 minutes, followed by a media wash. The cell pellets were finally suspended in 90% growth factor reduced Matrigel (Corning®, 354230) diluted with advanced DMEM F-12 media on ice and plated in a 24-well tissue culture plate (40 µL/drop). After solidifying the gel at 37 °C in the 5% CO2 incubator for 15 minutes, 0.5 mL/well of complete mouse prostate organoid culture media were added. The components of the mouse prostate media were adapted from previously published protocols(*47*).

Organoid Adenoviral Cre and CRISPR/Cas-9 lentiviral transduction

In vitro CRE recombination was achieved by adenoviral delivery of CRE recombinase-containing virus or empty vector based on previously published protocols (*48*). For the adenoviral transduction, 5µl of a 10X viral stock was added to 100k cells, along with 16µg/mL polybrene, and incubated for 1 hour at 37°C on a rotor. Following this, the tubes were centrifuged for one hour at 750xg at 25°C, washed twice with 1X PBS, and plated into four 40µl Matrigel domes. Immunoblots for V5 were run to confirm recombination and transgene expression.

Similarly, CRISPR/Cas-9 mediated knockout of *Trp53* and *Nsd2* was performed by lentiviral delivery of a CRISPR-V2 plasmid encoding an active Cas9 and sgRNA sequences. Briefly, 1 mL of the 10x lentivirus was added along with polybrene to 100k cells. Spinfection was performed as described above, and immunoblotting or immunohistochemistry was performed to validate protein knock-out. The sgRNAs used in this study are as follows:

sgTrp53 #1:5'-CACCGGCGGTTCATGCCCCCCATGC-3'

sgTrp53 #2:5'-CACCGGTGAAATACTCTCCATCAA G-3'

sgNsd2#1:5'-CACCGTTGATAGGTGTAGTATTGGG-3'

sgNsd2#2:5'-CACCGCCTTAGCTACTTGAGGGTTG-3'

Organoid growth and limiting dilution assays

Prior to plating for growth assays, organoids were digested with TrypLE for 10 minutes. 250-500 cells were plated per Matrigel dome in a 12-well plate in differing media conditions. Cell viability was assayed at Day 0 for six days according to the CellTiter-Glo 3D Cell Viability Assay Kit (Promega G9683).

For the organoid limited dilution assay, 50, 100, 500, and 1000 cells were plated in 40µl or 25µl Matrigel domes in a 24-well plate (manual) or 96-well plate (IncuCyte), respectively, in complete mouse prostate organoid media. For the manual assay, organoids were cultured for eight days before imaging, while for the IncuCyte, images were captured every day, and representative images from Day 8 were shown.

In situ proximity ligation assay (PLA)

PLA was performed using a previously described protocol (49). Briefly, FOXA1 Class 1 mutant NSD2 wild-type and NSD2 knock-out organoids were fixed for 10 minutes in 4% PFA and washed thrice with 1x PBS. After ethanol-based dehydration and sectioning, slides were deparaffinized and antigen retrieval was performed by boiling slides for 15 minutes in citrate buffer (pH 6.0). Next, slides were permeabilized using 0.5% Triton X-100% (in PBS) for 10 min. The NaveniFlex PLA kit staining kit was used with mild modifications. Briefly, slides were blocked in 10% goat serum, followed by incubation with the desired primary antibodies (AR abcam, Cat#ab108341; NSD2 abcam, Cat#ab75359) at four degrees overnight. Finally, slides were mounted using Prolong Gold Antifade mounting media and the PLA signal intensity/spot was determined with ImageJ software.

RNA isolation and sequencing

Prostate organoids were seeded at a density of 20,000 cells/40µl Matrigel dome in 2 replicates on Day 0 and incubated for 48 hours in complete organoid media, followed by a 12-hour starvation in media lacking EGF and DHT. Organoids were stimulated with 10 nM DHT for 24 hours to define a mouse-specific AR signature. RNA isolation and library preparation were performed using previously described protocols(*5*, *49*). Briefly, total RNA was then isolated using QIAzol Lysis Reagent (QIAGEN) and extracted using the miRNeasy Mini Kit (Qiagen). SuperScript III Reverse Transcriptase enzyme (Thermo Fisher Scientific) was used for conversion of 1ug RNA into cDNA following the manufacturer's instructions.

RiboErase RNA-seq libraries were prepared using 200–1,000 ng of total RNA extracted from mouse organoids. Ribosomal RNA was eliminated through enzymatic digestion using the KAPA RNA Hyper+RiboErase HMR Kit (Roche). RNA was then fragmented to approximately 200-300 bp using heat and double-stranded cDNA was synthesized after fragmentation, followed by end-repair and ligation using New England Biolabs adapters. The final library preparation involved amplification (2x KAPA HiFi HotStart mix) and addition of NEB dual barcodes, following the manufacturer's protocol. Quality control of the libraries were performed using the Agilent 2100 Bioanalyzer and paired-end sequencing was performed on the Illumina Novaseq, with 2 × 100 nucleotide read lengths and a sequence coverage of 15–20 million paired reads per sample.

Alignment was performed using Kallisto (v0.46.1) to the mouse (mm10/GRCm38) or human (hg38/GRCh38) references(*50*). Read counts were adjusted for normalization and filtered in R via package EdgeR (*51*). Package Limma-Voom (vlimma_3.53.10) was involved in differential expression analysis (*52*). GSEA analysis and corresponding heatmaps and figures were created using R package fgsea (vfgsea_1.24.0), ComplexHeatmap, and ggplot2 for signatures from MSigDB's hallmark MTORC1 and custom AR signatures based on our data (*53–55*).

ChIP-seq and analyses

Prostate organoids were seeded at a density of 30,000 cells/well and grown for 3-5 days in complete media. Organoids were stimulated with 1 nM of R1881 for 6 hours before harvest. Organoids were dissociated using TrypleLE for 10 minutes and split into 500,000 to 1 million cells per tube.

Chromatin immunoprecipitation (ChIP) experiments were performed as described earlier(*5, 56*). The Ideal ChIP-seq Kit for transcription factors (Diagenode) was used following the manufacturer's protocol. 500,000 to 1 million cells were used per ChIP reaction with 2µg of antibody. Briefly, cells were crosslinked, lysed and sonicated using a Bioruptor Pico, Diagenode machine to achieve a desired chromatin size of about 200bp. Sheared chromatin was incubated with desired antibodies overnight at 4 °C followed by de-crosslinking and DNA purification the next day.

Purified DNA was prepared for sequencing using the manufacturer's instructions (Illumina). Library preparation and processing of reads were performed as described earlier (*56*). Quality control of the final libraries were performed using the Bioanalyzer 2100 (Agilent) and sequenced on the Illumina Novaseq Sequencer (125-nucleotide read length).

Reads were processed as described in Parolia et al (*10*). Briefly, reads were trimmed using Trimmomatic version 0.39 (settings TruSeq3-PE-2.fa:2:30:10, minlen 50), and aligned using bwa ("bwa mem," options -5SP -T0, version 0.7.17-r1198-dirty) to mm10 (GRCm38) genome reference or hg38 (GRCh38) reference(*57, 58*). Next, reads were filtered using Markduplicates from Picard in addition to a quality score filtering of >20 via samtools(*59*). Peak calling was performed using MACS2, filtered using bedtools, and converted to bigwigs with UCSC

wigtoBigwig(*60*, *61*). Cistrome overlap analysis was performed in R (v3.6.0) using ChipSeekAnno (v3.0.0) and ChipSeeker (v1.29.1)(*62*, *63*). Enrichment heatmaps were generated using Deeptools(*64*). HOMER was used to perform motif enrichment analysis (*65*). Motif dot plot summaries were created in R (v3.6.0) using Sushi (*66*).

Single-cell RNA and multi-omics (ATAC+RNA) sample preparation

Mice were euthanized, and distinct prostate lobes were micro-dissected. Unless specified otherwise, all four prostate lobes were pooled and minced using a disposable scalpel into small pieces of approximately 1.0 mm^3 in a 10-cm culture dish under aseptic conditions. Subsequently, the minced tissues were transferred into 15 mL tubes containing prewarmed (37 °C) advanced DMEM F-12 +/+/+ media with 5 mg/mL collagenase Type II (Gibco, 17101015) and incubated on a rotator for 30 minutes at 37°C. Next, the tissues were digested with TrypLE (Thermo Fisher Scientific, 12605010) for 15 minutes, followed by a media wash. Single cells were strained using a 70 and 40-micron filter, followed by a 0.4% BSA in 1XPBS wash. Following this, cells were either processed for single-cell RNA-sequencing or 10x single-cell multi-omics (ATAC+RNAseq).

Single-cell RNA sequencing was performed using the Chromium Next GEM Single Cell 3' Kit v3.1 (10x Genomics) according to the manufacturer's instructions (*67*). Briefly, single-cell suspensions were loaded onto a Chromium Next GEM Chip along with reverse transcription reagents and Gel Beads containing barcoded oligonucleotides. The chip was then placed in a Chromium Controller to generate Gel Bead-in-Emulsions (GEMs), encapsulating individual cells with unique barcoded Gel Beads. Cellular mRNAs were reverse transcribed within each GEM to generate barcoded full-length cDNA. Following reverse transcription, the GEMs were broken, and the barcoded cDNA was amplified by PCR. The amplified cDNA underwent fragmentation, end repair, and A-tailing, followed by adapter ligation and PCR amplification to generate final sequencing libraries. Library quality was assessed using an Agilent Bioanalyzer.

Single-cell multi-omics was performed using the Chromium Next GEM Single-cell Multiome ATAC+Gene expression kit (10x Genomics) according to the manufacturer's protocol (*68*). Briefly, nuclei were isolated, followed by transposition and adapter ligation using a Tn5 transposase. Transposed nuclei were loaded onto a Chromium Next GEM microfluidic chip along

with gel beads containing 10x barcoded oligonucleotides. The GEMs were incubated to produce barcoded DNA from transposed chromatin (ATAC) and barcoded full-length cDNA from polyadenylated mRNA (Gene expression) within each GEM. Following this, the GEM was broken, cDNA was amplified, and pre-amplified cDNA was used as input for the ATAC and gene expression library preparation. For ATAC, sample indexing primers were used for further amplification while the amplified cDNA underwent fragmentation, end repair, and A-tailing, followed by adapter ligation and PCR amplification to generate final sequencing libraries. Library quality was assessed using an Agilent Bioanalyzer.

Single-cell RNA-seq and multiome (single nucleus RNA-seq and ATAC-seq) sequencing

Single-cell RNA-sequencing: Data processing was performed using a previously described protocol (*49*). Briefly, after sequencing, read demultiplexing, alignment, and gene quantification were conducted with the 10X Genomics Cell Ranger pipeline (v5.0) for scRNA libraries or Cell Ranger ARC pipeline (v2.0) for multiome libraries. Custom reference was built by adding sequences of human FOXA1 and puroR to the prebuilt mouse reference (mm10) downloaded from the 10X website. To reduce potential ambient RNA contamination, especially for nuclear libraries, SoupX (*69*) was applied to the raw gene count matrix obtained from Cell Ranger, and the corrected read count matrix was used for downstream analyses with Seurat (v4.1)(*70*) if not specified otherwise. Using the median number of detected genes and fraction of mitochondrial reads per cell, low-quality cells were further filtered (table S3). In addition, putative doublets were identified using the pre-SoupX matrix with scDblFinder (*71*) followed by removal of mitochondrial genes from the matrix. After QC steps, the SoupX corrected count matrices of replicates from paired libraries (e.g., case and control) were pooled and normalized. The top 2000 highly variable genes were identified followed by assignment of cells to clusters. Cell annotation was performed using the TransferData method and public datasets as reference (*32–34*), which were manually inspected and adjusted according to top marker gene expression. Sub-clustering was conducted with all cells annotated as Basal, Luminal, or LumP to capture the heterogeneity of epithelial cells.

To conduct gene set enrichment analysis (GSEA) between groups, pseudo-bulk gene expression profiles were generated and TPM-based normalized expression was calculated with edgeR by using the TMM scaling factors (*51*). In Class 1 case versus control comparison, luminal cells in

case that were puroR-;hFOXA1+ (based on imputed data) were compared to luminal cells in control that were puroR+; whereas for basal cells, puroR+ were used for case and control. Genes ranked by logFC were used as input for pre-ranked GSEA with fgsea (*72*). MsigDB was used to download Hallmark gene sets (*73*). When noted, MAGIC (*74*) imputed gene expression was used for data visualization with heatmap and violin plot. Pseudo-time analysis was performed with Slingshot (*75*) using MAGIC imputed gene expression matrix. The gene expression analyses of the multiome libraries followed the same pipeline as the single-cell RNA-seq libraries described above. Prediction of cell cycle position was performed with tricycle (*76*)

Gene signatures: Gene signatures from Class 1 and Class 2 single-cell RNA-sequencing or multi-omics data were defined by performing pseudo-bulk analyses between the case and control-specific clusters. Pure tumor cells in the pseudo-bulk analyses of 100-week-old single-cell RNA-sequencing data were identified by expression of transgenic FOXA1 and lack of puroR mRNA expression. This selection strategy enabled the removal of cells within case tissues that escaped Cre recombination, thus retaining puroR transcript expression.

The AR signature (only comprising activated genes) was defined using common genes up-regulated in wild-type mouse prostate organoids stimulated with DHT (10 nM) for 24 hours. mTOR signature from the MSigDB hallmark database (HALLMARK_MTORC1_SIGNALING) was used. Basal, Lum, and LumP signatures were selected from the top 25 up-regulated genes in differential analysis between one versus others (ranked by p values from Wilcoxon's test) using the single-cell data from a 100-week control mouse. Top marker genes of basal cells expressed in LumP cells were excluded to avoid overlapping signatures. Prostate adenocarcinoma-related transcription factors were obtained from a previous publication (*26*). The FOXA1-driven PCa signature was defined as the top 50 up-regulated genes in the Class 1 case vs control luminal cell clusters (fig. S3F). Markers for LumP (*32–34*),Luminal-2 (*32–34*), and Luminal-C (*32–34*) were identified using data and annotation from corresponding studies by comparing to all other epithelial cells. Consensus stemness (42 genes) signature was determined from the intersection of the top 100 up-regulated genes in LumP, Luminal-2, and Luminal-C (fig. S8B).
The castration signature was derived using multiome gene expression data collected from wild-type prostates or residual prostate tissues 2 weeks post-castration. Specifically, differential analysis

was conducted between castrated and wild-type (i.e., hormonally intact) epithelial cells at the pseudo-bulk level, and the top 20 up-regulated genes were identified. Genes with similar expression levels in basal cells were filtered out (the difference in mean expression between castrated luminal and basal was <1 log2TPM).

Single-cell ATAC-seq: The ATAC-seq data of the multiome libraries from the Class 2 case and control animals were pooled and analyzed together. To filter low-quality cells, only cell barcodes associated with more than 1000 fragments in peaks were kept. If not specified otherwise, all downstream analyses were performed with Signac (v1.5.0)(*77*) and Seurat (v4.1)(*70*). To integrate different libraries, a common set of peak regions was first created by merging peak regions of all libraries; fragments in the consensus peak set were then re-counted using the raw fragments data for each library. The resultant peak cell matrices from all libraries were pooled and subjected to term frequency-inverse document frequency (TF-IDF) normalization. Subsequently, linear dimension reduction data was performed based on the top 50% abundant peaks using singular value decomposition (SVD). The top 30 latent semantic indexing (LSI) components, except the first component, were used to perform nonlinear dimension reduction using UMAP and construct a shared nearest neighbor graph before clustering with the SLM algorithm. To find differentially accessible regions (DARs) between cell groups, the findMarker function was utilized with the LR (logistic regression) test method and the total number of fragments as a latent variable. Motif enrichment in the top 2000 DARs was calculated using a hypergeometric test. Motif activity score at the cell level was calculated by running chromVAR (*78*) with JASPAR2020 motifs (*79*). biomaRT was used to map the mouse homolog genes of the human signature (*80*).

Cohorts, datasets, and resources

Pie charts for FOXA1 mutations were made from previous studies (*2*, *3*, *5*, *6*, *8*, *81*).

For the TCGA prostate adenocarcinoma RNA-seq analysis, the TPM matrix was downloaded from UCSC Xena browser (https://xena.ucsc.edu/)(*82*). FOXA1 class 1 cases were identified using an in-house pipeline (*83*). PI3K altered cases include cases with driver mutations in AKT1, PIK3CA, PTEN or BRAF and cases with PTEN homozygous deletion; cases with ETS fusions were excluded. Cases that are not grouped into the major molecular subtypes (ETS, SPOP, FOXA1 and IDH1) were selected for comparison (the "Other" group). As we noticed substantial heterogeneity

within the “Other” group, we conducted k-means clustering based on single sample gene set enrichment scores (GSVA(*84*)) of hallmark pathways (*85*), and divided it into two groups. One group seemed to have fewer CNV events than the other and was used as a pure “wild-type” reference group in differential analysis. Differential analysis between molecular subtypes of interest (example: FOXA1 Class 1 and PI3K altered) and the wild-type group was performed with Limma after filtering out lowly expressed genes and log-transformation of the TPM matrix. The human FOXA1 Class 1 tumor signature was defined as top 300 up-regulated genes (ranked by p-value) compared to the pure wild-type tumors. The mouse FOXA1 Class 1 tumor signature was derived from differential analysis of pseudo-bulk profiles of the tumor cell cluster of FOXA1 Class1 versus luminal cells of the wild-type (single cell RNA-seq data); genes with higher expression in basal cells (by 2-fold) than in luminal cells were excluded; top 300 genes ranked by p-value were then selected as the respective human and mouse signatures.

For SU2C RNA-seq analysis, transcriptomic data from an extended SU2C cohort (n=757) were downloaded in the form of RPKM data. FOXA1 class 1 cases were identified by an in-house pipeline (*83*)(*86*). PI3K pathway alteration was defined by pathogenic mutations in AKT1, PIK3CA, PIK3CB, or PTEN losses. The wild-type control group was defined as samples lacking alterations in ETS, SPOP, FOXA1, IDH1, and PI3K pathway genes. Differential expression analysis was performed using limma-trend with empirical Bayes moderation. The top 200 significantly ranked genes underwent Gene Ontology enrichment analysis via clusterProfiler v4.10.1 using default parameters (*87*). All statistical analyses were conducted in R v4.3.3. Data visualization employed ggplot2 v3.5.1 and ggpubr v0.6.0, with boxplots displaying median ± interquartile range.

Similar transcriptomic analysis from Asian primary PCa was performed in the CPGEA cohort. FOXA1 Class 1 cases were identified using a published pipeline (*8*). PI3K-altered cases include cases with PTEN mono and biallaelic losses, AKT mutations or PIK3CA amplifications and hotspot mutations. Cases that are not grouped into the major molecular subtypes (SPOP, MAP3K7, BRAF) were selected for comparison as the “Other” group. Next, DEG methodology was used to compare transcriptomes between FOXA1 Class 1 and PTEN/AKT-altered primary PCa. Briefly, we first identified DEGs in the PI3K-mutant vs common “Other” tumor sub-group (i.e., cases without FOXA1 Class 1 or PI3K alterations omitting MAP3K7 deletions, BRAF

mutations and SPOP mutant cases). DEGs were also identified for FOXA1-Class 1 mutant vs the same “Other” tumor subtype and statistical significance was assessed using the prop.test. Next, to generate the transcriptomic concordance plot shown below, we plotted the fold-change of all DEGs (using the following threshold -1.5>log2FC>1.5) to assess transcriptomic concordance between FOXA1 Class1 and PI3K-altered cases. For GSEA analysis, we ranked genes based on their log_2FC values from the Class 1 (n=56) vs. Wildtype (n=62, omitting PI3K-altered cases, MAP3K7 deletions, BRAF mutations and SPOP mutant cases) comparison and performed GSEA analysis for hallmark AR signaling pathway.

TCGA PRAD protein analysis: Normalized RPPA value was downloaded from cBioportal (https://www.cbioportal.org/)(*88*). Boxplots of selected proteins in different molecular subtypes were generated with ggplot (*89*) and ggpubr (*90*); the non-parametric Wilcoxon signed-rank test was used for two-group comparison.

Asian primary PCa protein analysis: The phosphoproteomic data were obtained from the Figshare repository (DOI: 10.6084/m9.figshare.25706982)(*31*)(*91*). FOXA1 class 1 cases were identified by an in-house established pipeline (*86*). Cases with canonical alterations in ETS fusion genes, SPOP mutations, FOXA1 mutations outside Class 1, and IDH1 mutations were excluded. Remaining samples constituted the "Other" comparison group. Differential phosphorylation between FOXA1 Class 1 and Other groups was assessed using two-tailed Student's t-tests. Boxplots of selected proteins in different molecular subtypes were generated with ggplot(*89*) and ggpubr (*90*)

Signature activity scores of KLF5 and Club signatures were calculated using the Single Sample Gene Set Enrichment Analysis (ssGSEA) method using the GSVA package in R. FOXA1 Class 2 clinical cases were stratified into Klf5 high and low groups based on a bimodal distribution of the CRPC_KLF5 signature. Here, we selected all samples with the Class 2 mutation and ssgsea activity of above 0.45 as KLF-high, and any Class 2 samples with ssgsea lower than 0.45 as KLF5-low.

Tissue and organoid histology

Prostate tissue was fixed in 10% neutral buffered formalin overnight, dehydrated with ethanol, paraffin embedded, and serial sectioned at 4 μm thickness. Organoids were seeded at a density of

20,000 cells per 40 μL Matrigel dome and incubated in complete mouse prostate culture media for 5 to 7 days. Subsequently, organoid samples were harvested, embedded in Histogel, and fixed in 4% paraformaldehyde (PFA) for 1 hour, followed by dehydration with ethanol, paraffin embedding, and serial sectioning at five μm thickness.

Histological analysis

Two genitourinary pathologists examined H&E-stains from formalin-fixed paraffin-embedded (FFPE) tissue sections in a blinded manner. Histological grading was conducted following a pre-approved schema at the beginning of the study (*49*). Each biopsy was assessed based on five broad histological categories through microscopic evaluation. These categories were as follows: Category 0 - indicating unremarkable prostate histology; Category 1 - indicating hyperplasia; Category 2 - indicating low and high-grade prostatic intraepithelial neoplasia (PIN); Category 3 - indicating florid high-grade PIN/atypical intraductal proliferation (AIP)/intraductal carcinoma-like; and category 4 - indicating invasive adenocarcinoma prostate (**fig. S1H**). Each prostate sample was scored by assigning a percentage prevalence to the five subcategories. The total score was then assigned to each biopsy and calculated as the sum of the percentage of cells in different subcategories (PC) with a score of 0 to 4. The total score (out of a possible 400) was calculated using the formula: Total Score = (PC Category 0 * 0 + PC Category 1 * 1 + PC Category 2 * 2 + PC Category 3 * 3 + PC Category 4 * 4).

Immunohistochemistry

Immunohistochemistry was performed manually or using the Ventana automated staining system (Roche Tissue Diagnostics). Antibodies used for immunohistochemistry are as follows: Ki67 (BD Pharmingen™: 550609); V5 tag (Thermo Fisher Scientific: R960-25); AR (Millipore: 06-680); FOXA1 (Thermo Fisher Scientific: PA5-27157); TROP2 (Bio-Techne R&D Systems: BAF1122); SMA (Roche-Ventana Medical Systems, 760-2833); p53 (Leica: P53-CM5P-L).

Manual: Stained slides were deparaffinized and incubated with a Citrate-based Antigen Unmasking Solution (Vector Laboratories, H-3300-250). 3% hydrogen peroxide (Sigma) was used to inactivate endogenous peroxidases. Primary antibodies were diluted in either normal horse serum (Vector Laboratories, S-2012-50) or 10% normal goat serum (Vector Laboratories, S-1000-20) and incubated overnight. Antibodies were detected with DAB Peroxidase Substrate kit (Vector

Labs). The slides were counterstained with hematoxylin and cover-slipped with EcoMount (BioCare Medical).

Ventana: Ventana staining was carried out using a previously described protocol (*10*). Briefly, singleplex IHC was carried out on the Ventana ULTRA automated slide staining system (Roche-Ventana Medical Systems) using the OmniView Universal Diaminobenzidine Detection Kit (catalog number 760-500, Roche-Ventana) and Hematoxylin II (catalog number 790-2208, Roche-Ventana) for counterstain. Staining was evaluated under 10x and 20x magnification using a bright-field microscope.

Multiplex Immunofluorescence (IF)

Multiplex IF was performed following a protocol described previously (*10*). The consecutive IF detection system was developed using the Discovery FITC Kit (catalog number 760-232, Roche-Ventana), Discovery Red 610 (catalog number 760-245), and the Discovery Cy5 Kit (catalog number 760-238, Roche-Ventana) with DAPI (Prolong Gold anti-fade, catalog number P36931, Invitrogen/Fisher-Scientific). Imaging of multiplex staining was performed using the EVOS 7000 system. Antibodies used for immunofluorescence were as follows: CK8 (Abcam: ab53280); AR (Ventana: 760-4605); V5 tag (Thermo Fisher Scientific: R960-25); TROP2 (Bio-Techne R&D Systems: BAF1122); NSD2 (Abcam: ab75359); NKX3.1 (Athena Enzyme Systems/Athena Environmental Sciences 0317). Fluorescent imaging were performed at 10x and 20x magnification.

Chapter acknowledgments

Contents of Chapter 3 were published online as part of this article:. *Science*, 2025. doi: 10.1126/science.adv2367. **Eyunni S**, Young E, George J, Mannan R, Carson SE, Zhang Y, Tien J, Jaber M, Luo J, Pang M, Mehra R, Cao X, Su F, Wang R, Cieslik M, Lee DK, Xu J, Chinnaiyan AM, Parolia A. Divergently evolved FOXA1 mutations drive prostate tumorigenesis or therapy-resistant intra-luminal plasticity**.**

Author contributions

S.E., A.M.C., and A.P. conceived the study and the experiments; A.P. and S.E. designed and carried out the multi-omics and functional experiments with assistance from J.L., Q.Z., M.P., M.J.,

H.H., and S.E.C. Mouse colony maintenance and breeding were performed by S.E., A.P., Q.Z., M.P., and S.E.C. *In vivo* castration and tumor grafting experiments were performed by S.E., J.C.T., A.J.T. A.P., E.Y., and J.G. carried out all the bioinformatics analyses. Y.Z. analyzed the single-cell RNA/ATAC-seq datasets. Mouse histological evaluations were conducted by R. Mannan., S.M., and R. Mehra; N.H., R.J.R., M.G., J.L. and M.C. helped carry out the genomic and transcriptomic analyses of patient cohorts. X.C., R.W., and F.S. generated next-generation sequencing libraries and carried out the sequencing. J.L. carried out the analyses on a Chinese cohort of prostate cancer patients. S.E. and A.P. wrote the manuscript and developed the figures with feedback from A.M.C.; S.E. wrote the methods section. In addition, all authors discussed the results, provided feedback, and reviewed the final manuscript.

Chapter 4

Discussion

AR activity is reprogrammed to enable malignant phenotypes during prostate tumorigenesis

At 10 weeks of gestation in humans or 17.5 days post coitum in mice, budding of the prostate gland is triggered by a surge in testosterone in the urogenital sinus [1]. The expression of AR in the stromal and epithelial compartments, as well as paracrine signaling between these compartments, is crucial in determining the fate of prostate cells. This positions AR as a master regulator of the prostatic lineage. Notably, AR has context-dependent roles where it promotes proliferation during development but drives terminal differentiation and inhibits proliferation in the adult gland. Why then do we see such significant therapeutic benefit from inhibiting AR in prostate cancer patients? The answer lies in how AR function is altered in cancer. Through development and adult glandular homeostasis, AR activity proceeds largely unchallenged. However, during clonal evolution of prostate cancer, genetic events that facilitate the survival and proliferative components of AR activity in luminal epithelial cells are selectively favored. In this thesis, we elucidate the mechanisms underlying the cancer-specific reprogramming of AR and provide evidence from both human and murine models that support this salient feature.

Previous studies have shown that upon transformation, AR's cistrome undergoes a 2-3 fold expansion. While pioneer factors like FOXA1 and HOXB13 have been identified as key mediators of this reprogramming, both of these proteins are also expressed in normal prostate cells [2,3]. This suggests that additional cancer-specific cofactors cooperate with lineage-specifying factors

to reprogram AR and activate its proliferative gene program. Through a CRISPR screening approach, we identified NSD2, an H3K36 dimethyltransferase, as a cancer-specific subunit of AR[4]. Unlike FOXA1 and HOXB13, NSD2 is exclusively expressed in cancer cells. Here, NSD2, in collaboration with FOXA1, redirects AR away from its canonical binding sites to its cancer-specific sites observed in patient tumors. Notably, this phenomenon was reproduced in our genetically engineered mouse models, where overexpression of FOXA1 Class 1 mutant reprograms AR to bind its cancer-specific cistrome through aberrantly gained NSD2. Thus, reprogramming of AR is a defining feature of human and mouse prostate tumorigenesis, and the work highlighted in this thesis provides the first functional evidence of a cancer-specific cofactor of AR.

In light of these findings, it is interesting to note that cancer-specific reprogramming of AR doesn't imply amplified AR signaling. In fact, AR, unlike classical oncogenes, is rarely altered in the primary disease [5], and AR signaling is never amplified in the primary disease. However, two independent genomic subtypes that deviate from this norm are FOXA1 Class 1 and SPOP mutations, both of which are primary, truncal alterations. Transcriptomic data from our genetically engineered mouse models, tissue-derived 3D organoids, and human samples revealed that the overexpression of FOXA1 Class 1 mutants amplified AR signaling [6,7]. However, in addition to AR activation, we found that Class 1 mutants also upregulated the PI3K/mTORC1 pathway, independent of mutations in the PI3K pathway. This phenomenon of AR and PI3K co-activation is also observed in SPOP mutant tumors [8]. Prior studies have reported an antagonistic relationship between AR and PI3K signaling [9,10], which likely explains the lack of genomic alterations in the PI3K pathway through PTEN loss or AKT mutations in the primary disease, where AR activity is central to the survival of prostate cells [5,11]. Hence, we believe that activation of AR signaling in these two distinct genomic subtypes in primary PCa likely functions to offset the antagonistic properties of PI3K activation. Together, these findings suggest that FOXA1 Class 1 and SPOP mutants possess the ability to precisely regulate both AR and PI3K/mTORC1 signaling, enabling optimal growth while maintaining the necessary balance in

these pathways.

Targeting cancer-specific epigenetic cofactors to disrupt oncogenic lineage programs

Every cell in the human body has the same underlying genetic sequence encoded in the DNA, and yet, there are about 200 different cell types. Hence, the human body's cellular diversity arises not from differences in DNA sequence but from the differential accessibility of genomic regions, which define which proteins are expressed in each cell type. This accessibility is orchestrated through epigenetic regulation: pioneer transcription factors and associated epigenetic proteins collaborate to mark lineage-specific regulatory elements, remodel chromatin, and deposit chemical modifications, ultimately establishing unique gene expression programs for each lineage.

In the prostate, AR is the nodal master transcription factor (TF) that establishes cellular identity. AR's ability to bind DNA depends on FOXA1, a pioneer factor that opens chromatin at prostate-specific enhancers, enabling AR recruitment [12–14]. This functional dependency is mirrored in breast tissue, where the estrogen receptor (ER) similarly relies on FOXA1 for DNA binding and transcriptional activation [15,16]. Functional activation of these lineage-defining enhancers involves the coordinated action of histone methyltransferases (e.g., EZH2, NSD2, KMT2D, ASH1L), acetyltransferases (e.g., p300/CBP), and acetyl reader proteins (e.g., BRD4), which lay down the chromatin marks necessary for robust transcription, highlighting the critical interplay between lineage TFs and their epigenetic co-factors.

In cancer, this interplay offers substantial therapeutic promise. However, aside from AR, which is druggable due to its ligand-binding pocket, most co-factors or oncogenic TFs, including FOXA1, ERG, and MYC, lack such accessible druggable domains [17], making direct targeting challenging. Therapeutic targeting of AR has improved patient outcomes in PCa, but invariable resistance to anti-androgens mediated by alterations in AR or lineage plasticity—such as transdifferentiation to NEPC—remains a major incurable clinical challenge today [2,18,19]. Epigenetic co-regulators, such as EZH2, BRD4, SWI/SNF, and p300/CBP [10–13], offer novel

drug targets as they harbor domains (SET, HAT, HMG, and bromodomain) accessible to small-molecule inhibition. However, many of these proteins are essential for normal cell survival, and systemic inhibition often leads to unwanted toxicity due to loss of specificity for cancer cells. This clinical experience underscores the need to identify and target epigenetic proteins that are specifically expressed in cancer cells or that function uniquely in collaboration with lineage-defining TFs. This concept is best explored in Chapter 2, where we identified NSD2 as a novel cancer-specific co-factor of AR. In normal prostate cells, H3K36me2 methylation is primarily deposited by NSD1. However, in PCa, NSD2 is aberrantly expressed where it collaborates with FOXA1 to deposit H3K36me2 at oncogenic AR-bound enhancers, driving a cancer-specific gene program. Although NSD1 and NSD2 are functionally similar in their enzymatic capabilities in PCa cells, only NSD2 is able to physically interact with AR through its specialized HMG domain. This molecular specificity suggests that even in the presence of redundant epigenetic machinery, paralogs may not fully compensate in pathological settings if they lack critical interaction domains. Thus, in prostate cancer, targeting NSD2's interaction with AR could selectively disrupt oncogenic programs without compromising normal prostate cell function maintained by NSD1.

With this in mind, a valuable strategy for identifying novel epigenetic drug targets could lie in a functional domain-based CRISPR screening approach [20]. Here, a two-step approach can be used where genes are first prioritized based on their cancer-specific expression. Next, the CRISPR screen can be designed to target functional protein domains rather than the typical 5' coding exons. Systematic identification of cancer-specific epigenetic proteins with specialized domains that mediate direct binding to lineage TFS, like AR or ER, and/or their co-factors will enable the development of more precise epigenetic therapies that overcome systemic toxicities associated with prior epigenetic drugs like the BRD family of proteins.

Paralog co-targeting as a novel approach for cancer therapy

Through human evolution, new genes evolved primarily through the process of gene duplication. This process gave rise to paralogs, which are genes with identical or nearly identical sequences

[21–23]. While most paralogs are functionally redundant, some paralog pairs have acquired foreign genomic sequences through retrotransposition, allowing for divergence in function. Recently, in cancer biology, paralogs have gained significant interest owing to the limited therapeutic benefit derived from targeting singular proteins with functionally redundant partners. To address this, recent studies have conducted CRISPR screens to identify paralogs, thereby developing drugs that target both gene partners or common domains. For example, a recent study found that co-targeting paralog pairs, specifically DUSP4 and DUSP6, in BRAF- and NRAS-mutant melanomas markedly inhibited growth and prevented resistance to MAPK inhibitors [24].

In PCa, we uncovered a similar paralog co-dependency between two H3K36 methyltransferases, NSD1 and NSD2. We found that while NSD2 loss alone disrupts the cancer hallmark features of migration and invasion, and reverses the cancer-specific reprogramming of AR, it is unable to induce cytotoxicity on its own. Additionally, the loss of NSD2 alone did not result in a global decrease in H3K36me2 levels, suggesting the presence of possible compensatory mechanisms. Here, we uncovered that while NSD1 lacks the AR-interacting HMG domain, its catalytic SET-domain function deposits the dominant majority of the H3K36me2 epigenetic mark. Hence, dual loss of NSD1/2 had the most profound impact on global H3K36me2 levels and induced preferential cytotoxicity in AR-positive and AR-dependent prostate cancer cells. These findings also raise an interesting question about the evolutionary relationship between NSD1 and NSD2. During gene duplication, which paralog appeared first? While the NSD family of proteins contains the core methyltransferase SET domain, the PWWP domain, and the PHD zinc finger domains, NSD2 is distinct due to its HMG domain. This domain confers NSD2 with a unique ability to interact with additional proteins, such as AR, enabling certain cancer-specific functions not found in NSD1. Hence, if one were to speculate, it is likely that both NSD1 and NSD2 evolved from a common ancestor through gene duplication, after which NSD2 acquired the novel HMG domain that distinguishes it from other NSD family members, such as NSD1 or NSD3. In summary, paralog co-targeting represents a novel paradigm in cancer therapeutics, particularly for epigenetic proteins, where overlapping functions result from shared structural domains.

Distinct classes of FOXA1 mutations activate divergent oncogenic programs

A central theme emerging from this study is the concept of divergent oncogenic programs driven by distinct classes of FOXA1 mutations. While it is relatively uncommon to find "divergent" oncogenic programs in the same gene — GATA3 being a likely exception — "contrasting" phenotypic outcomes arising from gene alterations are more frequently observed. For example, wild-type TP53 has tumor-suppressive functions, whereas mutations that alter its contact or structural DNA-binding properties confer the protein with dominant-negative oncogenic properties [25]. However, in our study, we report two distinct classes of gain-of-function FOXA1 mutations that diverge in both clinical incidence and functional disease phenotypes. Using genetically engineered mice, we model FOXA1 Class 1 mutants, which are drivers of the primary disease, and Class 2 mutations, which are acquired in the advanced, metastatic disease. We find that while Class 1 mutations drive the development of high-grade AR-positive adenocarcinoma, Class 2 mutations are unable to initiate disease formation. This is consistent with their genomic acquisition in the later stages of the disease. However, class 2 mutants reprogram terminally differentiated AR-positive cells to acquire a stem-like progenitor state that is resistant to androgen deprivation. This makes class 2 mutant-expressing tissues less susceptible to castration-mediated cellular atrophy, and likely suggests that rather than initiating disease, class 2 mutants contribute to driving therapy resistance. Our findings provide the first functional, causal evidence linking FOXA1 mutations to the development of AR-dependent PCa and progression. Additionally, the mouse models developed in this study serve as a valuable bioresource for studying the pathobiology of breast, salivary gland, and bladder tumors harboring oncogenic FOXA1 alterations.

Future work

A significant limitation in the field is the paucity of relevant AR-dependent PCa cell lines. Additionally, models that recapitulate reliance on androgen signaling in the primary disease have been challenging to generate. For example, the most widely used PTEN inactivation-driven PCa mouse models typically drive high-grade, invasive prostate cancers that exhibit castration-resistant

phenotypes [26], thus representing a more advanced disease. Given that FOXA1-Class 1 tumors are castration-sensitive *in vivo*, we have established syngeneic 3D organoids and 2D cell lines derived from these tumors. Similar to the primary model, these cell lines exhibit key features of the human disease, such as NKX3.1 loss, AR expression, and dependence, making them suitable for studying AR biology and testing novel therapeutics in a syngeneic context [6].

Metastatic cancers reactivate features of their embryonic past

Proliferation, invasion, and migration are key hallmarks of cancer that mirror essential processes in embryonic development and organogenesis [27]. For example, during gastrulation, cells undergo an epithelial-to-mesenchymal transition to form the endoderm, mesoderm, and ectoderm germ layers [28]. Later in cancer, the same process, EMT, forms the basis of metastasis, where epithelial cells lose their polarity and migrate, often marking incurable stages of the disease. Recent studies linking developmental stages to human tumors have classified cancers into three broad groups [29,30]. Aggressive and poorly differentiated tumors, such as lung and colon cancer, tend to align with early developmental (Group 1) signatures, while slower-growing tumors like prostate and ovarian cancer show similarities to neonatal or late-stage developmental programs (Group 3).

A notable feature of embryonic development is aberrant proliferation, which is later restricted as differentiation and organ specification occur. In the prostate gland, AR demonstrates a biphasic effect. Early in development, stromal AR is necessary for the proliferation of the urogenital epithelium [31–33]. However, in the adult prostate, AR acts primarily as a driver of terminal differentiation, promoting anti-proliferative effects. This dichotomy raises a critical question: How do prostate cancers reinstate a proliferative state in such a differentiated context? Lineage tracing and tumor evolution studies in genetically engineered mouse models have revealed that AR+ luminal prostate cancers can dedifferentiate to adopt more basal and stem-like, proliferative gene programs, which confer aggressive properties, including therapy resistance [34,35]. The mechanisms governing this plasticity are not fully understood, but cancer may reactivate developmental proteins to initiate these processes. For instance, although aberrant NSD2

expression in tumors drives AR reprogramming at chimeric AR:FOXA1 sites and contributes to metastasis, it remains unclear whether NSD2 similarly regulates AR in the developing prostate, where AR drives proliferation. Thus, it is worth re-examining tumorigenesis as a process in which tumors do not simply activate foreign gene signatures, but rather re-engage developmental programs inherent to their lineage.

Cellular plasticity is recognized as a common feature in diverse cancers. In prostate cancer, plasticity can manifest as transdifferentiation into a different cell fate (or lineage plasticity, exemplified by the emergence of NEPC), or as dedifferentiation into a more precursor cell state within the luminal lineage, a phenomenon we refer to as "intra-luminal plasticity" [6,19,36]. While lineage plasticity into NEPC often enables escape from targeted therapies, intra-luminal plasticity provides another axis for therapy resistance. This phenomenon is best demonstrated in mouse prostate models expressing FOXA1 Class 2 mutations, which are clinically found in metastatic disease [7]. Expression of the Class 2 mutant in normal prostate tissue does not drive tumorigenesis; rather, it reprograms differentiated AR+ cells into a more stem-like progenitor state. These cells activate transcriptional programs governed by stemness-associated factors such as KLF5 and SOX9. Importantly, these progenitor-like cells retain AR positivity but exhibit increased resistance to androgen deprivation. This enables the survival of Class 2 mutant-expressing cells even in the absence of androgen. While the process of de-differentiation is critical in highly differentiated tumor types, it is worth questioning whether these selective pressures similarly shape cancers that are, by nature, poorly differentiated from the outset. An intriguing idea mirroring this theme was proposed by Lloyd J. Old. He described cancer as a "somatic cell pregnancy", owing to activation of gametogenic programs in cancer cells [37]. Therapeutically, uncovering conserved features of this "rebirth" may offer new avenues for intervention across various cancer types.

In summary, this thesis focuses on uncovering the functional consequences of FOXA1 mutations in prostate cancer and identifying cancer-specific epigenetic regulators that collaborate with FOXA1 to drive oncogenic transcriptional programs. In Chapter 2, we characterize NSD2, a

histone methyltransferase aberrantly expressed in prostate cancer, as a novel cancer-specific cofactor of AR. We demonstrate that NSD2 collaborates with FOXA1 to reprogram AR at chimeric AR:FOXA1 sites, thereby activating its oncogenic transcriptional program. Given the high incidence and diversity of FOXA1 mutations in prostate cancer, Chapter 3 presents the development of the first genetically engineered mouse models harboring class-specific FOXA1 mutations. Using these models, we provide direct causal evidence linking distinct FOXA1 mutations to driving prostate tumorigenesis or therapy resistance.

Collectively, this work reveals how mutations in key pioneer transcription factors collaborate with epigenetic cofactors to drive disease. By dissecting the molecular interplay between FOXA1 and cancer-specific chromatin regulators, such as NSD2, we establish a framework for understanding lineage-constrained tumor evolution and identifying new therapeutic vulnerabilities that may ultimately extend beyond prostate cancer.

Chapter acknowledgements:

This chapter was written by Sanjana Eyunni. The concepts highlighted here summarize the relevant findings presented in this thesis and are inspired by discussions with Abhijit Parolia.

References:

1. Toivanen R, Shen MM. Prostate organogenesis: tissue induction, hormonal regulation and cell type specification. Development. 2017;144: 1382–1398.

2. Mills IG. Maintaining and reprogramming genomic androgen receptor activity in prostate cancer. Nat Rev Cancer. 2014;14: 187–198.

3. 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.

4. Parolia A, Eyunni S, Verma BK, Young E, Liu Y, Liu L, et al. NSD2 is a requisite subunit of the AR/FOXA1 neo-enhanceosome in promoting prostate tumorigenesis. Nat Genet. 2024; 1–12.

5. Cancer Genome Atlas Research Network. The Molecular Taxonomy of Primary Prostate Cancer. Cell. 2015;163: 1011–1025.

6. Eyunni S, Mannan R, Zhang Y, Young E, Zhang Q, Luo J, et al. Divergent FOXA1 mutations drive prostate tumorigenesis and therapy-resistant cellular plasticity. Science. 2025; eadv2367.

7. 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.

8. Blattner M, Liu D, Robinson BD, Huang D, Poliakov A, Gao D, et al. SPOP Mutation Drives Prostate Tumorigenesis In Vivo through Coordinate Regulation of PI3K/mTOR and AR Signaling. Cancer Cell. 2017;31: 436–451.

9. Qi W, Morales C, Cooke LS, Johnson B, Somer B, Mahadevan D. Reciprocal feedback inhibition of the androgen receptor and PI3K as a novel therapy for castrate-sensitive and -resistant prostate cancer. Oncotarget. 2015;6: 41976–41987.

10. Carver BS, Chapinski C, Wongvipat J, Hieronymus H, Chen Y, Chandarlapaty S, et al. Reciprocal feedback regulation of PI3K and androgen receptor signaling in PTEN-deficient prostate cancer. Cancer Cell. 2011;19: 575–586.

11. 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.

12. 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.

13. Hankey W, Chen Z, Wang Q. Shaping chromatin states in prostate cancer by pioneer

transcription factors. Cancer Res. 2020;80: 2427–2436.

14. The Role of Hepatocyte Nuclear Factor-3α (Forkhead Box A1) and Androgen Receptor in Transcriptional Regulation of Prostatic Genes.

15. Seachrist DD, Anstine LJ, Keri RA. FOXA1: A pioneer of nuclear receptor action in breast cancer. Cancers (Basel). 2021;13: 5205.

16. Arruabarrena-Aristorena A, Maag JLV, Kittane S, Cai Y, Karthaus WR, Ladewig E, et al. FOXA1 Mutations Reveal Distinct Chromatin Profiles and Influence Therapeutic Response in Breast Cancer. Cancer Cell. 2020. doi:10.1016/j.ccell.2020.08.003

17. Arrowsmith CH, Bountra C, Fish PV, Lee K, Schapira M. Epigenetic protein families: a new frontier for drug discovery. Nat Rev Drug Discov. 2012;11: 384–400.

18. Beltran H, Tomlins S, Aparicio A, Arora V, Rickman D, Ayala G, et al. Aggressive variants of castration-resistant prostate cancer. Clin Cancer Res. 2014;20: 2846–2850.

19. Tang DG. Understanding and targeting prostate cancer cell heterogeneity and plasticity. Semin Cancer Biol. 2022;82: 68–93.

20. Shi J, Wang E, Milazzo JP, Wang Z, Kinney JB, Vakoc CR. Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. Nat Biotechnol. 2015;33: 661–667.

21. Iohannes SD, Jackson D. Tackling redundancy: genetic mechanisms underlying paralog compensation in plants. New Phytol. 2023;240: 1381–1389.

22. Tekaia F, Dujon B. Pervasiveness of gene conservation and persistence of duplicates in cellular genomes. J Mol Evol. 1999;49: 591–600.

23. Ryan CJ, Mehta I, Kebabci N, Adams DJ. Targeting synthetic lethal paralogs in cancer. Trends Cancer. 2023;9: 397–409.

24. Ito T, Young MJ, Li R, Jain S, Wernitznig A, Krill-Burger JM, et al. Paralog knockout profiling identifies DUSP4 and DUSP6 as a digenic dependence in MAPK pathway-driven cancers. Nat Genet. 2021;53: 1664–1672.

25. Wang X, Yang J, Yang W, Sheng H, Jia B, Cheng P, et al. Multiple roles of p53 in cancer development: Regulation of tumor microenvironment, m6A modification and diverse cell death mechanisms. J Adv Res. 2024 [cited 24 Jul 2025]. doi:10.1016/j.jare.2024.10.026

26. Wang S, Gao J, Lei Q, Rozengurt N, Pritchard C, Jiao J, et al. Prostate-specific deletion of the murine Pten tumor suppressor gene leads to metastatic prostate cancer. Cancer Cell. 2003;4: 209–221.

27. Hanahan D. Hallmarks of cancer: New dimensions. Cancer Discov. 2022;12: 31–46.

28. Aiello NM, Stanger BZ. Echoes of the embryo: using the developmental biology toolkit to study cancer. Dis Model Mech. 2016;9: 105–114.

29. Naxerova K, Bult CJ, Peaston A, Fancher K, Knowles BB, Kasif S, et al. Analysis of gene expression in a developmental context emphasizes distinct biological leitmotifs in human cancers. Genome Biol. 2008;9: R108.

30. Soares C. Cancer Clues from Embryonic Development. In: Scientific American [Internet]. 1 May 2009 [cited 21 Jul 2025]. Available: https://www.scientificamerican.com/article/cancer-clues-from-embryos/

31. Lai K-P, Yamashita S, Huang C-K, Yeh S, Chang C. Loss of stromal androgen receptor leads to suppressed prostate tumourigenesis via modulation of pro-inflammatory cytokines/chemokines. EMBO Mol Med. 2012;4: 791–807.

32. Simanainen U, Allan CM, Lim P, McPherson S, Jimenez M, Zajac JD, et al. Disruption of prostate epithelial androgen receptor impedes prostate lobe-specific growth and function.

Endocrinology. 2007;148: 2264–2272.

33. Wen S, Chang H-C, Tian J, Shang Z, Niu Y, Chang C. Stromal androgen receptor roles in the development of normal prostate, benign prostate hyperplasia, and prostate cancer. Am J Pathol. 2015;185: 293–301.

34. Feng W, Ladewig E, Salsabeel N, Zhao H, Lee YS, Gopalan A, et al. ERG activates a stem-like proliferation-differentiation program in prostate epithelial cells with mixed basal-luminal identity. bioRxivorg. 2024. doi:10.1101/2023.05.15.540839

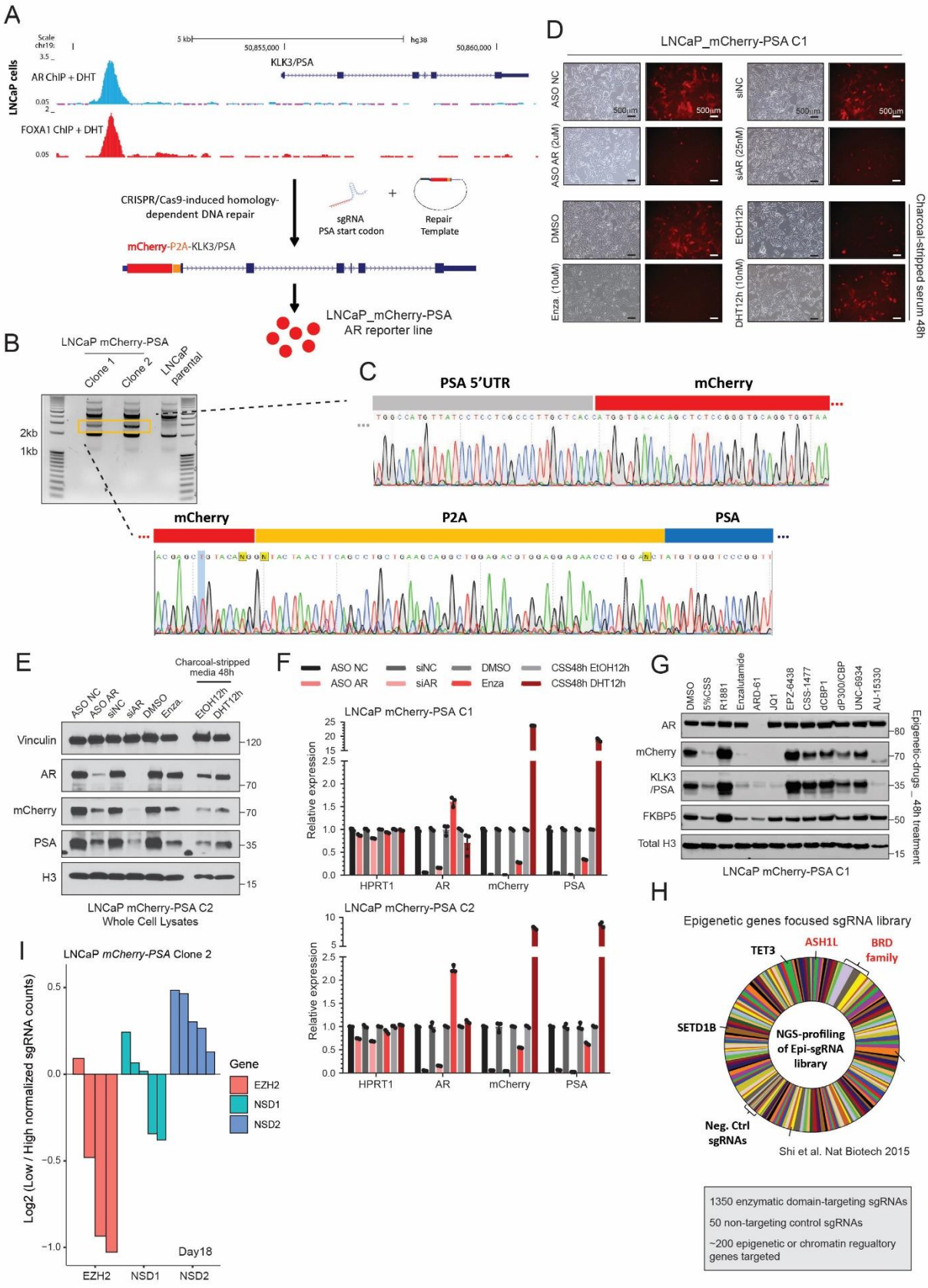
35. Jiang C, Song Y, Rorive S, Allard J, Tika E, Zahedi Z, et al. Innate immunity and the NF-κB pathway control prostate stem cell plasticity, reprogramming and tumor initiation. Nat Cancer. 2025. doi:10.1038/s43018-025-00994-3

36. Yamada Y, Beltran H. Clinical and biological features of neuroendocrine prostate cancer. Curr Oncol Rep. 2021;23: 15.

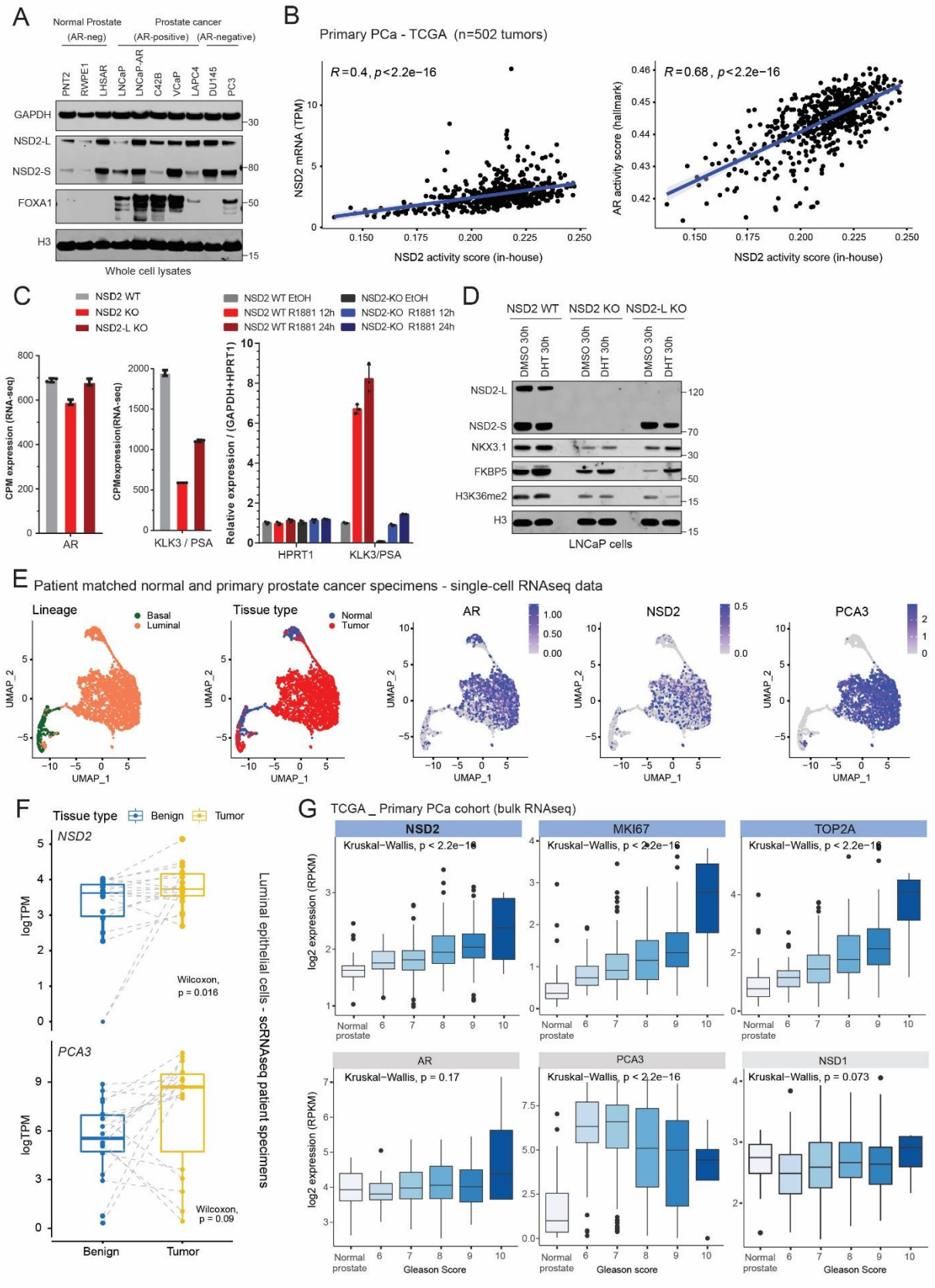
37. Old LJ. Cancer is a somatic cell pregnancy. Cancer Immun. 2007;7: 19.

Appendix A

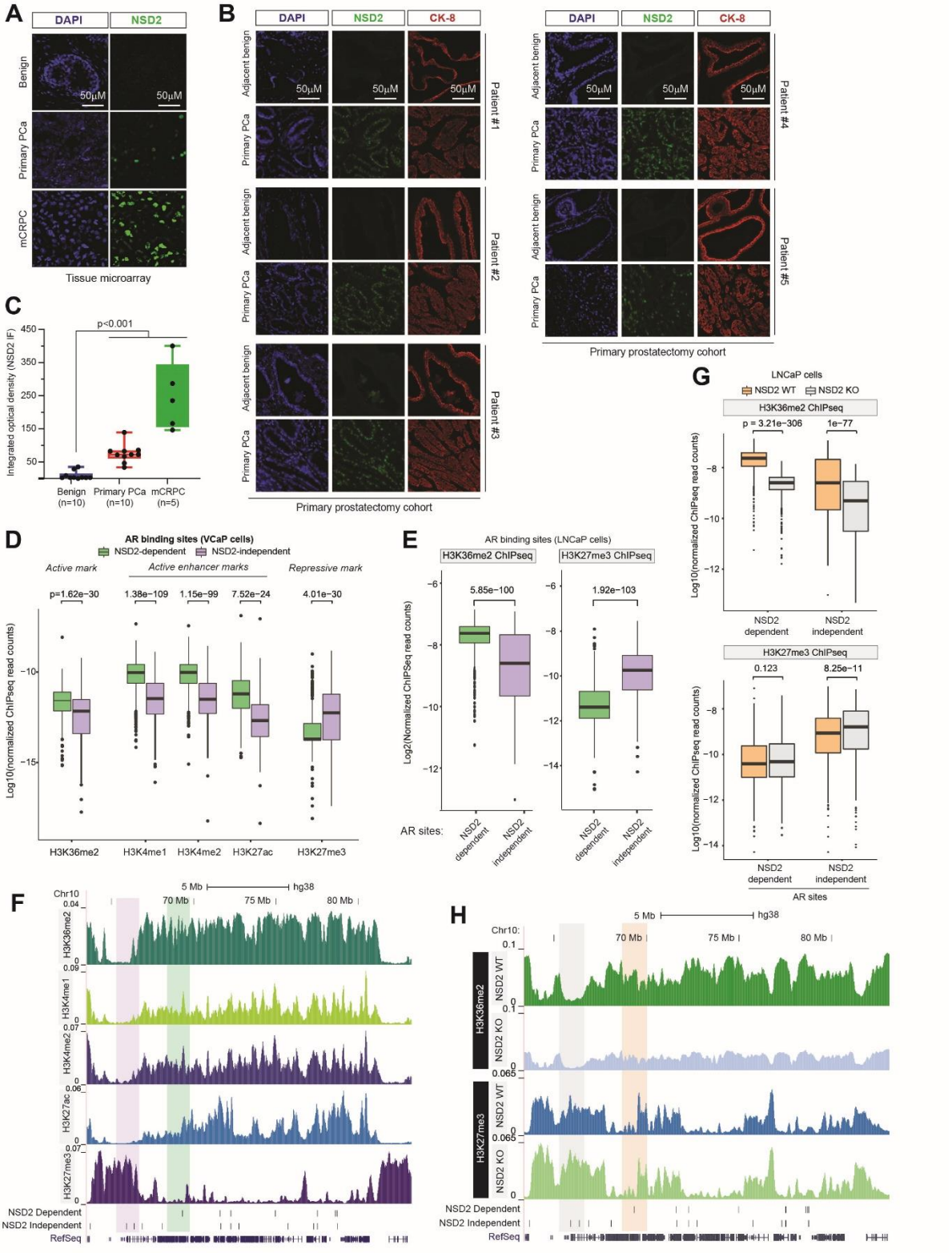
Supplementary Figures for Chapter 2


A
LNCaP cells
AR ChIP + DHT
FOXA1 ChIP + DHT
KLK3/PSA
CRISPR/Cas9-induced homology-dependent DNA repair
sgRNA PSA start codon
Repair Template
mCherry-P2A-KLK3/PSA
LNCaP_mCherry-PSA AR reporter line
B
LNCaP mCherry-PSA
Clone 1
Clone 2
LNCaP parental
2kb
1kb
C
PSA 5'UTR
mCherry
mCherry
P2A
PSA
D
LNCaP_mCherry-PSA C1
ASO NC
ASO AR (2uM)
siNC
siAR (25nM)
DMSO
Enza. (10uM)
EtOH12h
DHT12h (10nM)
Charcoal-stripped serum 48h
E
Charcoal-stripped media 48h
ASO NC
ASO AR
siNC
siAR
DMSO
Enza.
EtOH12h
DHT12h
Vinculin
AR
mCherry
PSA
H3
LNCaP mCherry-PSA C2
Whole Cell Lysates
F
ASO NC
ASO AR
siNC
siAR
DMSO
Enza
CSS48h EtOH12h
CSS48h DHT12h
LNCaP mCherry-PSA C1
Relative expression
HPRT1
AR
mCherry
PSA
LNCaP mCherry-PSA C2
G
DMSO
5%CSS
R1881
Enzalutamide
ARD-61
JQ1
EPZ-6438
CSS-1477
dCBP1
dP300/CBP
UNC-6934
AU-15330
AR
mCherry
KLK3 /PSA
FKBP5
Total H3
Epigenetic-drugs _ 48h treatment
LNCaP mCherry-PSA C1
H
Epigenetic genes focused sgRNA library
TET3
ASH1L
BRD family
SETD1B
NGS-profiling of Epi-sgRNA library
Neg. Ctrl sgRNAs
Shi et al. Nat Biotech 2015
I
LNCaP mCherry-PSA Clone 2
Log2 (Low / High normalized sgRNA counts)
Gene
EZH2
NSD1
NSD2
Day18
EZH2
NSD1
NSD2
1350 enzymatic domain-targeting sgRNAs
50 non-targeting control sgRNAs
~200 epigenetic or chromatin regualtory genes targeted

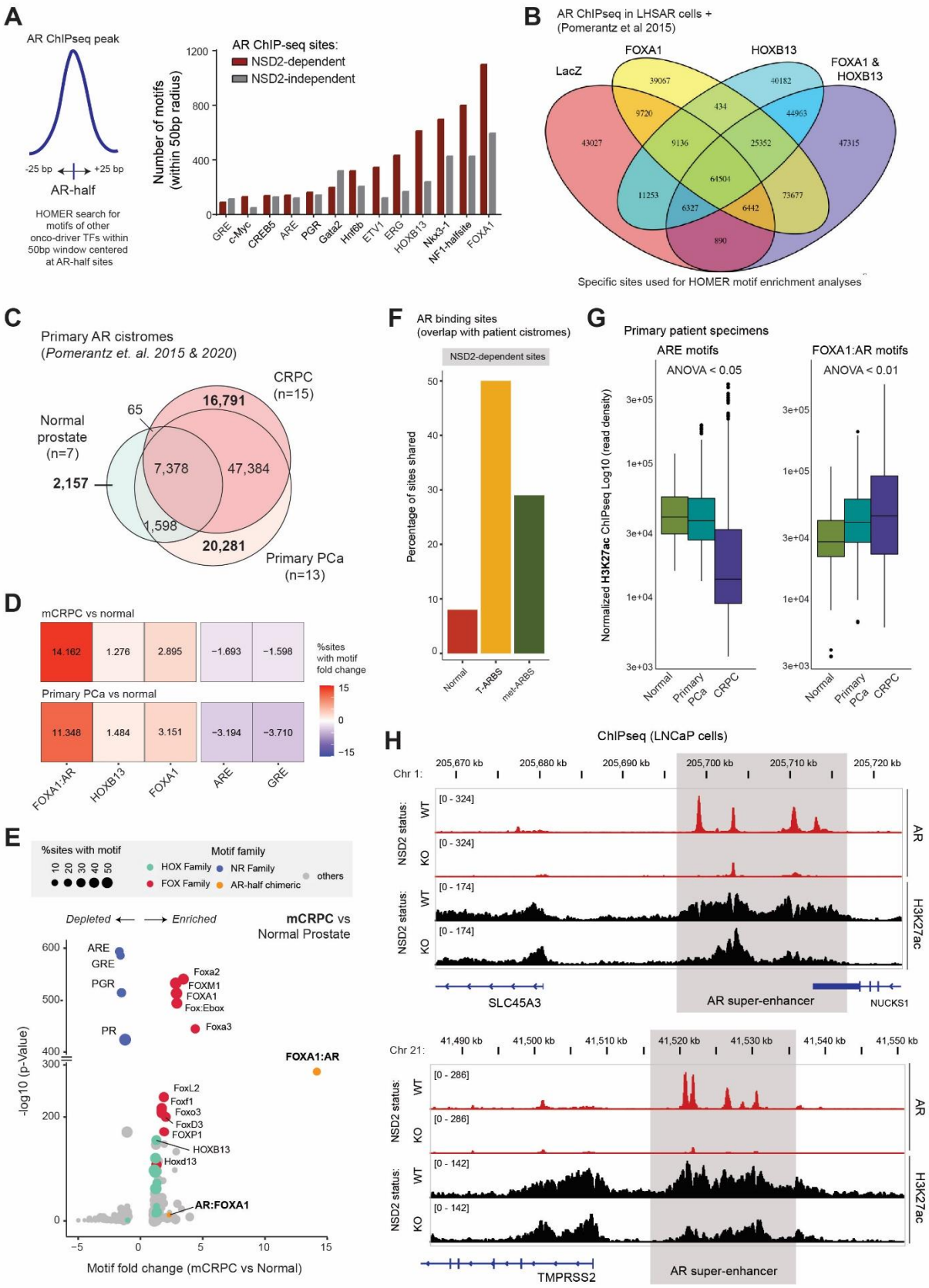
Supplementary Figure A2-1: Generation and characterization of the endogenous mCherry-PSA AR reporter cell lines. A) Schematic representation of the workflow of LNCaP-mCherry-PSA AR reporter cell line generation. B) DNA gel electrophoresis image showing the exogenously inserted mCherry amplicon in the LNCaP-mCherry-PSA lines. Clones 1 and 2 were used for the functional CRISPR screen. C) Sanger sequencing chromatograms of the PCR amplicon from reporter cells in panel (b) showing the KLK3/PSA gene promoter and exon 1 start codon junctions. D) Representative brightfield and mCherry immunofluorescence images of the LNCaP-mCherry-PSA clone 1 treated with (top) AR-targeting siRNA or antisense oligonucleotides (ASOs) (siAR and ASO AR respectively) or enzalutamide (bottom left). Reporter cells were also serum starved for 48h and stimulated with DHT (10nM for 12h) to showcase gain in signal (bottom right). All treatments were repeated at least twice. Scale bar: 500µm. E) Immunoblots of noted proteins in LNCaP reporter cells as in panel (d). F) Expression (qPCR) of noted genes in reporter monoclones treated as in panel (d) to manipulate AR signaling (n = 3 biological replicates). Mean +/- SEM is shown. G) Immunoblots of noted proteins, including the exogenously introduced mCherry protein, in LNCaP reporter cells treated with AR-targeting epigenetic drugs. Total H3 is used as a loading control. H) Next-generation sequencing-based abundance of sgRNAs in the epigenetic-focused library used in the CRISPR screen highlighting some of the known epigenetic regulators of AR. I) Individual NSD1, NSD2, or EZH2-targeting sgRNA ratios in mCherry-LOW to mCherry-HIGH cells in the CRISPR screen.


A
Normal Prostate
Prostate cancer
(AR-neg)
(AR-positive)
(AR-negative)
PNT2
RWPE1
LHSAR
LNCaP
LNCaP-AR
C42B
VCaP
LAPC4
DU145
PC3
GAPDH
NSD2-L
NSD2-S
FOXA1
H3
Whole cell lysates
B
Primary PCa - TCGA (n=502 tumors)
R=0.4, p<2.2e−16
NSD2 mRNA (TPM)
NSD2 activity score (in-house)
R=0.68, p<2.2e−16
AR activity score (hallmark)
NSD2 activity score (in-house)
C
NSD2 WT
NSD2 KO
NSD2-L KO
NSD2 WT EtOH
NSD2 WT R1881 12h
NSD2 WT R1881 24h
NSD2-KO EtOH
NSD2-KO R1881 12h
NSD2-KO R1881 24h
CPM expression (RNA-seq)
AR
CPMexpression(RNA-seq)
KLK3 / PSA
Relative expression / (GAPDH+HPRT1)
HPRT1
KLK3/PSA
D
NSD2 WT
NSD2 KO
NSD2-L KO
DMSO 30h
DHT 30h
NSD2-L
NSD2-S
NKX3.1
FKBP5
H3K36me2
H3
LNCaP cells
E
Patient matched normal and primary prostate cancer specimens - single-cell RNAseq data
Lineage
Basal
Luminal
Tissue type
Normal
Tumor
AR
NSD2
PCA3
UMAP_1
UMAP_2
F
Tissue type
Benign
Tumor
NSD2
PCA3
logTPM
Wilcoxon, p = 0.016
Wilcoxon, p = 0.09
Benign
Tumor
Luminal epithelial cells - scRNAseq patient specimens
G
TCGA _ Primary PCa cohort (bulk RNAseq)
NSD2
MKI67
TOP2A
AR
PCA3
NSD1
Kruskal−Wallis, p < 2.2e−16
Kruskal−Wallis, p = 0.17
Kruskal−Wallis, p = 0.073
log2 expression (RPKM)
Normal prostate
Gleason Score
Gleason score

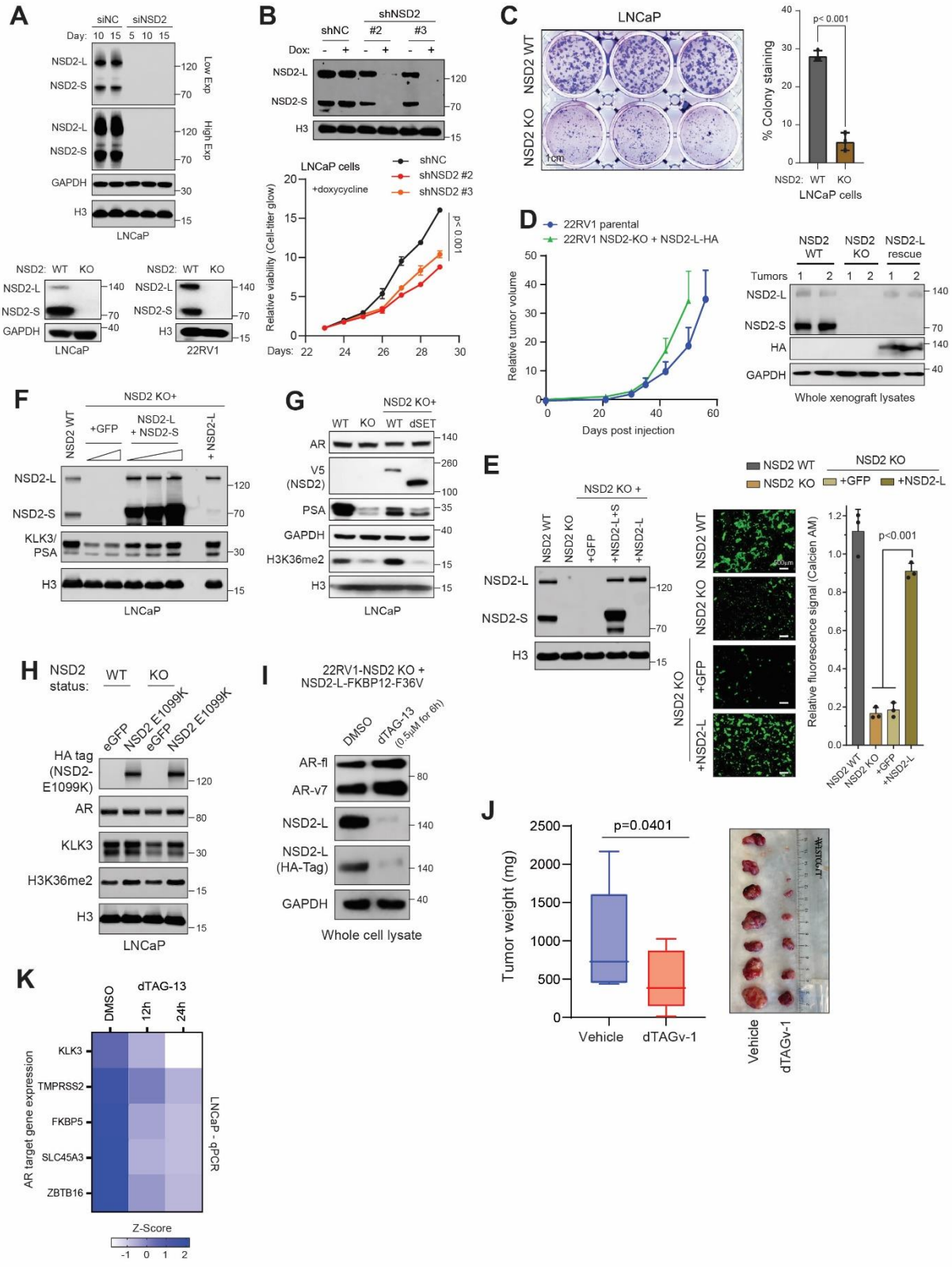
Supplementary Figure A2-2: NSD2 transcript and protein expression in primary patient specimens. A) Immunoblot of labeled proteins in a collection of AR-positive and AR-negative prostate cell lines. GAPDH and H3 are used as a loading control. B) ***Left:*** Correlation plots showing the *NSD2* transcript expression and gene signature-based "NSD2 activity score" in primary prostate cancers from the TCGA cohort (n=502 tumors). ***Right:*** Correlation plots showing NSD2 activity score and the widely-used hallmark AR activity score in primary PCa tumors. (Pearson's linear correlation coefficient, permutation test). Line, mean; shaded region, SEM. C) Relative expression (qPCR or RNAseq) of *AR* and *KLK3* transcripts in CRISPR-edited NSD2-KO or NSD2-L-KO LNCaP cells (***left;*** n = 2 biological replicates) or NSD2 CRISPR-edited cells stimulated with R1881 for 12 or 24 hours (***right*;** n = 3 biological replicates). *HPRT1* is used as a loading control. Mean +/- SEM is shown. D) Immunoblot of labeled proteins in LNCaP NSD2 WT and KO cells stimulated with DHT for 30 hours. E) UMAP plots from patient-matched normal and primary prostate cancer single-cell RNA-seq data. F) *NSD2* and *PCA3* transcript expression in patient-matched normal and primary prostate cancer luminal epithelial cells (pseudo-bulk analyses from single cell data; n=15 biological replicates, two-sided Wilcoxon test). Box plot center, median; box, quartiles 1-3; whiskers, quartiles 1-3 ± 1.5 × interquartile range; dot are outliers. G) Boxplot showing RNA expression of labeled genes in primary prostate cancer specimens (TCGA cohort) stratified by the Gleason score (normal = 52; Gleason 6 = 46; Gleason 7 = 249, Gleason 8 = 65; Gleason 9 = 138, Gleason 10 = 4 tumor specimens. One-way ANOVA and Kruskal-Wallis test). Box plot center, median; box, quartiles 1-3; whiskers, quartiles 1-3 ± 1.5 × interquartile range; dot, outliers.


A
DAPI
NSD2
Benign
Primary PCa
mCRPC
50μM
Tissue microarray
B
DAPI
NSD2
CK-8
Adjacent benign
Primary PCa
Patient #1
Patient #2
Patient #3
Patient #4
Patient #5
Primary prostatectomy cohort
C
p<0.001
Integrated optical density (NSD2 IF)
Benign (n=10)
Primary PCa (n=10)
mCRPC (n=5)
D
AR binding sites (VCaP cells)
NSD2-dependent
NSD2-independent
Active mark
Active enhancer marks
Repressive mark
p=1.62e−30
1.38e−109
1.15e−99
7.52e−24
4.01e−30
Log10(normalized ChIPseq read counts)
H3K36me2
H3K4me1
H3K4me2
H3K27ac
H3K27me3
E
AR binding sites (LNCaP cells)
H3K36me2 ChIPseq
H3K27me3 ChIPseq
5.85e−100
1.92e−103
Log2(Normalized ChIPSeq read counts)
AR sites:
NSD2 dependent
NSD2 independent
G
LNCaP cells
NSD2 WT
NSD2 KO
H3K36me2 ChIPseq
p = 3.21e−306
1e−77
H3K27me3 ChIPseq
0.123
8.25e−11
Log10(normalized ChIPSeq read counts)
AR sites
F
Chr10
5 Mb
hg38
70 Mb
75 Mb
80 Mb
H3K36me2
H3K4me1
H3K4me2
H3K27ac
H3K27me3
NSD2 Dependent
NSD2 Independent
RefSeq
H
Chr10:
NSD2 WT
NSD2 KO
H3K36me2
H3K27me3

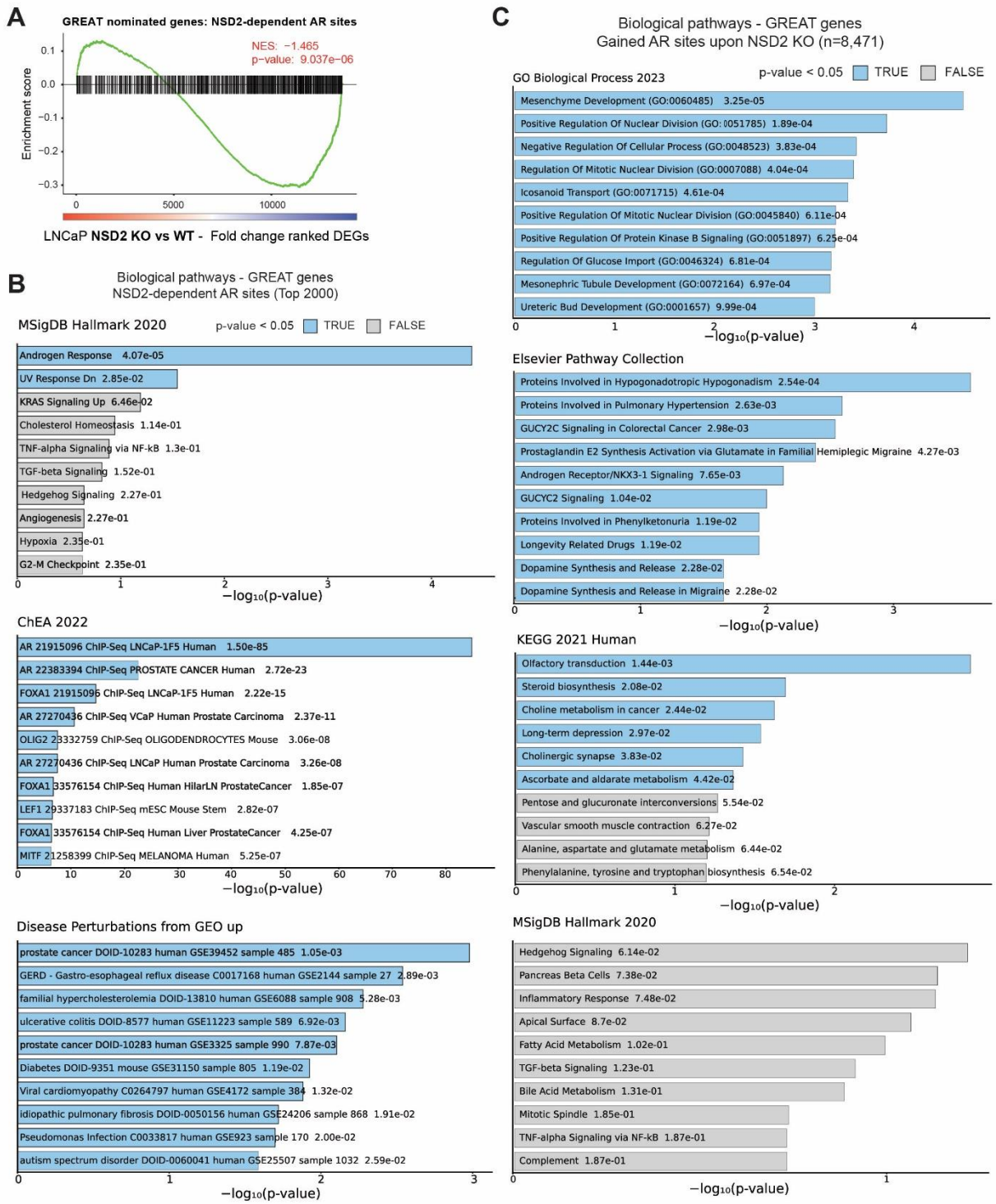
Supplementary Figure A2-3: NSD2 and H3K36me2 expression in patient tumors and prostate cancer cell lines. A) Representative immunofluorescence (IF) images of NSD2 in benign prostate, primary prostate cancer (PCa), and metastatic CRPC tissue microarray.Scale bar:50μm. B) Representative multiplex IF images of NSD2 and CK-8 in adjacent benign and primary prostate cancer lesions in patient prostatectomies (n=5 biological replicates, Scale bar:50μm). C) Integrated optical density quantification of NSD2 IF staining in benign (n=10), primary PCa (n=10), and mCRPC (n=5) tissues. Box plot center, median; box, quartiles 1-3; whiskers, min and max values. D) Boxplots of normalized ChIP-seq reads of distinct activating and repressive histone modifications at NSD2-dependent and NSD2-independent AR sites in VCaP cells (n = top 2000 sites, two-sided t-test). Box plot center, median; box, quartiles 1-3; whiskers, quartiles 1-3 ± 1.5 × interquartile range; dot, outliers. E) Boxplots of normalized ChIP-seq reads of NSD2-catalyzed H3K36me2 and EZH2/PRC2-catalyzed H3K27me3 histone marks at NSD2-dependent and NSD2-independent AR sites in LNCaP cells (n = top 2000 sites, two-sided t-test). Box plot center, median; box, quartiles 1-3; whiskers, quartiles 1-3 ± 1.5 × interquartile range; dot, outliers. F) ChIP-seq read-density tracks of histone modification within a Chr10 locus in VCaP cells. NSD2-dependent and independent AR sites are marked in the tracks below with representative enhancers highlighted. G) ChIP-seq read density boxplots showing H3K36m2 (*top*) and H3K27me3 (*bottom*) signals at AR sites in NSD2 KO or WT LNCaP cells (n = top 2000 sites, two-sided t-test). Box plot center, median; box, quartiles 1-3; whiskers, quartiles 1-3 ± 1.5 × interquartile range; dot, outliers. H) ChIP-seq read-density tracks of H3K36me2 and H3K27ac within a Chr10 locus in NSD2 WT and KO LNCaP cell lines. NSD2-dependent and independent AR sites are marked and highlighted.


A
AR ChIPseq peak
-25 bp
+25 bp
AR-half
HOMER search for motifs of other onco-driver TFs within 50bp window centered at AR-half sites
AR ChIP-seq sites:
NSD2-dependent
NSD2-independent
Number of motifs (within 50bp radius)
GRE
c-Myc
CREB5
ARE
PGR
Gata2
Hnf6b
ETV1
ERG
HOXB13
Nkx3-1
NF1-halfsite
FOXA1
B
AR ChIPseq in LHSAR cells + (Pomerantz et al 2015)
LacZ
FOXA1
HOXB13
FOXA1 & HOXB13
39067
40182
9720
434
44963
43027
9136
25352
47315
64504
11253
6327
6442
73677
890
Specific sites used for HOMER motif enrichment analyses
C
Primary AR cistromes
(Pomerantz et. al. 2015 & 2020)
CRPC (n=15)
Normal prostate (n=7)
Primary PCa (n=13)
65
16,791
7,378
47,384
2,157
1,598
20,281
D
mCRPC vs normal
14.162
1.276
2.895
−1.693
−1.598
Primary PCa vs normal
11.348
1.484
3.151
−3.194
−3.710
%sites with motif fold change
15
0
−15
FOXA1:AR
HOXB13
FOXA1
ARE
GRE
E
%sites with motif
Motif family
HOX Family
NR Family
others
FOX Family
AR-half chimeric
Depleted
Enriched
mCRPC vs Normal Prostate
-log10 (p-Value)
ARE
GRE
PGR
PR
Foxa2
FOXM1
FOXA1
Fox:Ebox
Foxa3
FOXA1:AR
FoxL2
Foxf1
Foxo3
FoxD3
FOXP1
HOXB13
Hoxd13
AR:FOXA1
Motif fold change (mCRPC vs Normal)
F
AR binding sites
(overlap with patient cistromes)
NSD2-dependent sites
Percentage of sites shared
Normal
T-ARBS
met-ARBS
G
Primary patient specimens
ARE motifs
ANOVA < 0.05
FOXA1:AR motifs
ANOVA < 0.01
Normalized H3K27ac ChIPseq Log10 (read density)
Normal
Primary PCa
CRPC
H
ChIPseq (LNCaP cells)
Chr 1:
205,670 kb
205,680 kb
205,690 kb
205,700 kb
205,710 kb
205,720 kb
NSD2 status:
WT
KO
[0 - 324]
[0 - 174]
AR
H3K27ac
SLC45A3
AR super-enhancer
NUCKS1
Chr 21:
41,490 kb
41,500 kb
41,510 kb
41,520 kb
41,530 kb
41,540 kb
41,550 kb
[0 - 286]
[0 - 142]
TMPRSS2
AR super-enhancer

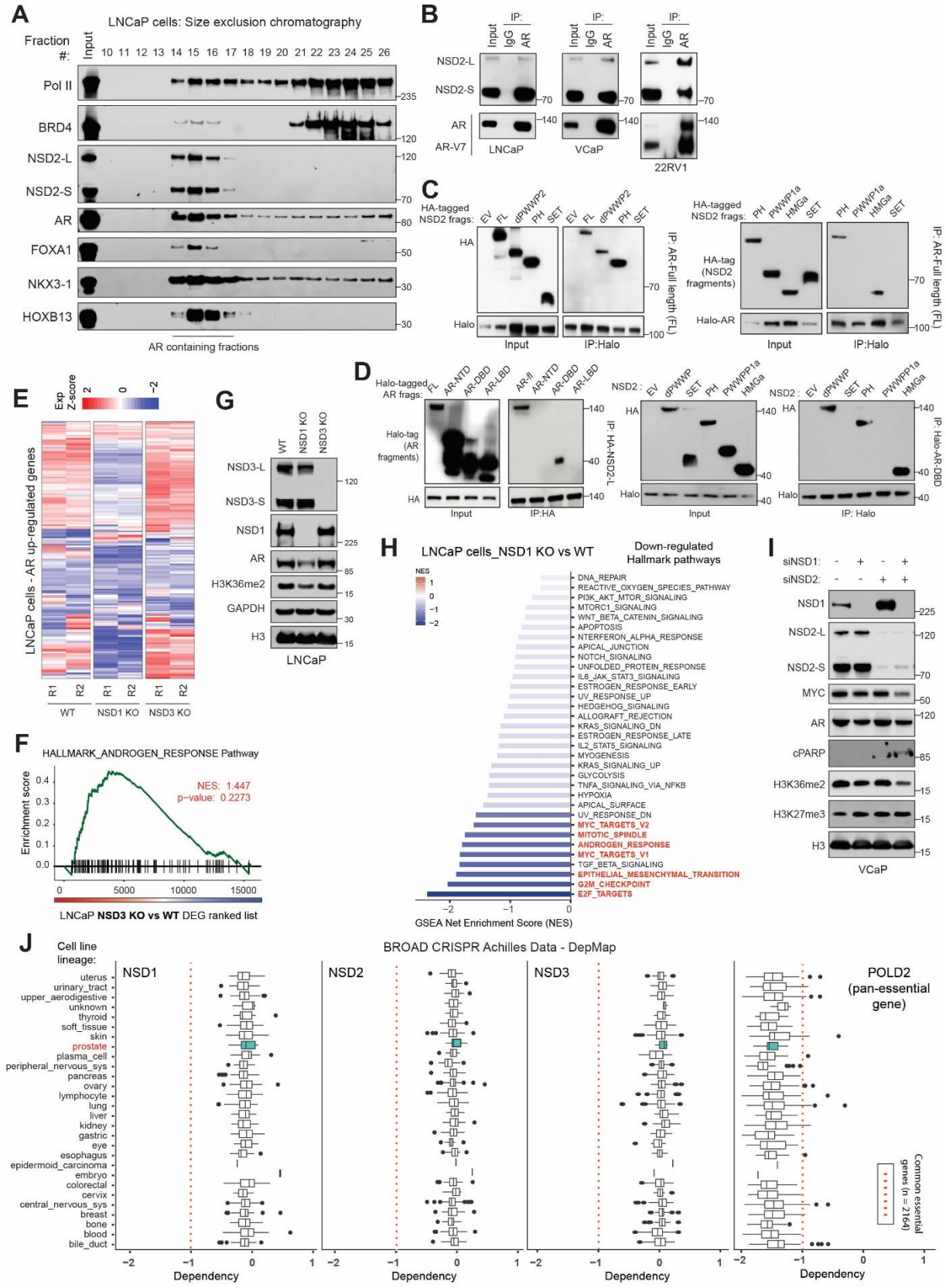
Supplementary Figure A2-4: Motif characterization of the NSD2-enabled AR neo-cistrome in prostate cancer cells. A) *Left*: Schematic representation of the half-motif enrichment analysis. *Right*: Motif enrichment plot of AR-half motifs with neighboring motifs of other transcription factors at NSD2-dependent and independent AR sites in LNCaP cells. **B)** Venn diagram showing overlaps between AR ChIP-seq sites in LHSAR cells with LacZ (control), FOXA1, HOXB13, FOXA1+HOXB13 overexpression. **C)** Venn diagram showing overlap of AR cistromes (ChIP-seq) in normal prostate, primary prostate cancer, and castration-resistant prostate cancer specimens. (Pomerantz et. al. 2015 & 2020) **D)** Motif fold-change heatmap in normal, primary cancer, and castration-resistant prostate cancer specimens. **E)** Fold change and significance of HOMER motifs enriched within mCRPC cancer-specific AR sites over normal tissue-specific AR elements (data from Pomerantz, 2015 and 2020). **F)** Barplot showing percentage of shared sites between the NSD2-dependent AR sites and AR cistromes from the normal prostate, primary PCa (T-ARBS), or metastatic CRPC (met-ARBS) patient tumors. **G)** Boxplot showing H3K27ac ChIP-seq read density at sites containing the ARE or the FOXA1:AR motif in normal and tumor patient samples (normal prostate, n=7; primary prostate cancer, n=13; castration-resistant prostate cancer - CRPC, n=15; one-way ANOVA and Tukey's test). Box plot center, median; box, quartiles 1-3; whiskers, quartiles 1-3 ± 1.5 × interquartile range; dot, outliers. **H)** ChIP-seq read density tracks of AR and H3K27ac within the *SLC45A3* and *TMPRSS2* loci in NSD2 WT and NSD2 KO LNCaP cells. Super-enhancer clusters are highlighted in a gray box.


A
siNC
siNSD2
Day: 10 15 5 10 15
NSD2-L
NSD2-S
Low Exp
High Exp
GAPDH
H3
LNCaP
NSD2: WT KO
22RV1
B
shNSD2
shNC
#2
#3
Dox: - + - + - +
LNCaP cells
shNC
shNSD2 #2
shNSD2 #3
+doxycycline
Relative viability (Cell-titer glow)
Days:
p< 0.001
C
LNCaP
NSD2 WT
NSD2 KO
% Colony staining
p< 0.001
LNCaP cells
D
22RV1 parental
22RV1 NSD2-KO + NSD2-L-HA
Relative tumor volume
Days post injection
NSD2 WT
NSD2 KO
NSD2-L rescue
Tumors
HA
Whole xenograft lysates
E
NSD2 KO +
+GFP
+NSD2-L+S
+NSD2-L
Relative fluorescence signal (Calcien AM)
p<0.001
F
NSD2 KO+
NSD2 WT
+GFP
NSD2-L + NSD2-S
+ NSD2-L
KLK3/ PSA
G
WT KO WT dSET
AR
V5 (NSD2)
PSA
H3K36me2
H
NSD2 status: WT KO
eGFP
NSD2 E1099K
HA tag (NSD2-E1099K)
AR
KLK3
H3K36me2
I
22RV1-NSD2 KO + NSD2-L-FKBP12-F36V
DMSO
dTAG-13 (0.5μM for 6h)
AR-fl
AR-v7
NSD2-L
NSD2-L (HA-Tag)
Whole cell lysate
J
p=0.0401
Tumor weight (mg)
Vehicle
dTAGv-1
K
dTAG-13
DMSO
12h
24h
AR target gene expression
KLK3
TMPRSS2
FKBP5
SLC45A3
ZBTB16
LNCaP - qPCR
Z-Score

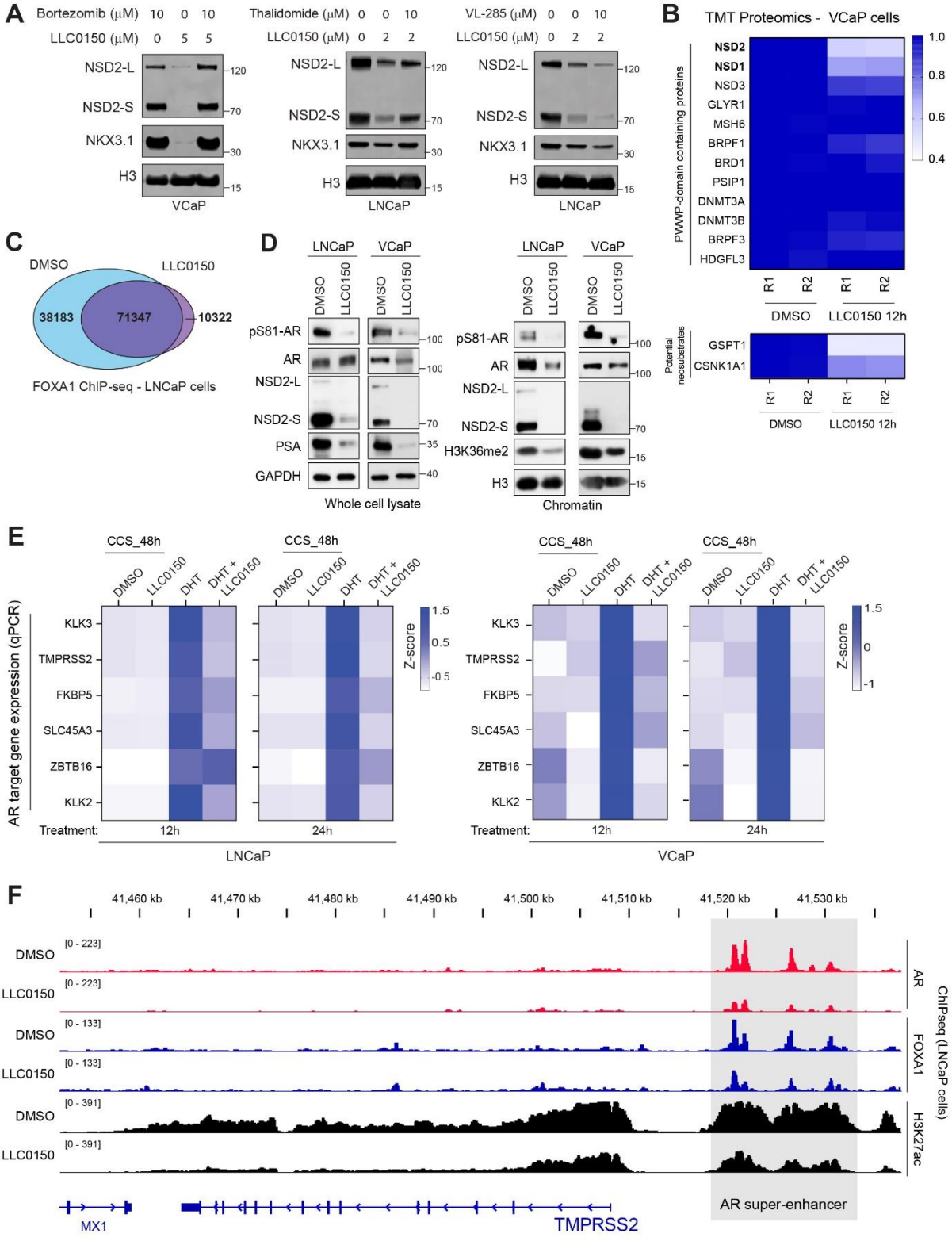
Supplementary Figure A2-5: Molecular characterization of the NSD2-rescued prostate cancer cells. A) *Top*: Immunoblots of noted proteins upon long-term treatment with control (siNC) or NSD2-targeting siRNA (siNSD2). *Bottom*: Immunoblot of NSD2 in LNCaP and 22RV1 cells treated with a control sgRNA or sgRNA targeting NSD2. GAPDH and H3 are used as loading controls. B) *Top:* Immunoblots of NSD2 and H3 from stable shNSD2-expressing LNCaP cells +/- doxycycline (1ug/ml for 72h). *Bottom*: Growth curves (CTG) of control shRNA or shNSD2-expressing LNCaP cells +/- doxycycline (n = 4 biological replicates, two-sided t-test). Mean +/- SEM are shown. C) *Left:* Representative images of colonies of control or NSD2-null LNCaP cells. *Right*: Quantification of stained colonies from left panel (n = 3 biological replicates, two-sided t-test). Mean +/- SEM are shown. Scale bar:1 cm. D) *Left*: Tumor volumes of 22RV1 parental or NSD2-KO+HA-tagged NSD2-L xenografts in mice. *Right*: Immunoblot of noted proteins from the 22RV1 xenograft tumors. (parental, n = 8; NSD2 KO, n=7). Mean +/- SEM are shown. E) *Left*: Immunoblots showing expression of the KLK3 (aka PSA) protein in the eGFP or NSD2 overexpressing LNCaP cells. *Right:* Representative images from the Boyden chamber assay in the LNCaP NSD2 WT and KO or NSD2-L rescued lines. Fluorescence signal from invaded cells is shown (n = 3 biological replicates; one-way ANOVA and Tukey's test, Scale bar:500μm). F) Immunoblot of listed proteins in wildtype or NSD2 KO LNCaP cells with stable exogenous overexpression of NSD2-L and/or NSD2-S isoforms. eGFP is used as control. G) Immunoblots of noted proteins in the NSD2 wildtype or NSD2-KO LNCaP cells rescued with exogenous WT or SET domain-deleted mutants. H) Immunoblots of noted proteins in LNCaP cells with hyper-catalytic NSD2 SET domain E1099K mutant. I) Immunoblots of noted proteins in the 22RV1 NSD2-FKBP12-F36V engineered cell lines +/- dTAG-13 treatment. J) *Left*: Tumor weights of 22RV1+NSD2-FKBP12-F36V xenografts at endpoint (day 18) +/- dTAGv-1 (n = 10 biological replicates; two-sided t-test). *Right*: Tumor images at the endpoint from the animal growth studies. Box plot center, median; box, quartiles 1-3; whiskers, min and max values. K) Expression of AR target genes in the 22RV1 NSD2-KO + NSD2-FKBP12-F36V cell line +/- dTAG-13 treatment for 12h or 24h.


A
GREAT nominated genes: NSD2-dependent AR sites
NES: −1.465
p−value: 9.037e−06
Enrichment score
LNCaP NSD2 KO vs WT - Fold change ranked DEGs
B
Biological pathways - GREAT genes
NSD2-dependent AR sites (Top 2000)
MSigDB Hallmark 2020
p-value < 0.05 TRUE FALSE
Androgen Response 4.07e-05
UV Response Dn 2.85e-02
KRAS Signaling Up 6.46e-02
Cholesterol Homeostasis 1.14e-01
TNF-alpha Signaling via NF-kB 1.3e-01
TGF-beta Signaling 1.52e-01
Hedgehog Signaling 2.27e-01
Angiogenesis 2.27e-01
Hypoxia 2.35e-01
G2-M Checkpoint 2.35e-01
−log10(p-value)
ChEA 2022
AR 21915096 ChIP-Seq LNCaP-1F5 Human 1.50e-85
AR 22383394 ChIP-Seq PROSTATE CANCER Human 2.72e-23
FOXA1 21915096 ChIP-Seq LNCaP-1F5 Human 2.22e-15
AR 27270436 ChIP-Seq VCaP Human Prostate Carcinoma 2.37e-11
OLIG2 23332759 ChIP-Seq OLIGODENDROCYTES Mouse 3.06e-08
AR 27270436 ChIP-Seq LNCaP Human Prostate Carcinoma 3.26e-08
FOXA1 33576154 ChIP-Seq Human HilarLN ProstateCancer 1.85e-07
LEF1 29337183 ChIP-Seq mESC Mouse Stem 2.82e-07
FOXA1 33576154 ChIP-Seq Human Liver ProstateCancer 4.25e-07
MITF 21258399 ChIP-Seq MELANOMA Human 5.25e-07
−log10(p-value)
Disease Perturbations from GEO up
prostate cancer DOID-10283 human GSE39452 sample 485 1.05e-03
GERD - Gastro-esophageal reflux disease C0017168 human GSE2144 sample 27 2.89e-03
familial hypercholesterolemia DOID-13810 human GSE6088 sample 908 5.28e-03
ulcerative colitis DOID-8577 human GSE11223 sample 589 6.92e-03
prostate cancer DOID-10283 human GSE3325 sample 990 7.87e-03
Diabetes DOID-9351 mouse GSE31150 sample 805 1.19e-02
Viral cardiomyopathy C0264797 human GSE4172 sample 384 1.32e-02
idiopathic pulmonary fibrosis DOID-0050156 human GSE24206 sample 868 1.91e-02
Pseudomonas Infection C0033817 human GSE923 sample 170 2.00e-02
autism spectrum disorder DOID-0060041 human GSE25507 sample 1032 2.59e-02
−log10(p-value)
C
Biological pathways - GREAT genes
Gained AR sites upon NSD2 KO (n=8,471)
GO Biological Process 2023
p-value < 0.05 TRUE FALSE
Mesenchyme Development (GO:0060485) 3.25e-05
Positive Regulation Of Nuclear Division (GO:0051785) 1.89e-04
Negative Regulation Of Cellular Process (GO:0048523) 3.83e-04
Regulation Of Mitotic Nuclear Division (GO:0007088) 4.04e-04
Icosanoid Transport (GO:0071715) 4.61e-04
Positive Regulation Of Mitotic Nuclear Division (GO:0045840) 6.11e-04
Positive Regulation Of Protein Kinase B Signaling (GO:0051897) 6.25e-04
Regulation Of Glucose Import (GO:0046324) 6.81e-04
Mesonephric Tubule Development (GO:0072164) 6.97e-04
Ureteric Bud Development (GO:0001657) 9.99e-04
−log10(p-value)
Elsevier Pathway Collection
Proteins Involved in Hypogonadotropic Hypogonadism 2.54e-04
Proteins Involved in Pulmonary Hypertension 2.63e-03
GUCY2C Signaling in Colorectal Cancer 2.98e-03
Prostaglandin E2 Synthesis Activation via Glutamate in Familial Hemiplegic Migraine 4.27e-03
Androgen Receptor/NKX3-1 Signaling 7.65e-03
GUCYC2 Signaling 1.04e-02
Proteins Involved in Phenylketonuria 1.19e-02
Longevity Related Drugs 1.19e-02
Dopamine Synthesis and Release 2.28e-02
Dopamine Synthesis and Release in Migraine 2.28e-02
−log10(p-value)
KEGG 2021 Human
Olfactory transduction 1.44e-03
Steroid biosynthesis 2.08e-02
Choline metabolism in cancer 2.44e-02
Long-term depression 2.97e-02
Cholinergic synapse 3.83e-02
Ascorbate and aldarate metabolism 4.42e-02
Pentose and glucuronate interconversions 5.54e-02
Vascular smooth muscle contraction 6.27e-02
Alanine, aspartate and glutamate metabolism 6.44e-02
Phenylalanine, tyrosine and tryptophan biosynthesis 6.54e-02
−log10(p-value)
MSigDB Hallmark 2020
Hedgehog Signaling 6.14e-02
Pancreas Beta Cells 7.38e-02
Inflammatory Response 7.48e-02
Apical Surface 8.7e-02
Fatty Acid Metabolism 1.02e-01
TGF-beta Signaling 1.23e-01
Bile Acid Metabolism 1.31e-01
Mitotic Spindle 1.85e-01
TNF-alpha Signaling via NF-kB 1.87e-01
Complement 1.87e-01
−log10(p-value)

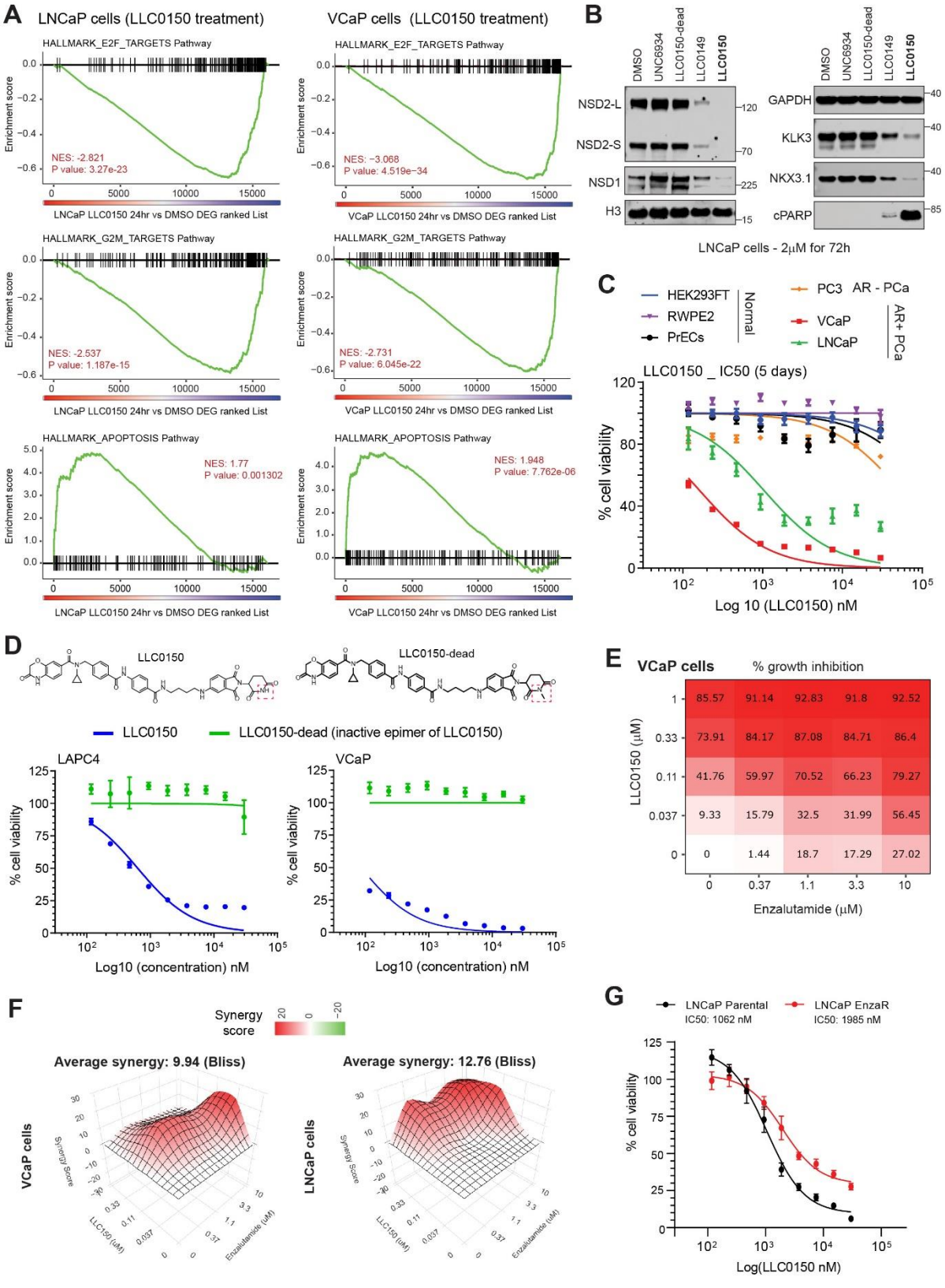
Supplementary Figure A2-6: GREAT neighboring genes and pathway enrichment analyses. A) Gene set enrichment analyses (GSEA) of GREAT nominated genes associated with the NSD2-dependent chimeric AR sites in NSD2 KO vs WT LNCaP cells (n = 2 biological replicates; GSEA enrichment test). B) GREAT and Enrichr analyses of putative chimeric AR gene targets in molecular signature and biology pathway databases (Fisher's exact test). C) GREAT and Enrichr analyses of putative gene targets of gained AR sites in the NSD2 KO LNCaP cells in pathway collections and databases (Fisher's exact test).


A
LNCaP cells: Size exclusion chromatography
Fraction #:
Input
10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
Pol II
BRD4
NSD2-L
NSD2-S
AR
FOXA1
NKX3-1
HOXB13
AR containing fractions
B
Input
IP:
IgG
AR
NSD2-L
NSD2-S
AR
AR-V7
LNCaP
VCaP
22RV1
C
HA-tagged NSD2 frags:
EV
FL
dPWWP2
PH
SET
PWWP1a
HMGa
HA
Halo
HA-tag (NSD2 fragments)
Halo-AR
Input
IP:Halo
IP: AR-Full length (FL)
D
Halo-tagged AR frags:
AR-NTD
AR-DBD
AR-LBD
AR-fl
NSD2 :
Halo-tag (AR fragments)
IP:HA
IP: HA-NSD2-L
IP: Halo-AR-DBD
E
Exp Z-score
LNCaP cells - AR up-regulated genes
R1
R2
WT
NSD1 KO
NSD3 KO
G
NSD3-L
NSD3-S
NSD1
AR
H3K36me2
GAPDH
H3
LNCaP
H
LNCaP cells_NSD1 KO vs WT
Down-regulated Hallmark pathways
NES
DNA_REPAIR
REACTIVE_OXYGEN_SPECIES_PATHWAY
PI3K_AKT_MTOR_SIGNALING
MTORC1_SIGNALING
WNT_BETA_CATENIN_SIGNALING
APOPTOSIS
NTERFERON_ALPHA_RESPONSE
APICAL_JUNCTION
NOTCH_SIGNALING
UNFOLDED_PROTEIN_RESPONSE
IL6_JAK_STAT3_SIGNALING
ESTROGEN_RESPONSE_EARLY
UV_RESPONSE_UP
HEDGEHOG_SIGNALING
ALLOGRAFT_REJECTION
KRAS_SIGNALING_DN
ESTROGEN_RESPONSE_LATE
IL2_STAT5_SIGNALING
MYOGENESIS
KRAS_SIGNALING_UP
GLYCOLYSIS
TNFA_SIGNALING_VIA_NFKB
HYPOXIA
APICAL_SURFACE
UV_RESPONSE_DN
MYC_TARGETS_V2
MITOTIC_SPINDLE
ANDROGEN_RESPONSE
MYC_TARGETS_V1
TGF_BETA_SIGNALING
EPITHELIAL_MESENCHYMAL_TRANSITION
G2M_CHECKPOINT
E2F_TARGETS
GSEA Net Enrichment Score (NES)
I
siNSD1:
siNSD2:
NSD1
NSD2-L
NSD2-S
MYC
AR
cPARP
H3K36me2
H3K27me3
H3
VCaP
F
HALLMARK_ANDROGEN_RESPONSE Pathway
Enrichment score
NES: 1.447
p-value: 0.2273
LNCaP NSD3 KO vs WT DEG ranked list
J
BROAD CRISPR Achilles Data - DepMap
Cell line lineage:
uterus
urinary_tract
upper_aerodigestive
unknown
thyroid
soft_tissue
skin
prostate
plasma_cell
peripheral_nervous_sys
pancreas
ovary
lymphocyte
lung
liver
kidney
gastric
eye
esophagus
epidermoid_carcinoma
embryo
colorectal
cervix
central_nervous_sys
breast
bone
blood
bile_duct
NSD1
NSD2
NSD3
POLD2 (pan-essential gene)
Dependency
Common essential genes (n = 2164)

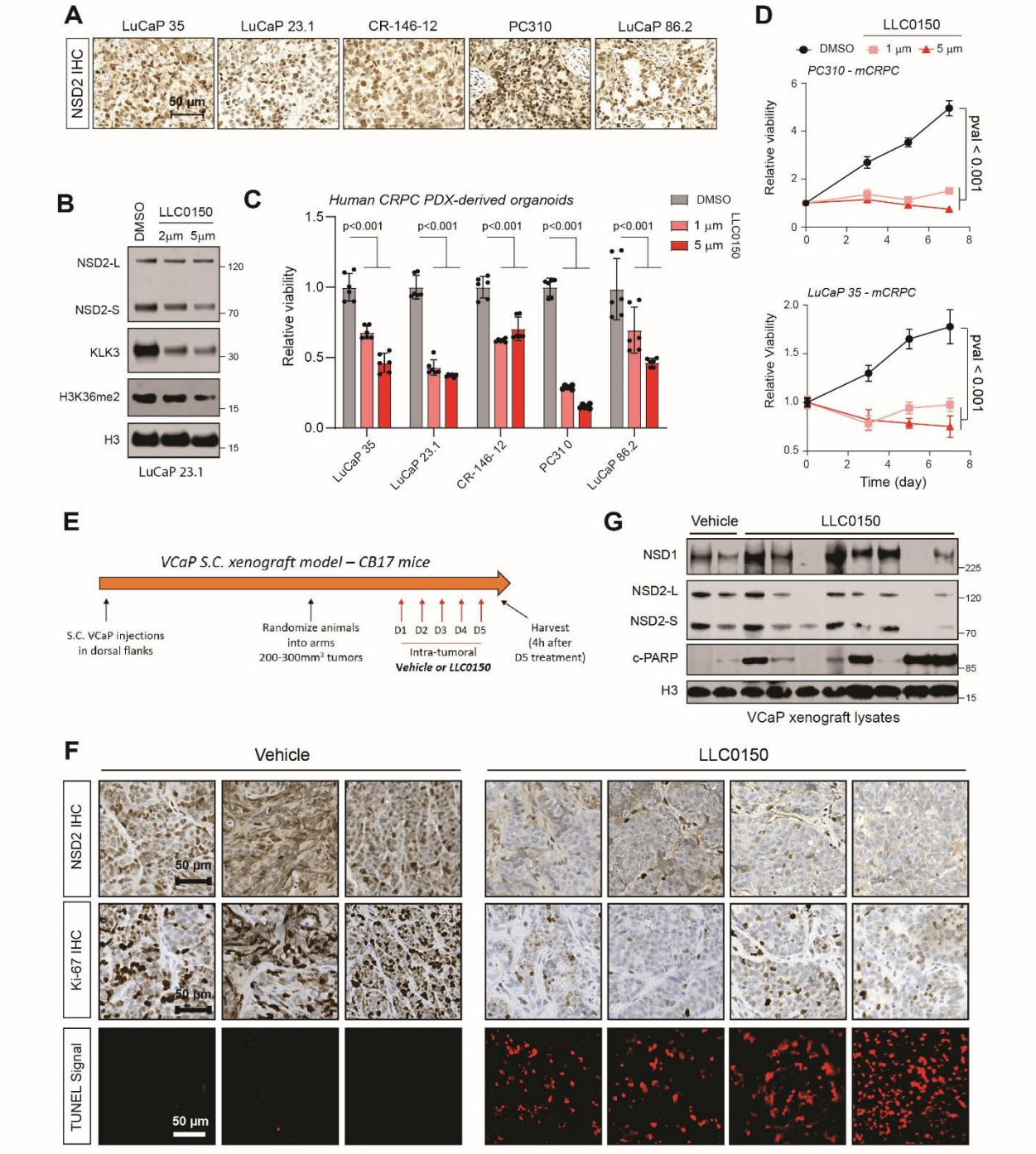
Supplementary Figure A2-7: Fragment-based NSD2–AR co-immunoprecipitation and characterization of the NSD paralog knockout prostate cancer cells. A) Immunoblots of noted proteins in size-exclusion chromatography fractions of nuclear lysate extracted from wild-type LNCaP cells. Fractions containing the AR protein are marked. B) Immunoblots of indicated proteins upon co-immunoprecipitation of AR in prostate cancer cells. C) Immunoblots of indicated proteins upon immunoprecipitation of exogenously expressed Halo-tagged full-length AR protein in HEK293FT cells that express HA-tagged NSD2 fragments. Both input (left) and immunoprecipitation (right) blots are shown. D) *Left*: Immunoblots of HA-tag-based immunoprecipitation of full length NSD2 in HEK293FT cells that express the Halo-tagged AR protein fragments. *Right*: Immunoblots of Halo-tag-based immunoprecipitation of the DNA-binding domain (DBD) of AR in HEK293FT cells that overexpress different HA-tagged NSD2 fragments. For both experiments, input and immunoprecipitation blots are shown. E) Heatmap of AR upregulated genes (z-score) in NSD1 or NSD3 knockout (KO) LNCaP cells. F) GSEA plots for AR-regulated genes using the fold change rank-ordered genes from LNCaP NSD3 knock-out (NSD3 KO) vs control cell lines. DEGS, differentially expressed genes (n = 2 biological replicates; GSEA enrichment test). G) Immunoblot of indicated proteins in NSD1 or NSD3-deficient LNCaP cells. H) GSEA net enrichment score (NES) plot of downregulated hallmark pathways in LNCaP NSD1 knocked out (KO) vs wild-type control cells. I) Immunoblot of indicated proteins upon treatment with NSD1 and NSD2-targeting siRNAs (labeled as siNSD1 and siNSD2) independently or in combination in VCaP cells. J) Dependency map (DepMap) plots showing the dependency scores for NSD1, NSD2, NSD3, and POLD2 (positive control; pan-essential gene) across cell lines from distinct originating tissues. The red dotted line indicates pan-essentiality z-score cutoff. Box plot center, median; box, quartiles 1-3; whiskers, quartiles 1-3 ± 1.5 × interquartile range; dot, outliers.


A
Bortezomib (μM) 10 0 10
LLC0150 (μM) 0 5 5
NSD2-L
NSD2-S
NKX3.1
H3
VCaP
Thalidomide (μM) 0 0 10
LLC0150 (μM) 0 2 2
LNCaP
VL-285 (μM) 0 0 10
LLC0150 (μM) 0 2 2
LNCaP
B
TMT Proteomics - VCaP cells
PWWP-domain containing proteins
NSD2
NSD1
NSD3
GLYR1
MSH6
BRPF1
BRD1
PSIP1
DNMT3A
DNMT3B
BRPF3
HDGFL3
R1 R2 R1 R2
DMSO LLC0150 12h
Potential neosubstrates
GSPT1
CSNK1A1
C
DMSO
LLC0150
38183
71347
10322
FOXA1 ChIP-seq - LNCaP cells
D
LNCaP VCaP
DMSO LLC0150 DMSO LLC0150
pS81-AR
AR
NSD2-L
NSD2-S
PSA
GAPDH
Whole cell lysate
H3K36me2
H3
Chromatin
E
CCS_48h
DMSO LLC0150 DHT DHT + LLC0150
AR target gene expression (qPCR)
KLK3
TMPRSS2
FKBP5
SLC45A3
ZBTB16
KLK2
Z-score
Treatment: 12h 24h
LNCaP
VCaP
F
41,460 kb 41,470 kb 41,480 kb 41,490 kb 41,500 kb 41,510 kb 41,520 kb 41,530 kb
DMSO
LLC0150
[0 - 223]
[0 - 133]
[0 - 391]
AR
FOXA1
H3K27ac
ChIPseq (LNCaP cells)
MX1
TMPRSS2
AR super-enhancer

Supplementary Figure A2-8: Mechanistic characterization of the NSD1/2 PROTAC degrader LLC0150. A) Immunoblots of indicated proteins in VCaP and LNCaP cells pre-treated with bortezomib, thalidomide, or VL-285 followed by treatment with LLC0150 at noted concentrations. B) Heatmap of relative abundance of several PWWP-domain-containing and known neo-substrate proteins detected via Tandem Mass Tag (TMT) based quantitative MS upon 12h treatment with LLC0150 in VCaP cells. C) Genome-wide changes in FOXA1 ChIP-seq peaks in LNCaP cells treated with LLC0150 (2uM for 48h). D) Immunoblots of noted proteins in whole-cell or chromatin lysates from VCaP and LNCaP cells treated with LLC0150 (2uM) for 24h. E) Heatmap of z-score normalized expression (qRT-PCR) of AR target genes in LNCaP and VCaP cells treated with LLC0150 followed by DHT stimulation (10nM for 24h). Treatment with DHT alone is used as a control. F) Read density ChIP-seq tracks of AR, FOXA1, and H3K27ac within the *TMPRSS2* super-enhancer in LNCaP cells treated with LLC0150 (2uM for 24h). Super-enhancer cluster is highlighted in a gray box.


A
LNCaP cells (LLC0150 treatment)
VCaP cells (LLC0150 treatment)
HALLMARK_E2F_TARGETS Pathway
NES: -2.821
P value: 3.27e-23
NES: -3.068
P value: 4.519e-34
HALLMARK_G2M_TARGETS Pathway
NES: -2.537
P value: 1.187e-15
NES: -2.731
P value: 6.045e-22
HALLMARK_APOPTOSIS Pathway
NES: 1.77
P value: 0.001302
NES: 1.948
P value: 7.762e-06
Enrichment score
LNCaP LLC0150 24hr vs DMSO DEG ranked List
VCaP LLC0150 24hr vs DMSO DEG ranked List
B
DMSO
UNC6934
LLC0150-dead
LLC0149
LLC0150
NSD2-L
NSD2-S
NSD1
H3
GAPDH
KLK3
NKX3.1
cPARP
LNCaP cells - 2μM for 72h
C
HEK293FT
RWPE2
PrECs
Normal
PC3 AR - PCa
VCaP
LNCaP
AR+ PCa
LLC0150 _ IC50 (5 days)
% cell viability
Log 10 (LLC0150) nM
D
LLC0150
LLC0150-dead
LLC0150-dead (inactive epimer of LLC0150)
LAPC4
VCaP
% cell viability
Log10 (concentration) nM
E
VCaP cells
% growth inhibition
LLC0150 (μM)
Enzalutamide (μM)
1 85.57 91.14 92.83 91.8 92.52
0.33 73.91 84.17 87.08 84.71 86.4
0.11 41.76 59.97 70.52 66.23 79.27
0.037 9.33 15.79 32.5 31.99 56.45
0 0 1.44 18.7 17.29 27.02
0 0.37 1.1 3.3 10
F
Synergy score
Average synergy: 9.94 (Bliss)
Average synergy: 12.76 (Bliss)
VCaP cells
LNCaP cells
Synergy Score
LLC150 (uM)
Enzalutamide (uM)
G
LNCaP Parental
IC50: 1062 nM
LNCaP EnzaR
IC50: 1985 nM
% cell viability
Log(LLC0150 nM)

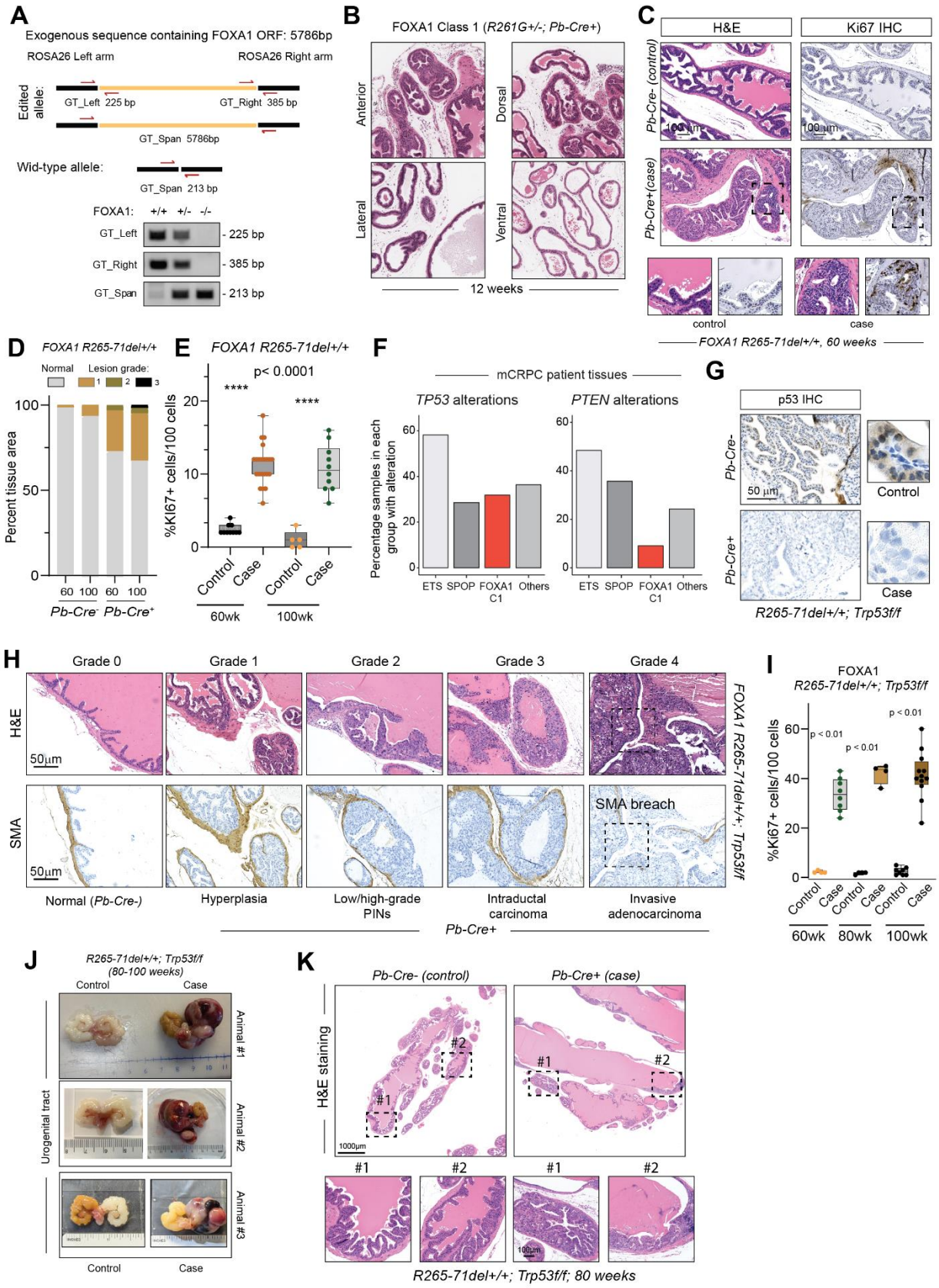
Supplementary Figure A2-9: Transcriptomic effect and drug synergism of LLC0150 in prostate cancer cells. A) GSEA plots for E2F, G2M, and apoptosis pathway genes using the fold change rank-ordered genes from the LLC0150 vs DMSO treated LNCaP (left) or VCaP (right) cell lines. DEGS, differentially expressed genes (n = 2 biological replicates; GSEA enrichment test). B) Immunoblot of noted proteins in LNCaP cells treated with LLC0150 (2uM for 72h), dead-analog (LLC0150-dead), or the warhead alone (UNC6934). LLC0149 is an independent NSD1/2 PROTAC. C) Dose-response curves of LLC0150 in normal prostate, AR-positive, or AR-negative prostate cancer cell lines at the indicated concentrations for five days. (PrECs, n = 3 biological replicates; others, n = 6 biological replicates). Mean +/- SEM are shown. D) Dose-response curves of LLC0150 and its inactive epimer control (LLC0150-dead) in LAPC4 and VCaP cell lines (n = 6 biological replicates). Mean +/- SEM are shown. E) Percent growth inhibition (Cell-titer Glo) of VCaP cells upon co-treatment with varying concentrations of LLC0150 and enzalutamide for 5 days. F) 3D synergy plots of LLC0150 and enzalutamide co-treated LNCaP and VCaP cells. Red peaks in the 3D plots denote synergy with the average synergy scores noted above. G) Dose-response curves of LLC0150 in LNCaP parental and enzalutamide-resistant cell lines at varying concentrations for five days. Half maximal inhibitory concentrations (IC50) are noted (n = 5 biological replicates). Mean +/- SEM are shown.


A
LuCaP 35
LuCaP 23.1
CR-146-12
PC310
LuCaP 86.2
NSD2 IHC
50 µm
B
DMSO
LLC0150
2µm 5µm
NSD2-L
NSD2-S
KLK3
H3K36me2
H3
120
70
30
15
15
LuCaP 23.1
C
Human CRPC PDX-derived organoids
DMSO
1 µm
5 µm
LLC0150
Relative viability
p<0.001
LuCaP 35
LuCaP 23.1
CR-146-12
PC310
LuCaP 86.2
D
LLC0150
DMSO
1 µm
5 µm
PC310 - mCRPC
Relative viability
pval < 0.001
LuCaP 35 - mCRPC
Relative Viability
pval < 0.001
Time (day)
E
VCaP S.C. xenograft model – CB17 mice
S.C. VCaP injections
in dorsal flanks
Randomize animals
into arms
200-300mm³ tumors
D1 D2 D3 D4 D5
Intra-tumoral
Vehicle or LLC0150
Harvest
(4h after
D5 treatment)
G
Vehicle
LLC0150
NSD1
NSD2-L
NSD2-S
c-PARP
H3
225
120
70
85
15
VCaP xenograft lysates
F
Vehicle
LLC0150
NSD2 IHC
Ki-67 IHC
TUNEL Signal
50 µm

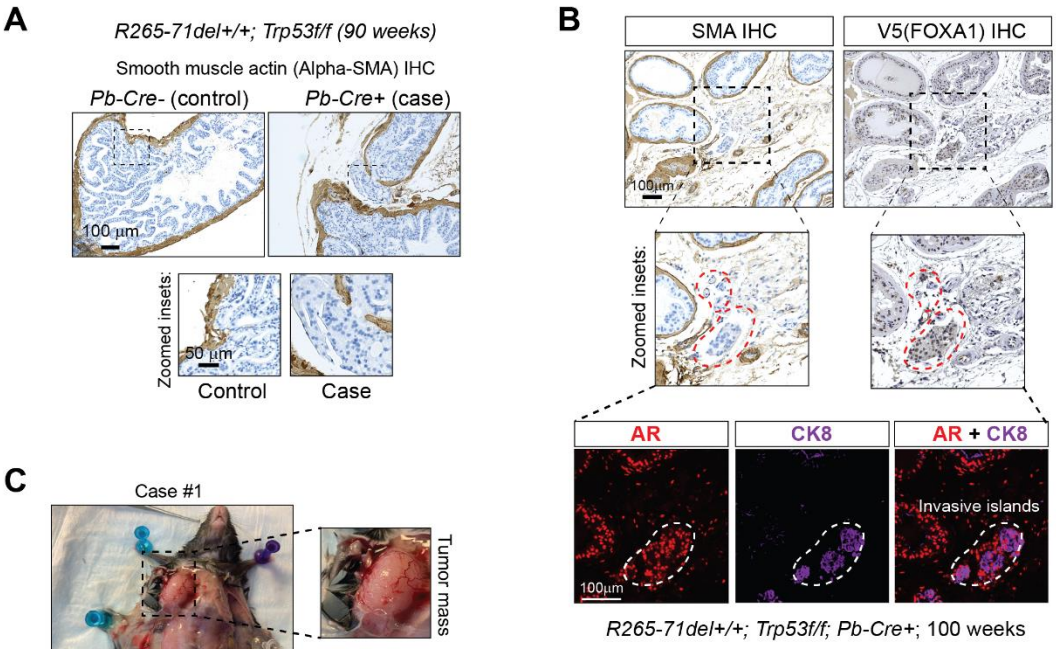
Supplementary Figure A2-10: Efficacy assessment of LLC0150 in prostate cancer organoids and xenografts. A) Representative images of NSD2 IHC in a panel of patient-derived xenografts (PDXs). Scale bar:50µm B) Immunoblot of NSD2, AR targets, and histone marks in the LuCaP 23.1 PDX-derived organoid line treated with LLC0150, the NSD1/2 degrader. Total H3 is the loading control. C) Barplots showing relative viability of the PDX-derived organoid lines treated with two doses of LLC0150. In all lines, degradation of NSD1/2 reduces cell viability (n = 6 biological replicates; one-way ANOVA and Tukey's test). Mean +/- SEM are shown. D) Growth curves (Cell-titer Glo) of two representative AR/NSD2-positive PDX-derived organoid lines treated with DMSO or LLC0150 (n = 6 biological replicates; one-way ANOVA and Tukey's test). Mean +/- SEM are shown. E) Schematic overview of the LLC0150 intratumoral injection study in a VCaP xenograft model. F) *Top*: Representative NSD2 and KI67 IHC images in the vehicle and LLC0150 treated tumors. Loss of NSD2 correlates with a reduction in proliferating KI67-positive cells. *Bottom*: TUNEL assay in the vehicle and LLC0150 treated tumors reveal a high number of apoptotic (TUNEL-positive) cells in the drug-treated tumors.Vehicle, n = 3 tumors; LLC0150, n = 4 tumors. Scale bar: 50µm. G) Immunoblot of noted proteins in matched tumor lysates from f. Total H3 is used as a loading control.

Appendix B

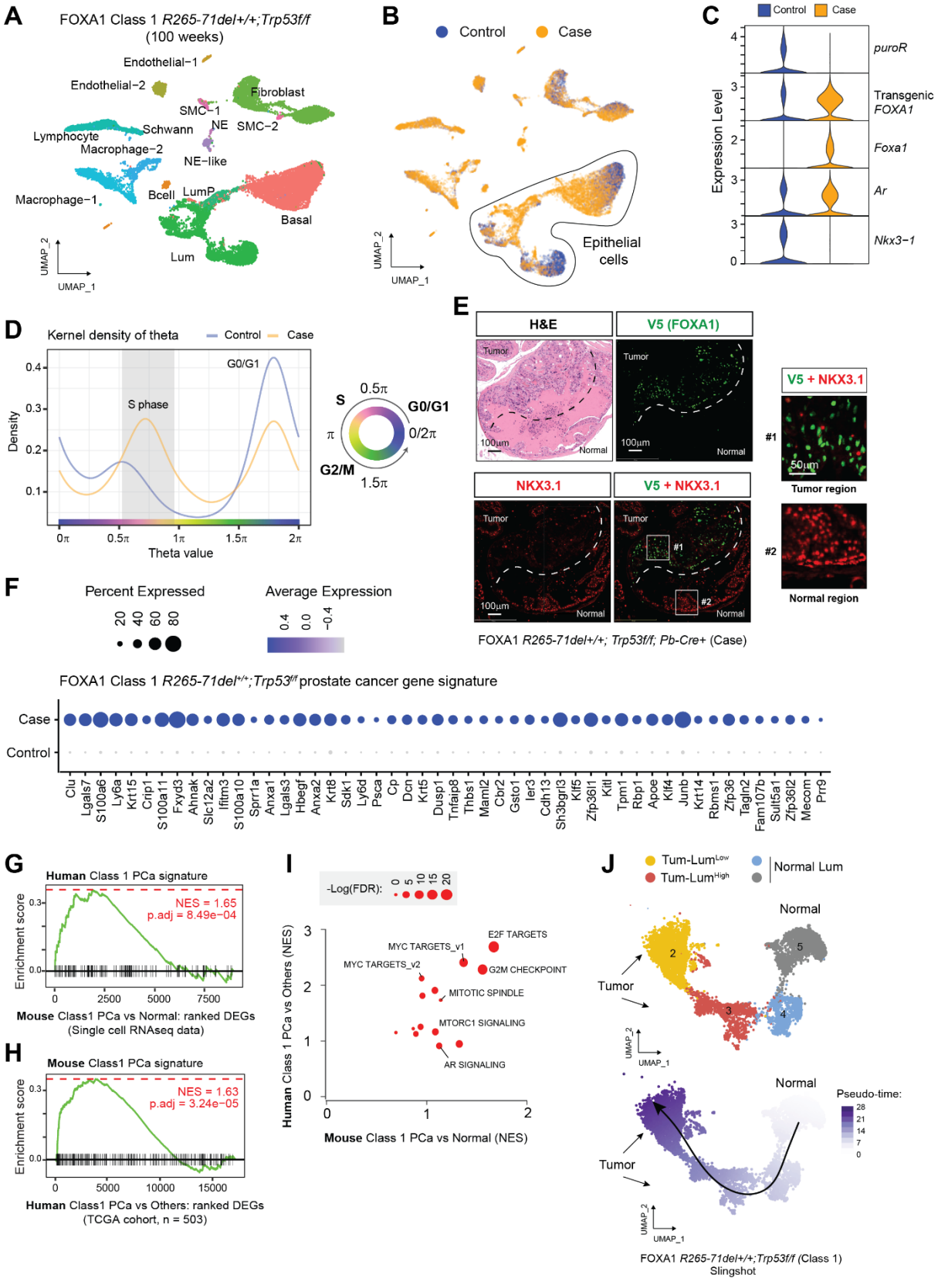
Supplementary Figures for Chapter 3


A
Exogenous sequence containing FOXA1 ORF: 5786bp
ROSA26 Left arm
ROSA26 Right arm
Edited allele:
GT_Left 225 bp
GT_Right 385 bp
GT_Span 5786bp
Wid-type allele:
GT_Span 213 bp
FOXA1: +/+ +/- -/-
GT_Left - 225 bp
GT_Right - 385 bp
GT_Span - 213 bp
B
FOXA1 Class 1 (R261G+/-; Pb-Cre+)
Anterior
Dorsal
Lateral
Ventral
12 weeks
C
H&E
Ki67 IHC
Pb-Cre- (control)
Pb-Cre+(case)
100 µm
control
case
FOXA1 R265-71del+/+, 60 weeks
D
FOXA1 R265-71del+/+
Normal
Lesion grade: 1 2 3
Percent tissue area
Pb-Cre- Pb-Cre+
E
FOXA1 R265-71del+/+
p< 0.0001
%KI67+ cells/100 cells
Control Case Control Case
60wk 100wk
F
mCRPC patient tissues
TP53 alterations
PTEN alterations
Percentage samples in each group with alteration
ETS SPOP FOXA1 C1 Others
G
p53 IHC
Pb-Cre-
Pb-Cre+
50 µm
Control
Case
R265-71del+/+; Trp53f/f
H
Grade 0
Grade 1
Grade 2
Grade 3
Grade 4
H&E
SMA
50µm
SMA breach
FOXA1 R265-71del+/+; Trp53f/f
Normal (Pb-Cre-)
Hyperplasia
Low/high-grade PINs
Intraductal carcinoma
Invasive adenocarcinoma
Pb-Cre+
I
FOXA1
R265-71del+/+; Trp53f/f
%Ki67 + cells/100 cells
p < 0.01
Control Case Control Case Control Case
60wk 80wk 100wk
J
R265-71del+/+; Trp53f/f
(80-100 weeks)
Control Case
Animal #1
Animal #2
Animal #3
Urogenital tract
Control Case
K
Pb-Cre- (control)
Pb-Cre+ (case)
H&E staining
#1 #2
1000µm
100µm
R265-71del+/+; Trp53f/f; 80 weeks

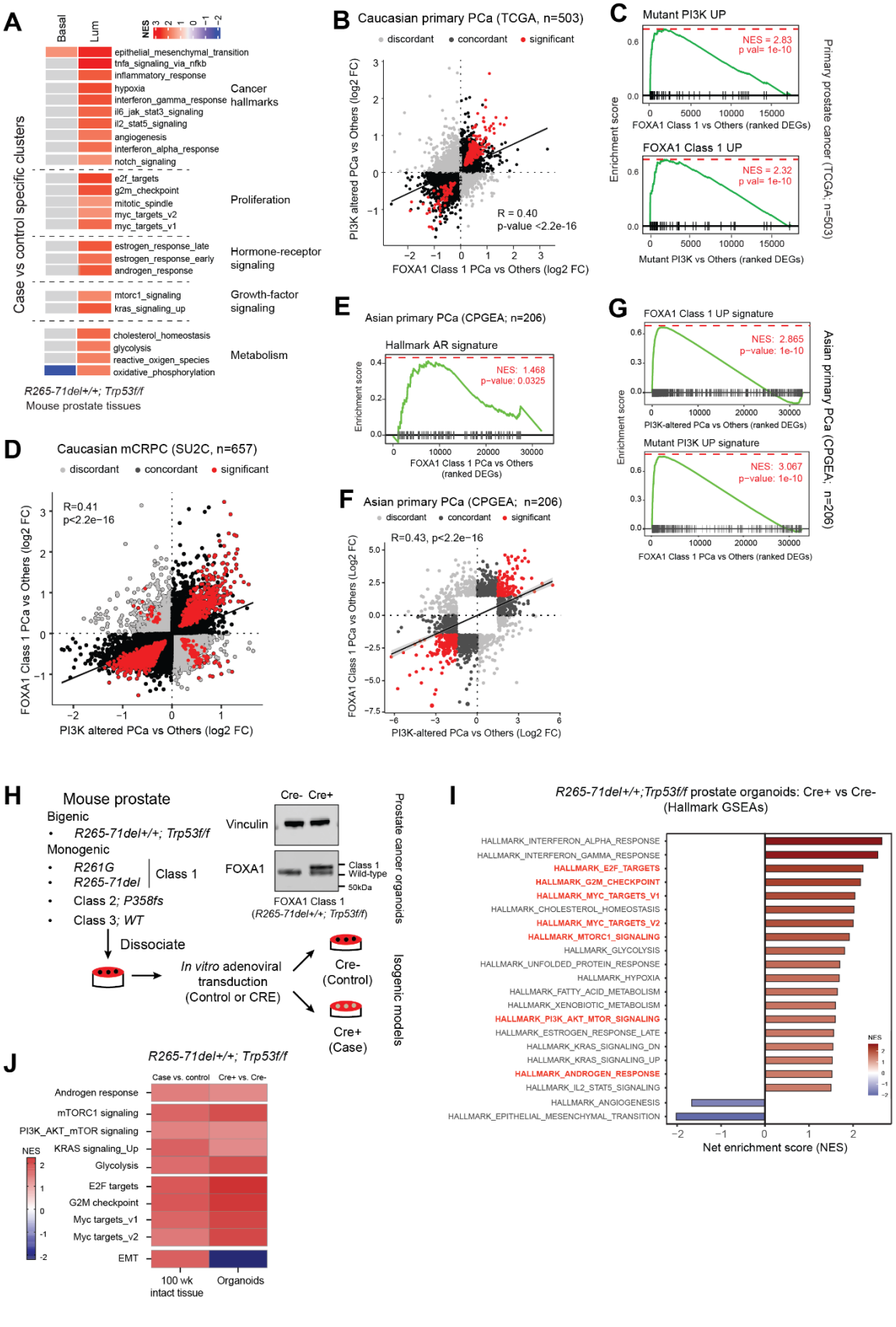
Supplementary Figure A3-1: Generation and characterization of FOXA1 Class 1 mutant transgenic mice. A) Top: Schematic of the FOXA1 knock-in construct at the ROSA26 locus with genotyping primers highlighted in red. Bottom: Representative genotyping agarose gel distinguishing transgenic mice that are wild-type (-/-), heterozygous (+/-), or homozygous (+/+). GTL and GTR bands indicate the presence of the transgene, while the absence of a span band indicates a homozygous mutant animal. B) Representative H&E-stained cross-sections of *FOXA1 R261G+/-* Cre+ case prostate lobes from 12-week-old mice. Scale =100µm. C) Representative H&E and Ki67 stained cross-sections of *FOXA1 R265-71del+/+* control and case prostate tissues (anterior lobe) from 60-week-old mice. Scale =100µm. D) Histopathological grading of tissues from panel C. E) Percentage of Ki67-positive cells from 60 and 100-week-old tissues from *FOXA1 R265-71del+/+* control-case prostate tissues (two-sided t-test). Box plot: center line, median; box, interquartile range (Q1–Q3); whiskers, minimum to maximum values; all individual data points shown. F) Co-recurrence of *TP53* or *PTEN* alterations in CRPC tumors harboring mutually exclusive ETS fusions, SPOP mutations, or FOXA1 Class 1 mutations (n=371). G) Representative p53 IHC stained cross-sections of *FOXA1 R265-71del+/+;Trp53f/f* control and case animals. Scale =50µm. H) Representative H&E and alpha-SMA stained cross sections in the *FOXA1 R265-71del+/+;Trp53f/f* tissues representing the distinct grades used in this study. SMA staining is retained and highlighted as continuous brown expression at the parabasal regions in Grade 1:Hyperplasia, Grade 2: Low/high-grade prostate intraepithelial neoplasia (PINs), Grade 3: Florid high-grade PIN, Atypical intraductal proliferation and Intraductal carcinoma, but absent in Grade 4: Invasive adenocarcinoma. I) Percentage of Ki67-positive cells from 60, 80, and 100 week-old tissues from *FOXA1 R265-71del+/+;Trp53f/f* control-case prostate tissues (two-sided t-test). Box plot: center line, median; box, interquartile range (Q1–Q3); whiskers, minimum to maximum values; all individual data points shown. J) Representative images of the urogenital tract (prostate gland and seminal vesicles) in three distinct FOXA1 *R265-71del+/+;Trp53f/f* animals. K) Representative H&E-stained cross-sections of *FOXA1 R265-71del+/+;Trp53f/f* control-case prostate tissues (anterior lobe) from 80 week-old mice. Scale =1000um. Zoomed insets show normal and tumor lesions. Scale =100um.


A
R265-71del+/+; Trp53f/f (90 weeks)
Smooth muscle actin (Alpha-SMA) IHC
Pb-Cre- (control)
Pb-Cre+ (case)
100 μm
Zoomed insets:
50 μm
Control
Case
B
SMA IHC
V5(FOXA1) IHC
100μm
Zoomed insets:
AR
CK8
AR + CK8
Invasive islands
100μm
R265-71del+/+; Trp53f/f; Pb-Cre+; 100 weeks
C
Case #1
Tumor mass
R265-71del+/+; Trp53f/f; Pb-Cre+ (100 weeks)
D
H&E
V5 (FOXA1)
FOXA1
100μm
Tumor mass
R265-71del+/+ ;TP53f/f;Pb-Cre+

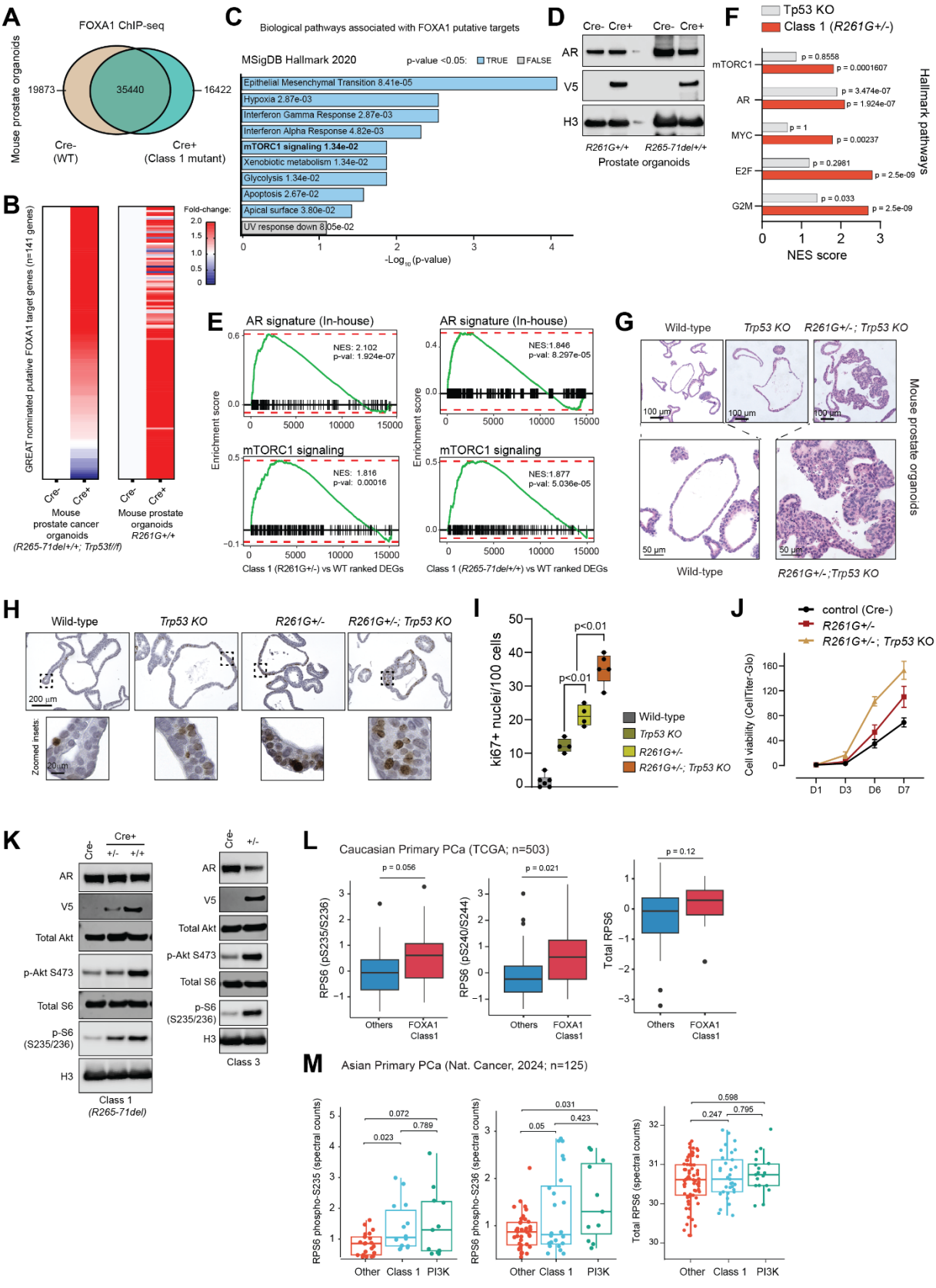
Supplementary Figure A3-2: High-grade invasive carcinoma in *FOXA1 R265-71del+/+;Trp53f/f* animals. A) Representative alpha-SMA IHC in *FOXA1 R265-71del+/+;Trp53f/f* 90-week-old control and case tissues. Scale =100µm. Zoomed insets of selected regions (highlighted in a black, dotted box) are shown below. B) Representative alpha-SMA and V5 IHC and multiplex- IF of CK8 and AR in *FOXA1 R265-71del+/+;Trp53f/f* 100-week-old case tissues. Zoomed insets of delaminated/invasive regions (highlighted in a black, dotted box and red, dotted circles) are shown below. Scale =100µm. C) Image of a tumor mass in a 100-week-old *FOXA1 R265-71del+/+;Trp53f/f* case animal. D) Representative H&E, V5, and FOXA1 stained cross sections from the tumor cavity mass shown in panel C. Scale =100µm.


A
FOXA1 Class 1 R265-71del+/+;Trp53f/f
(100 weeks)
Endothelial−1
Endothelial−2
Fibroblast
SMC−1
NE
SMC−2
Schwann
Lymphocyte
Macrophage−2
NE−like
Macrophage−1
Bcell
LumP
Basal
Lum
UMAP_2
UMAP_1
B
Control
Case
Epithelial cells
C
Control
Case
Expression Level
puroR
Transgenic FOXA1
Foxa1
Ar
Nkx3−1
D
Kernel density of theta
Control
Case
G0/G1
S phase
Density
Theta value
S
G0/G1
G2/M
E
H&E
V5 (FOXA1)
NKX3.1
V5 + NKX3.1
Tumor
Normal
100μm
50μm
#1
#2
Tumor region
Normal region
FOXA1 R265-71del+/+; Trp53f/f; Pb-Cre+ (Case)
F
Percent Expressed
Average Expression
FOXA1 Class 1 R265-71del+/+;Trp53f/f prostate cancer gene signature
Case
Control
Clu
Lgals7
S100a6
Ly6a
Krt15
Crip1
S100a11
Fxyd3
Ahnak
Slc12a2
Ifitm3
S100a10
Sprr1a
Anxa1
Lgals3
Hbegf
Anxa2
Krt8
Sdk1
Ly6d
Psca
Cp
Dcn
Krt5
Dusp1
Tnfaip8
Thbs1
Maml2
Cbr2
Gsto1
Ier3
Cdh13
Sh3bgrl3
Klf5
Zfp36l1
Kitl
Tpm1
Rbp1
Apoe
Klf4
Junb
Krt14
Rbms1
Zfp36
Tagln2
Fam107b
Sult5a1
Zfp36l2
Mecom
Prr9
G
Human Class 1 PCa signature
NES = 1.65
p.adj = 8.49e−04
Enrichment score
Mouse Class1 PCa vs Normal: ranked DEGs
(Single cell RNAseq data)
H
Mouse Class1 PCa signature
NES = 1.63
p.adj = 3.24e−05
Enrichment score
Human Class1 PCa vs Others: ranked DEGs
(TCGA cohort, n = 503)
I
-Log(FDR):
E2F TARGETS
MYC TARGETS_v1
MYC TARGETS_v2
G2M CHECKPOINT
MITOTIC SPINDLE
MTORC1 SIGNALING
AR SIGNALING
Human Class 1 PCa vs Others (NES)
Mouse Class 1 PCa vs Normal (NES)
J
Tum-Lum^Low
Tum-Lum^High
Normal Lum
Normal
Tumor
UMAP_2
UMAP_1
Pseudo-time:
FOXA1 R265-71del+/+;Trp53f/f (Class 1)
Slingshot

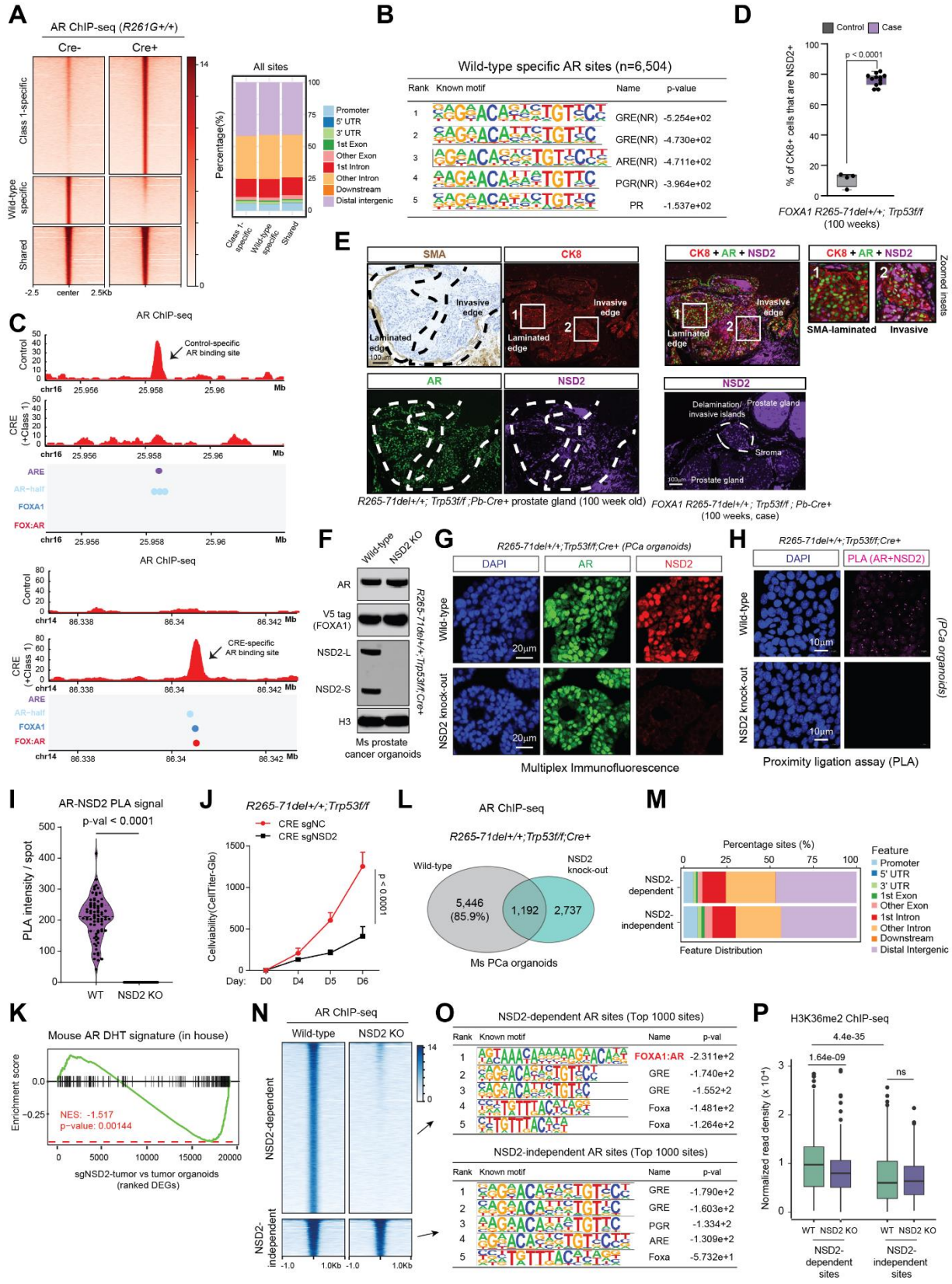
Supplementary Figure A3-3: Single-cell transcriptomic characterization of *FOXA1 R265-71del+/+;Trp53f/f* tumor tissues. A) UMAP plots of the distinct cell types present in the 100-week-old *FOXA1 R265-71del+/+;Trp53f/f* case and control animals (10x single-cell RNA-sequencing). B) UMAP plots of the distinct cell types from panel (A) with case or control tissue-derived cells colored across clusters. Epithelial cells used for further analyses are marked with a solid black line. C) Violin plots showing expression of select genes in the epithelial cells from control or case *FOXA1 R265-71del+/+;Trp53f/f* animals. D) Cell cycle plot (using the Tricycle R package) shows distribution of case or control cells across the distinct cell cycle phases as denoted by the theta values. E) Representative H&E and multiplex-IF of V5 and NKX3.1 in *FOXA1 R265-71del+/+;Trp53f/f* case tissues. Normal and tumor regions are annotated along with zoomed insets. F) Dot plot of the prostate cancer signature genes (defined in this study, see methods) between *FOXA1 R265-71del+/+;Trp53f/f* case and control cells. G) Gene set enrichment analysis (GSEA) plots for human Class 1 PCa activated signature from the rank-ordered genes from mouse Class 1 mutant tumor tissues. DEGS, differentially expressed genes (GSEA enrichment test). H) Gene set enrichment analysis (GSEA) plots for mouse Class 1 PCa activated signature from the rank-ordered genes of human Class 1 mutant tumor tissues. DEGS, differentially expressed genes (GSEA enrichment test). I) Hallmark gene set net enrichment score (NES) between mouse and human Class 1 ranked gene lists. Each point denotes a hallmark gene set, and the size of the points corresponds to the false discovery rate (FDR). J) UMAP plot of distinct clusters in the 100-week-old *FOXA1 R265-71del+/+;Trp53f/f* single-cell RNA-seq data (Top, basal cells removed) and pseudotime trajectory analyses (Bottom, Slingshot, see methods). The solid black arrow denotes the trajectory from normal (cluster 4,5) to tumor populations (cluster 2,3).


A
Basal
Lum
NES
epithelial_mesenchymal_transition
tnfa_signaling_via_nfkb
inflammatory_response
hypoxia
interferon_gamma_response
il6_jak_stat3_signaling
il2_stat5_signaling
angiogenesis
interferon_alpha_response
notch_signaling
e2f_targets
g2m_checkpoint
mitotic_spindle
myc_targets_v2
myc_targets_v1
estrogen_response_late
estrogen_response_early
androgen_response
mtorc1_signaling
kras_signaling_up
cholesterol_homeostasis
glycolysis
reactive_oxigen_species
oxidative_phosphorylation
Cancer hallmarks
Proliferation
Hormone-receptor signaling
Growth-factor signaling
Metabolism
Case vs control specific clusters
R265-71del+/+; Trp53f/f
Mouse prostate tissues
B
Caucasian primary PCa (TCGA, n=503)
discordant
concordant
significant
PI3K altered PCa vs Others (log2 FC)
FOXA1 Class 1 PCa vs Others (log2 FC)
R = 0.40
p-value <2.2e-16
C
Mutant PI3K UP
NES = 2.83
p val= 1e-10
FOXA1 Class 1 vs Others (ranked DEGs)
FOXA1 Class 1 UP
NES = 2.32
p val= 1e-10
Enrichment score
Mutant PI3K vs Others (ranked DEGs)
Primary prostate cancer (TCGA; n=503)
D
Caucasian mCRPC (SU2C, n=657)
discordant
concordant
significant
R=0.41
p<2.2e−16
FOXA1 Class 1 PCa vs Others (log2 FC)
PI3K altered PCa vs Others (log2 FC)
E
Asian primary PCa (CPGEA; n=206)
Hallmark AR signature
NES: 1.468
p-value: 0.0325
Enrichment score
FOXA1 Class 1 PCa vs Others
(ranked DEGs)
F
Asian primary PCa (CPGEA; n=206)
discordant
concordant
significant
R=0.43, p<2.2e−16
FOXA1 Class 1 PCa vs Others (Log2 FC)
PI3K-altered PCa vs Others (Log2 FC)
G
FOXA1 Class 1 UP signature
NES: 2.865
p−value: 1e-10
PI3K-altered PCa vs Others (ranked DEGs)
Mutant PI3K UP signature
NES: 3.067
p−value: 1e-10
Enrichment score
FOXA1 Class 1 PCa vs Others (ranked DEGs)
Asian primary PCa (CPGEA; n=206)
H
Mouse prostate
Bigenic
R265-71del+/+; Trp53f/f
Monogenic
R261G
R265-71del
Class 1
Class 2; P358fs
Class 3; WT
Dissociate
In vitro adenoviral transduction (Control or CRE)
Cre- (Control)
Cre+ (Case)
Isogenic models
Cre- Cre+
Vinculin
FOXA1
Class 1
Wild-type
50kDa
FOXA1 Class 1
(R265-71del+/+; Trp53f/f)
Prostate cancer organoids
I
R265-71del+/+;Trp53f/f prostate organoids: Cre+ vs Cre- (Hallmark GSEAs)
HALLMARK_INTERFERON_ALPHA_RESPONSE
HALLMARK_INTERFERON_GAMMA_RESPONSE
HALLMARK_E2F_TARGETS
HALLMARK_G2M_CHECKPOINT
HALLMARK_MYC_TARGETS_V1
HALLMARK_CHOLESTEROL_HOMEOSTASIS
HALLMARK_MYC_TARGETS_V2
HALLMARK_MTORC1_SIGNALING
HALLMARK_GLYCOLYSIS
HALLMARK_UNFOLDED_PROTEIN_RESPONSE
HALLMARK_HYPOXIA
HALLMARK_FATTY_ACID_METABOLISM
HALLMARK_XENOBIOTIC_METABOLISM
HALLMARK_PI3K_AKT_MTOR_SIGNALING
HALLMARK_ESTROGEN_RESPONSE_LATE
HALLMARK_KRAS_SIGNALING_DN
HALLMARK_KRAS_SIGNALING_UP
HALLMARK_ANDROGEN_RESPONSE
HALLMARK_IL2_STAT5_SIGNALING
HALLMARK_ANGIOGENESIS
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION
NES
Net enrichment score (NES)
J
R265-71del+/+; Trp53f/f
Case vs. control
Cre+ vs. Cre-
Androgen response
mTORC1 signaling
PI3K_AKT_mTOR signaling
KRAS signaling_Up
Glycolysis
E2F targets
G2M checkpoint
Myc targets_v1
Myc targets_v2
EMT
NES
100 wk intact tissue
Organoids

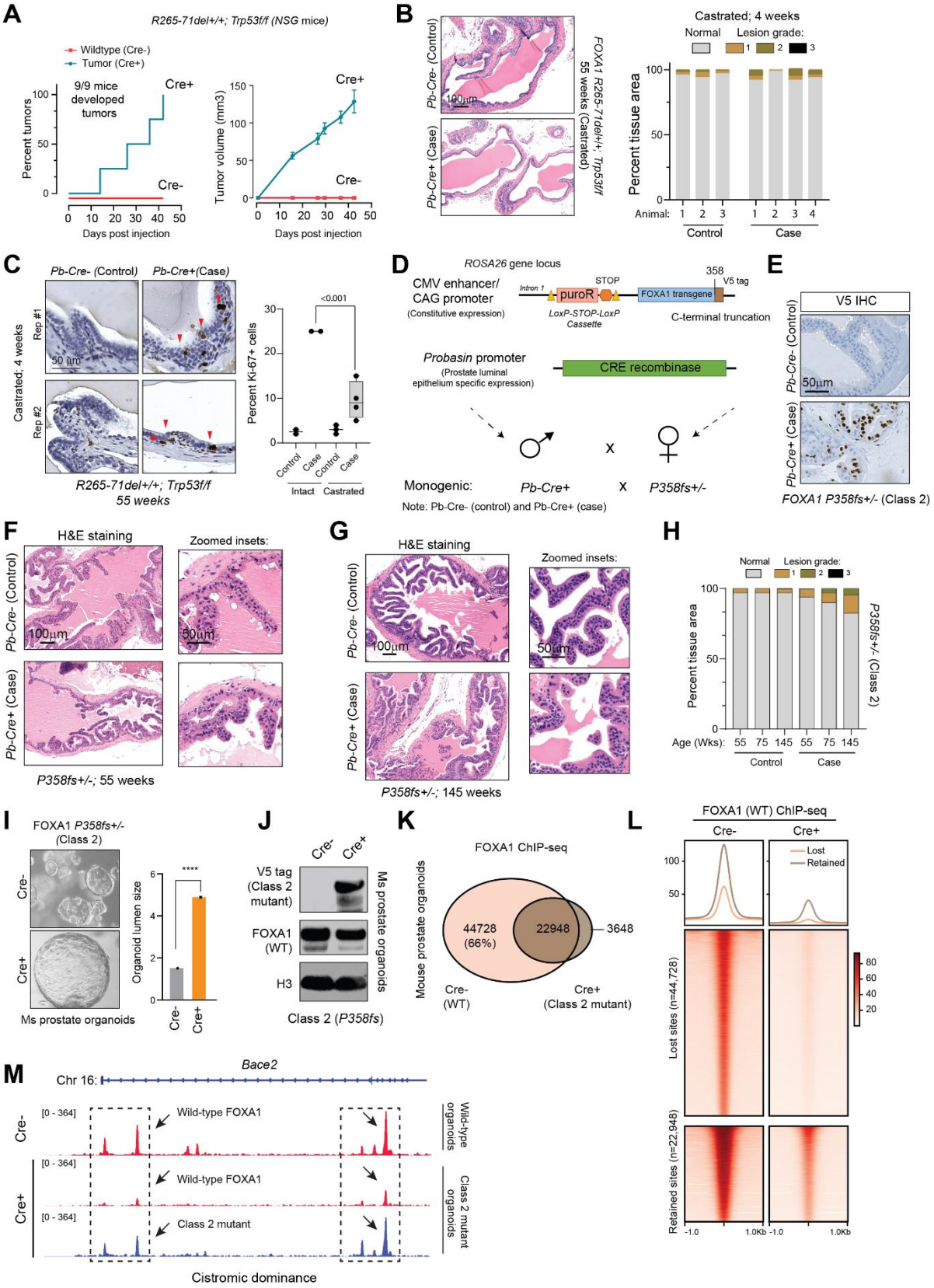
Supplementary Figure A3-4: FOXA1 Class 1 tumors are enriched for hyper-proliferative and oncogenic signatures. A) Heatmap of significantly enriched hallmark pathways from single-cell RNA-seq of *FOXA1 R265-71del+/+;Trp53f/f* tissues. Enriched pathways are clubbed by distinct cancer-relevant themes. B) Correlation plot of differentially expressed genes between the PI3K and FOXA1 class 1 mutant tumors from the TCGA primary PCa clinical cohort (n=503, pearson correlation test). C) Gene set enrichment analysis (GSEA) plots for PI3K and FOXA1 Class 1 activated genes from FOXA1 Class 1 or PI3K mutant human primary PCa samples (TCGA, n=503). DEGS, differentially expressed genes (GSEA enrichment test). D) Correlation plot of differentially expressed genes between the PI3K and FOXA1 class 1 mutant tumors from the SU2C CRPC PCa clinical cohort (n=657, pearson correlation test). E) GSEA plots of AR up-regulated genes using the fold change from Asian Class 1 mutant patient tumor tissues. DEGS, differentially expressed genes (n=206, GSEA enrichment test). F) Correlation plot of differentially expressed genes between the PI3K and FOXA1 class 1 mutant tumors from the CPGEA primary PCa clinical cohort (n=206, pearson correlation test). G) Gene set enrichment analysis (GSEA) plots for PI3K and FOXA1 Class 1 activated genes from FOXA1 Class 1 or PI3K mutant CPGEA patient tumor tissues. DEGS, differentially expressed genes (n=206, GSEA enrichment test). H) Left: Schematic of the *in vitro* generation of isogenic mouse prostate organoids. Control organoids denoted as Cre- do not express the FOXA1 mutants, while Cre+ organoids express the distinct mutations. Right: Immunoblots of noted proteins in bigenic (*R265-71del+/+;Trp53f/f)* cancer organoids. I) Heatmap of the net enrichment scores (NES) for noted hallmark pathways in the Cre+ vs Cre- *R265-71del+/+;Trp53f/f* prostate cancer organoids. J) Heatmap of NES scores showing concordance of noted cancer-relevant pathways between the FOXA1 *R265-71del+/+;Trp53f/f* tumor tissues (single-cell RNA seq) and tumor organoids.


A
FOXA1 ChIP-seq
Mouse prostate organoids
19873
35440
16422
Cre-
(WT)
Cre+
(Class 1 mutant)
B
GREAT nominated putative FOXA1 target genes (n=141 genes)
Fold-change:
Cre-
Cre+
Mouse prostate cancer organoids (R265-71del+/+; Trp53f/f)
Mouse prostate organoids R261G+/+
C
Biological pathways associated with FOXA1 putative targets
MSigDB Hallmark 2020
p-value <0.05: TRUE FALSE
Epithelial Mesenchymal Transition 8.41e-05
Hypoxia 2.87e-03
Interferon Gamma Response 2.87e-03
Interferon Alpha Response 4.82e-03
mTORC1 signaling 1.34e-02
Xenobiotic metabolism 1.34e-02
Glycolysis 1.34e-02
Apoptosis 2.67e-02
Apical surface 3.80e-02
UV response down 8.05e-02
-Log10 (p-value)
D
Cre- Cre+ Cre- Cre+
AR
V5
H3
R261G+/+
R265-71del+/+
Prostate organoids
E
AR signature (In-house)
NES: 2.102
p-val: 1.924e-07
NES:1.846
p-val: 8.297e-05
mTORC1 signaling
NES: 1.816
p-val: 0.00016
NES:1.877
p-val: 5.036e-05
Enrichment score
Class 1 (R261G+/-) vs WT ranked DEGs
Class 1 (R265-71del+/+) vs WT ranked DEGs
F
Tp53 KO
Class 1 (R261G+/-)
mTORC1
AR
MYC
E2F
G2M
p = 0.8558
p = 0.0001607
p = 3.474e-07
p = 1.924e-07
p = 1
p = 0.00237
p = 0.2981
p = 2.5e-09
p = 0.033
p = 2.5e-09
NES score
Hallmark pathways
G
Wild-type
Trp53 KO
R261G+/-; Trp53 KO
100 µm
50 µm
Mouse prostate organoids
Wild-type
R261G+/-;Trp53 KO
H
Wild-type
Trp53 KO
R261G+/-
R261G+/-; Trp53 KO
200 µm
Zoomed insets:
20 µm
I
ki67+ nuclei/100 cells
p<0.01
p<0.01
Wild-type
Trp53 KO
R261G+/-
R261G+/-; Trp53 KO
J
control (Cre-)
R261G+/-
R261G+/-; Trp53 KO
Cell viability (CellTiter-Glo)
D1
D3
D6
D7
K
Cre-
Cre+
+/-
+/+
AR
V5
Total Akt
p-Akt S473
Total S6
p-S6 (S235/236)
H3
Class 1 (R265-71del)
Class 3
L
Caucasian Primary PCa (TCGA; n=503)
p = 0.056
p = 0.021
p = 0.12
RPS6 (pS235/S236)
RPS6 (pS240/S244)
Total RPS6
Others
FOXA1 Class1
M
Asian Primary PCa (Nat. Cancer, 2024; n=125)
RPS6 phospho-S235 (spectral counts)
RPS6 phospho-S236 (spectral counts)
Total RPS6 (spectral counts)
0.072
0.789
0.023
0.031
0.423
0.05
0.598
0.795
0.247
Other
Class 1
PI3K

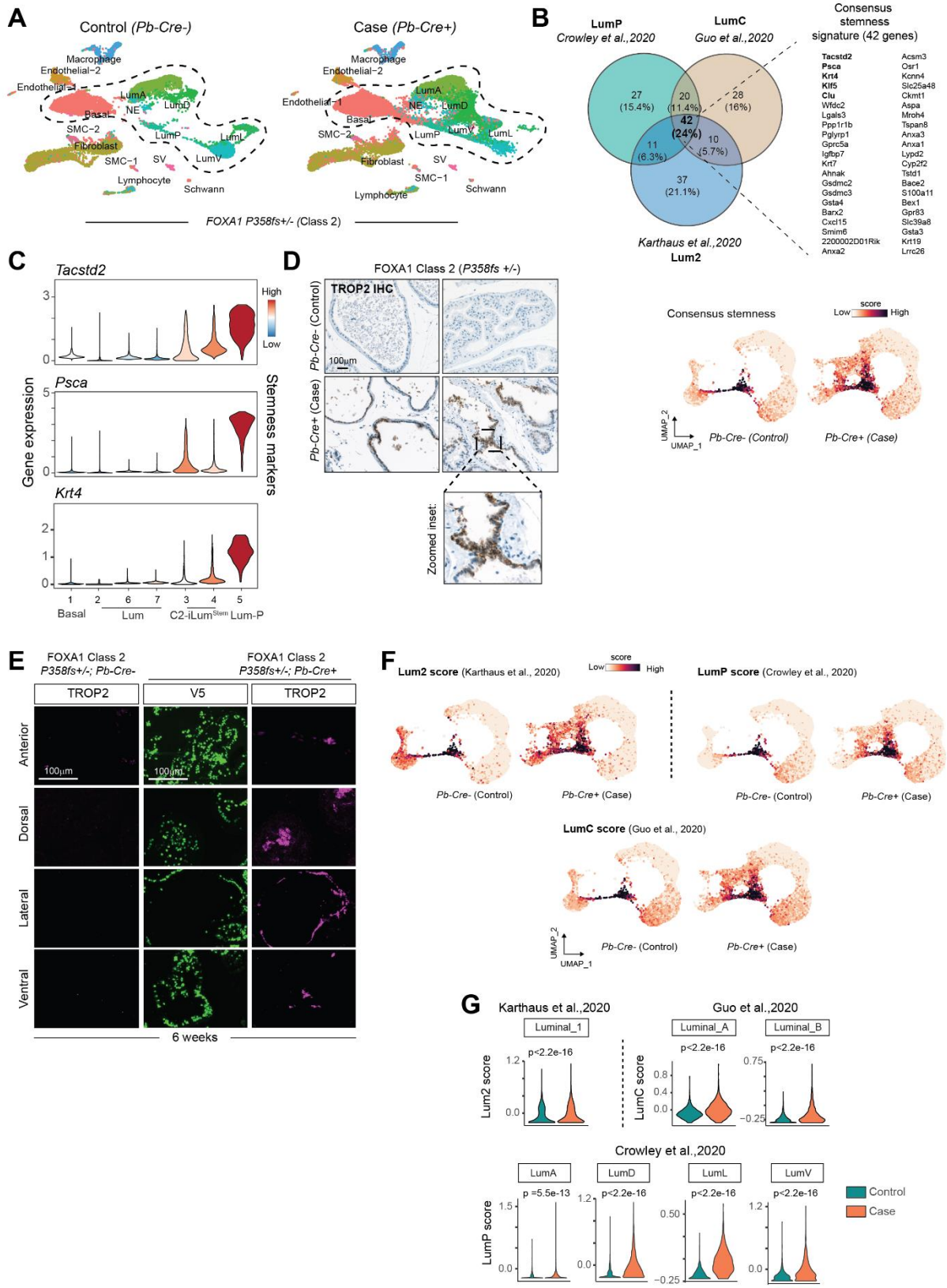
Supplementary Figure A3-5: FOXA1 Class 1 mutants reprogram AR and activate mTORC1 signaling. A) Overlap of the wild-type and Class 1 mutant FOXA1 cistrome in the Class 1 Cre- and Cre+ mouse prostate organoids. B) Fold-change expression of GREAT-nominated genes in Class 1-mutant expressing normal mouse prostate or PCa organoids (RNA-seq, n = 2 biological replicates). C) Enrichr analyses of putative FOXA1 gene targets using biology pathways in the MSigDB database (Fisher's exact test). D) Immunoblot of noted proteins in the monogenic Class 1 prostate organoids (*R265-71del, R261G*). E) GSEA plots of AR and mTORC1 up-regulated genes using the fold change from *FOXA1 R261G* (left) and *R265-71del* (right) Cre+ vs Cre- tumor organoids. DEGS, differentially expressed genes. (n=2 biological replicates, GSEA enrichment test). F) Barplots of the NES scores for the labeled hallmark pathways from RNA-seq data of *Trp53 KO* or FOXA1 Class 1 *R261G+/-* mouse prostate organoids (GSEA enrichment test). G) Representative H&E images of the wild-type, *Trp53* knock-out (KO), and *FOXA1 R261G+/-;Trp53* KO mouse prostate organoids. Scale=100μm; inset scale =50μm. H) Representative Ki67 IHC images of the distinct *R261G* isogenic mouse prostate organoids. I) Semi-quantification using percentage positivity of nuclear Ki67 expression in the four sub-cohorts shown in panel H (unpaired t-test. Box plot: center line, median; box, interquartile range (Q1–Q3); whiskers, minimum to maximum values; all individual data points shown). J) Cell growth (Cell Titer Glow) plots of the control, *R261G+/-* or *R261G+/-;Trp53 KO* organoid lines in complete media (n= 4 biological replicates). K) Immunoblots of labeled proteins in Cre- and Cre+ Class 1 (*R265-71del*, left) and Class 3 (wild-type overexpression, right) mouse prostate organoids. L) Reverse phase protein array-based expression of phosphorylated RPS6 (Serine 235 and 236) and total RPS6 in FOXA1 Class 1 mutant tumors from TCGA primary PCa samples (n=503, wilcoxon signed-rank test). M) Protein expression levels of phosphorylated RPS6 (Serine 235 and Serine 236) and total RPS6 in FOXA1 Class 1 and PI3K mutant tumors from Asian primary PCa samples (n=125, student t-test).


A
AR ChIP-seq (R261G+/+)
Cre-
Cre+
Class 1-specific
Wild-type specific
Shared
-2.5
center
2.5Kb
All sites
Percentage(%)
Promoter
5' UTR
3' UTR
1st Exon
Other Exon
1st Intron
Other Intron
Downstream
Distal intergenic
Class 1-specific
Wild-type specific
Shared
B
Wild-type specific AR sites (n=6,504)
Rank
Known motif
Name
p-value
1 GRE(NR) -5.254e+02
2 GRE(NR) -4.730e+02
3 ARE(NR) -4.711e+02
4 PGR(NR) -3.964e+02
5 PR -1.537e+02
D
Control
Case
p < 0.0001
% of CK8+ cells that are NSD2+
FOXA1 R265-71del+/+; Trp53f/f
(100 weeks)
C
AR ChIP-seq
Control
Control-specific AR binding site
CRE (+Class 1)
chr16
25.956
25.958
25.96
Mb
ARE
AR-half
FOXA1
FOX:AR
AR ChIP-seq
chr14
86.338
86.34
86.342
CRE-specific AR binding site
E
SMA
CK8
CK8 + AR + NSD2
CK8 + AR + NSD2
Zoomed insets
Invasive edge
Laminated edge
SMA-laminated
Invasive
AR
NSD2
Delamination/ invasive islands
Prostate gland
Stroma
R265-71del+/+; Trp53f/f ;Pb-Cre+ prostate gland (100 week old)
FOXA1 R265-71del+/+; Trp53f/f ; Pb-Cre+
(100 weeks, case)
F
Wild-type
NSD2 KO
AR
V5 tag (FOXA1)
NSD2-L
NSD2-S
H3
R265-71del+/+;Trp53f/f;Cre+
Ms prostate cancer organoids
G
R265-71del+/+;Trp53f/f;Cre+ (PCa organoids)
DAPI
AR
NSD2
Wild-type
NSD2 knock-out
20μm
Multiplex Immunofluorescence
H
R265-71del+/+;Trp53f/f;Cre+
DAPI
PLA (AR+NSD2)
Wild-type
NSD2 knock-out
10μm
(PCa organoids)
Proximity ligation assay (PLA)
I
AR-NSD2 PLA signal
p-val < 0.0001
PLA intensity / spot
WT
NSD2 KO
J
R265-71del+/+;Trp53f/f
CRE sgNC
CRE sgNSD2
Cellviability(CellTiter-Glo)
p < 0.0001
Day:
D0
D4
D5
D6
L
AR ChIP-seq
R265-71del+/+;Trp53f/f;Cre+
Wild-type
NSD2 knock-out
5,446 (85.9%)
1,192
2,737
Ms PCa organoids
M
Percentage sites (%)
NSD2-dependent
NSD2-independent
Feature
Promoter
5' UTR
3' UTR
1st Exon
Other Exon
1st Intron
Other Intron
Downstream
Distal Intergenic
Feature Distribution
K
Mouse AR DHT signature (in house)
Enrichment score
NES: -1.517
p-value: 0.00144
sgNSD2-tumor vs tumor organoids
(ranked DEGs)
N
AR ChIP-seq
Wild-type
NSD2 KO
NSD2-dependent
NSD2-independent
-1.0
1.0Kb
O
NSD2-dependent AR sites (Top 1000 sites)
Rank
Known motif
Name
p-val
1 FOXA1:AR -2.311e+2
2 GRE -1.740e+2
3 GRE -1.552+2
4 Foxa -1.481e+2
5 Foxa -1.264e+2
NSD2-independent AR sites (Top 1000 sites)
Rank
Known motif
Name
p-val
1 GRE -1.790e+2
2 GRE -1.603e+2
3 PGR -1.334+2
4 ARE -1.309e+2
5 Foxa -5.732e+1
P
H3K36me2 ChIP-seq
4.4e-35
1.64e-09
ns
Normalized read density (x 10⁻⁴)
WT
NSD2 KO
NSD2-dependent sites
NSD2-independent sites

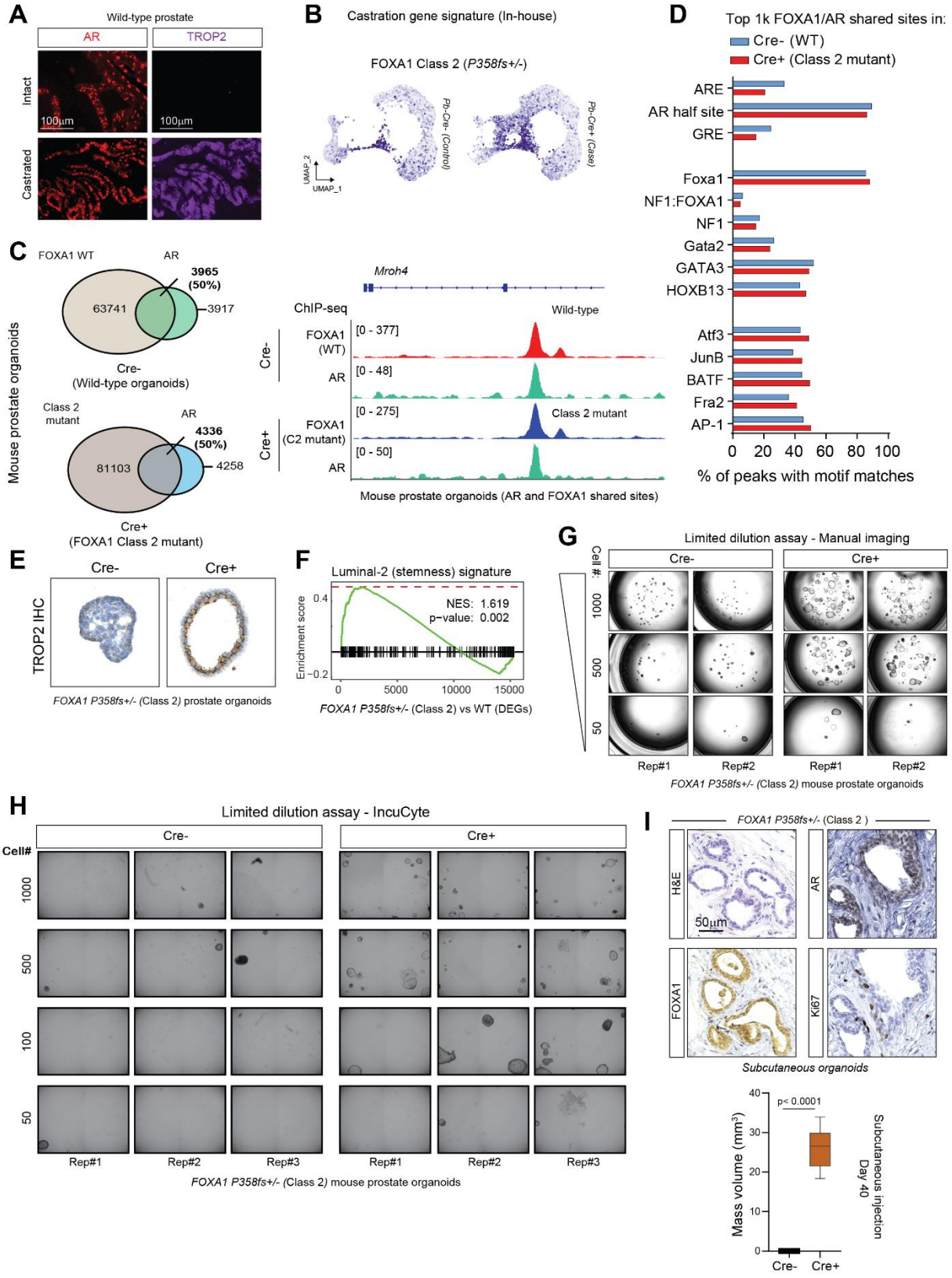
Supplementary Figure A3-6: NSD2 reprograms AR at chimeric AR half-sites. A) Left: ChIP-seq read-density heatmaps of AR at the Class 1-specific, wild-type, or shared AR binding sites in the *FOXA1 R261G+/+* mouse prostate organoids. Right: Genomic location of the wild-type, Class 1-specific, and shared AR binding sites from the heatmap in the left panel. B) Top five HOMER known motifs (ranked by p-value) enriched within Cre- i.e., wild-type specific AR binding sites. (HOMER, hypergeometric test). C) ChIP-seq read-density tracks of AR binding within loci in chromosome 16 and 14 in Cre- and Cre+ FOXA1 Class 1 *R261G+/+* prostate organoids. HOMER motifs detected within AR peaks are shown below. D) Boxplot showing the percentage of CK8+ cells that are NSD2+ cells in 100-week-old *FOXA1 R265-71del+/+;Trp53f/f* control and case tissues as quantified through multiplex-IF sections (two-sided t-test). Box plot: center line, median; box, interquartile range (Q1–Q3); whiskers, minimum to maximum values; all individual data points shown. E) Representative multiplex immunofluorescence of noted proteins in *FOXA1 R265-71del+/+;Trp53f/f* 100-week-old case tissues. Invasive edges and delaminated islands have elevated NSD2 expression. Scale =100µm. F) Immunoblot of noted proteins in NSD2 wild-type (WT) and NSD2 knock-out (KO) class 1 (*R265-71del+/+;Trp53f/f;Cre+*) tumor organoids. G) Representative multiplex immunofluorescence images of AR and NSD2 in the NSD2 WT and KO tumor organoids. H) Representative images from NSD2 and AR PLA assay in NSD2 WT and KO class 1 tumor organoids. AR-NSD2 interaction is denoted by the pink dots. I) Area-based quantification of NSD2+/AR+ PLA foci from panel H (unpaired t-test). J) Growth curves (cell titer glow) of NSD2 WT and NSD2 KO Class 1 tumor organoids (n=6 biological replicates, two-sided t-test, mean with SD is shown). K) GSEA plots of AR up-regulated genes using the fold change from NSD2 KO (sgNSD2) vs WT tumor organoids. DEGS, differentially expressed genes. (n=2 biological replicates, GSEA enrichment test). L) Overlap of AR cistrome in NSD2 WT and KO class 1 mutant tumor organoids. M) Genomic location of AR at NSD2-dependent and independent sites defined from the overlap analysis in panel L. N) ChIP-seq read density heatmaps of AR at AR enhancers in NSD2 WT and KO class 1 tumor organoids. O) Top five known HOMER motifs present within NSD2-dependent and independent AR sites in class 1 mutant tumor organoids (HOMER, hypergeometric test). P) Box plots of normalized ChIP-seq reads of H3K36me2 at NSD2-dependent and NSD2-independent AR sites in class 1 mutant tumor organoids (n = top 1000 sites, two-sided t-test). Box plot center, median; box, quartiles 1-3; whiskers, quartiles 1-3 ± 1.5 × interquartile range; dot, outliers.


A
R265-71del+/+; Trp53f/f (NSG mice)
Wildtype (Cre-)
Tumor (Cre+)
9/9 mice developed tumors
Cre+
Cre-
Percent tumors
Days post injection
Tumor volume (mm3)
B
Pb-Cre- (Control)
Pb-Cre+ (Case)
FOXA1 R265-71del+/+; Trp53f/f
55 weeks (Castrated)
100μm
Castrated; 4 weeks
Normal
Lesion grade:
Percent tissue area
Animal:
Control
Case
C
Pb-Cre- (Control)
Pb-Cre+(Case)
Castrated; 4 weeks
Rep #1
Rep #2
50 μm
R265-71del+/+; Trp53f/f
55 weeks
<0.001
Percent Ki-67+ cells
Intact
Castrated
D
ROSA26 gene locus
CMV enhancer/ CAG promoter
(Constitutive expression)
Intron 1
STOP
puroR
FOXA1 transgene
358
V5 tag
LoxP-STOP-LoxP Cassette
C-terminal truncation
Probasin promoter
(Prostate luminal epithelium specific expression)
CRE recombinase
Monogenic:
Pb-Cre+
x
P358fs+/-
Note: Pb-Cre- (control) and Pb-Cre+ (case)
E
V5 IHC
Pb-Cre- (Control)
Pb-Cre+ (Case)
50μm
FOXA1 P358fs+/- (Class 2)
F
H&E staining
Zoomed insets:
Pb-Cre- (Control)
Pb-Cre+ (Case)
100μm
50μm
P358fs+/-; 55 weeks
G
H&E staining
Zoomed insets:
Pb-Cre- (Control)
Pb-Cre+ (Case)
100μm
50μm
P358fs+/-; 145 weeks
H
Normal
Lesion grade:
Percent tissue area
P358fs+/- (Class 2)
Age (Wks):
55 75 145 55 75 145
Control
Case
I
FOXA1 P358fs+/- (Class 2)
Cre-
Cre+
Ms prostate organoids
Organoid lumen size
****
J
Cre-
Cre+
V5 tag (Class 2 mutant)
FOXA1 (WT)
H3
Ms prostate organoids
Class 2 (P358fs)
K
FOXA1 ChIP-seq
Mouse prostate organoids
44728 (66%)
22948
3648
Cre- (WT)
Cre+ (Class 2 mutant)
L
FOXA1 (WT) ChIP-seq
Cre-
Cre+
Lost
Retained
Lost sites (n=44,728)
Retained sites (n=22,948)
-1.0
1.0Kb
M
Bace2
Chr 16:
Cre-
Cre+
[0 - 364]
Wild-type FOXA1
Class 2 mutant
Wild-type organoids
Class 2 mutant organoids
Cistromic dominance

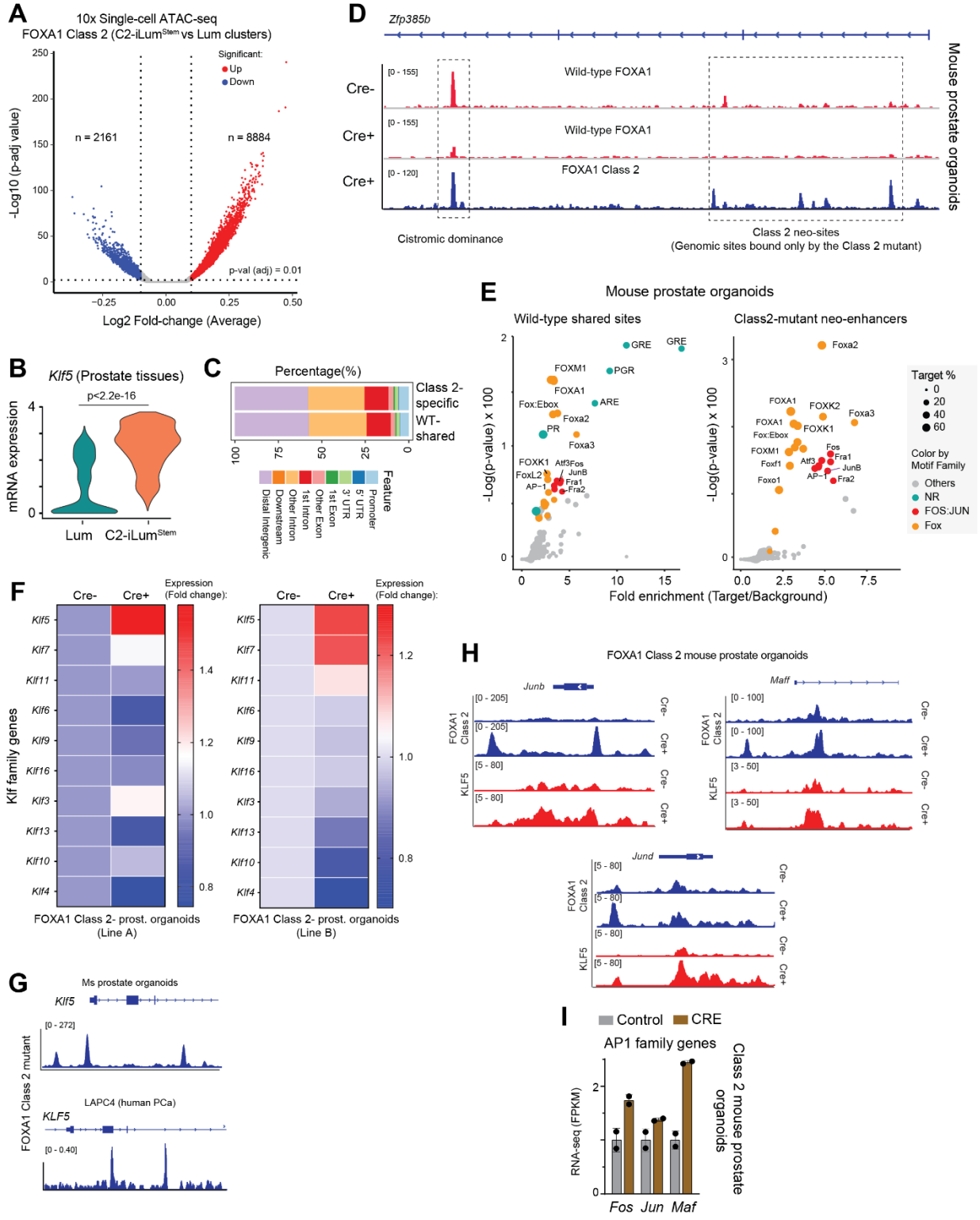
Supplementary Figure A3-7: FOXA1 Class 2 mutants cannot initiate pre-cancerous lesions. A) Left: Reverse Kaplan-Meier plot of subcutaneous organoid grafting of *R265-71del+/+;Trp53f/f* tumor organoids in *NSG* mice. Right: Tumor volumes of *R265-71del+/+;Trp53f/f* Cre- and Cre+ tumor organoids grafted in the subcutaneous flanks of *NSG* mice. B) Left: Representative H&E stained sections of control and case *R265-71del+/+;Trp53f/f* 55-week-old castrated prostate tissues. Scale =100µm.Right: Histopathological grading of tissues from the left panel. C) Left: Representative Ki67 IHC stained sections of control and case *R265-71del+/+;Trp53f/f* 55-week-old castrated prostate tissues. Red solid arrows highlight a few Ki67+ nuclei. Scale=50µm. Right: Percentage of Ki67+ cells in the *R265-71del+/+;Trp53f/f* 55-week-old intact and castrated case/control tissues (unpaired t-test). Box plot: center line, median; box, interquartile range (Q1–Q3); whiskers, minimum to maximum values; all individual data points shown. D) Schematic of *FOXA1* Class 2 (*P358fs*) knock-in transgenic mouse model and mating strategies for monogenic (*P358fs+/-*) mouse lines. Probasin-Cre (*Pb-Cre*) ensures prostate-specific expression of FOXA1. Control and case animals are denoted as *Pb-Cre-* and *Pb-Cre+*, respectively. E) Representative V5-epitope tag IHC in control (*Pb-Cre-*) and case (*Pb-Cre+*) Class 2 animals. Scale =50µm. F) Representative H&E-stained sections of 55-week-old FOXA1 *P358fs+/-* case and control tissues. Scale =100µm. G) Representative H&E-stained sections of 145-week-old FOXA1 *P358fs+/-* case and control tissues. Scale =100µm. H) Histopathological grading of 55, 75, and 145-week-old FOXA1 *P358fs+/-* case and control tissues. I) Left: Representative images of *FOXA1 P358fs+/-* Cre- and Cre+ mouse prostate organoids. Cre- (control) organoids lack expression of the Class 2 mutation, while Cre+ (case) organoids express the Class 2 mutation. Right: Lumen size of the Cre- and Cre+ FOXA1 *P358fs+/-* organoids (unpaired t-test). J) Immunoblots of noted proteins in the Cre- and Cre+ *FOXA1 P358fs+/-* organoids. C-terminal truncated FOXA1 Class 2 protein can be detected with the V5-epitope tag antibody. K) Overlap of the wild-type FOXA1 cistrome in the Cre- and Cre+ Class 2 mutant mouse prostate organoids. L) ChIP-seq read-density heatmaps of FOXA1 wild-type at lost and retained sites in Class 2 *P358fs+/-* Cre- and Cre+ organoids. M) ChIP-seq read-density tracks of the wild-type and Class 2 mutant in Cre- and Cre+ FOXA1 *P358fs+/-* mouse prostate organoids. The black dotted box highlights regions of cistromic dominance.


A
Control (Pb-Cre-)
Case (Pb-Cre+)
Macrophage
Endothelial−2
Endothelial−1
LumA
NE
LumD
Basal
SMC−2
LumP
LumL
Fibroblast
SMC−1
SV
LumV
Lymphocyte
Schwann
FOXA1 P358fs+/- (Class 2)
B
LumP
Crowley et al.,2020
LumC
Guo et al.,2020
27 (15.4%)
20 (11.4%)
28 (16%)
42 (24%)
11 (6.3%)
10 (5.7%)
37 (21.1%)
Karthaus et al.,2020
Lum2
Consensus stemness signature (42 genes)
Tacstd2
Psca
Krt4
Klf5
Clu
Wfdc2
Lgals3
Ppp1r1b
Pglyrp1
Gprc5a
Igfbp7
Krt7
Ahnak
Gsdmc2
Gsdmc3
Gsta4
Barx2
Cxcl15
Smim6
2200002D01Rik
Anxa2
Acsm3
Osr1
Kcnn4
Slc25a48
Ckmt1
Aspa
Mrohh4
Tspan8
Anxa3
Anxa1
Lypd2
Cyp2f2
Tstd1
Bace2
S100a11
Bex1
Gpr83
Slc39a8
Gsta3
Krt19
Lrrc26
C
Tacstd2
Psca
Krt4
High
Low
Gene expression
Stemness markers
Basal
Lum
C2-iLumStem
Lum-P
D
FOXA1 Class 2 (P358fs +/-)
TROP2 IHC
Pb-Cre- (Control)
Pb-Cre+ (Case)
100µm
Zoomed inset:
Consensus stemness
score
Low
High
UMAP_2
UMAP_1
Pb-Cre- (Control)
Pb-Cre+ (Case)
E
FOXA1 Class 2
P358fs+/-; Pb-Cre-
FOXA1 Class 2
P358fs+/-; Pb-Cre+
TROP2
V5
TROP2
Anterior
Dorsal
Lateral
Ventral
100µm
6 weeks
F
Lum2 score (Karthaus et al., 2020)
score
Low
High
LumP score (Crowley et al., 2020)
Pb-Cre- (Control)
Pb-Cre+ (Case)
LumC score (Guo et al., 2020)
UMAP_2
UMAP_1
G
Karthaus et al.,2020
Guo et al.,2020
Luminal_1
Luminal_A
Luminal_B
p<2.2e-16
Lum2 score
LumC score
Crowley et al.,2020
LumA
LumD
LumL
LumV
p =5.5e-13
LumP score
Control
Case

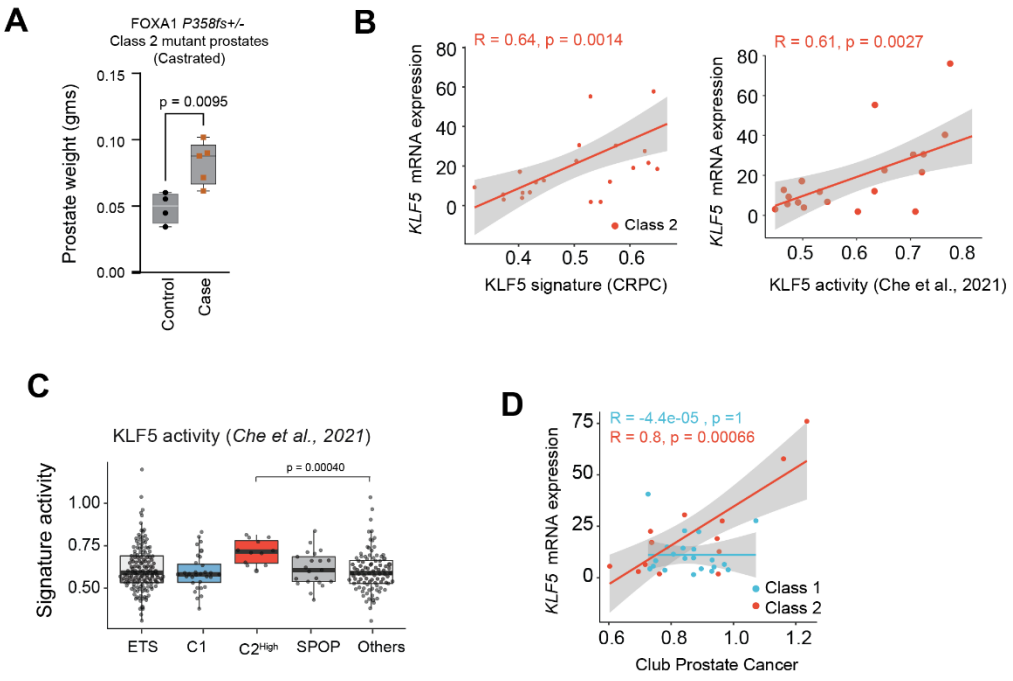
Supplementary Figure A3-8: Cistromically dominant FOXA1 Class 2 mutants activate stemness gene programs. A) Split UMAP plots of the distinct cell types from 10x single-cell RNA-sequencing data of FOXA1 *P358fs+/-* case and control animals. B) Top: Venn diagram showing the overlap of genes across three studies representing Luminal-P (*Crowley et al., 2020)*, Luminal-2 (*Karthaus et al., 2020*), and Luminal-C (*Guo et al., 2020*) stem-like clusters. The consensus stemness signature score is a 42-gene list. Bottom: Split UMAP plots with case or control tissue-derived cells colored by the consensus stemness score defined in the top panel. C) Violin plots of select stemness genes (single cell multiome gene expression, imputed) in the distinct Class 2 mutant case and control clusters. D) Representative TROP2 IHC in lateral and ventral prostatic lobes of FOXA1 *P358fs+/-* case and control tissue sections. Scale =100µm. E) Representative multiplex-IF of TROP2 and V5 in 6-week-old FOXA1 *P358fs+/-* case and control prostate lobes. Scale =100µm. F) Split UMAP plots of the case or control tissue-derived cells colored by the Lum2, LumP, and LumC scores. G) Violin plots of the LumP, LumC, and Lum2 scores in Class 2 mutant case and control luminal cells split by subpopulations annotated using three public datasets as reference (p-values obtained from Wilcoxon test).


A
Wild-type prostate
AR
TROP2
Intact
Castrated
100μm
B
Castration gene signature (In-house)
FOXA1 Class 2 (P358fs+/-)
Pb-Cre- (Control)
Pb-Cre+ (Case)
UMAP_2
UMAP_1
C
FOXA1 WT
AR
3965
(50%)
63741
3917
Cre-
(Wild-type organoids)
Mouse prostate organoids
Class 2
mutant
AR
4336
(50%)
81103
4258
Cre+
(FOXA1 Class 2 mutant)
Mroh4
ChIP-seq
Wild-type
Cre-
FOXA1
(WT)
[0 - 377]
AR
[0 - 48]
Cre+
FOXA1
(C2 mutant)
[0 - 275]
Class 2 mutant
AR
[0 - 50]
Mouse prostate organoids (AR and FOXA1 shared sites)
D
Top 1k FOXA1/AR shared sites in:
Cre- (WT)
Cre+ (Class 2 mutant)
ARE
AR half site
GRE
Foxa1
NF1:FOXA1
NF1
Gata2
GATA3
HOXB13
Atf3
JunB
BATF
Fra2
AP-1
0
20
40
60
80
100
% of peaks with motif matches
E
Cre-
Cre+
TROP2 IHC
FOXA1 P358fs+/- (Class 2) prostate organoids
F
Luminal-2 (stemness) signature
Enrichment score
0.4
−0.2
NES: 1.619
p-value: 0.002
0
5000
10000
15000
FOXA1 P358fs+/- (Class 2) vs WT (DEGs)
G
Limited dilution assay - Manual imaging
Cre-
Cre+
Cell #:
1000
500
50
Rep#1
Rep#2
Rep#1
Rep#2
FOXA1 P358fs+/- (Class 2) mouse prostate organoids
H
Limited dilution assay - IncuCyte
Cre-
Cre+
Cell#
1000
500
100
50
Rep#1
Rep#2
Rep#3
Rep#1
Rep#2
Rep#3
FOXA1 P358fs+/- (Class 2) mouse prostate organoids
I
FOXA1 P358fs+/- (Class 2)
H&E
50μm
AR
FOXA1
Ki67
Subcutaneous organoids
p< 0.0001
Mass volume (mm³)
0
10
20
30
40
Cre-
Cre+
Subcutaneous injection
Day 40

Supplementary Figure A3-9: FOXA1 Class 2 mutants drive higher stemness potential. A) Multiplex IF images of AR and TROP2 in intact and castrated wild-type mouse prostate tissues. Scale =100µm. B) Split UMAP plots of the case or control tissue-derived cells colored by the castration signature (defined in-house). C) Left: Overlap analysis of the AR cistrome with wild-type FOXA1 (top) and FOXA1 class 2 mutant (bottom) in Class 2 Cre- and Cre+ mouse prostate organoids. Right: ChIP-seq read-density tracks of AR and FOXA1 wild-type and Class 2 mutant proteins in FOXA1 *P358fs*+/- Cre- and Cre+ Class 2 organoids. D) Percentage of shared AR/FOXA1 wild-type sites or shared AR/FOXA1 class-2 mutant binding sites defined from panel c, containing known motifs of the labelled FOXA1 or AR cofactors. E) Representative TROP2 IHC positivity noted in FOXA1 *P358fs*+/- Cre- and Cre+ mouse prostate organoids. F) GSEA plots of the Luminal 2 up-regulated genes using the fold change from FOXA1 *P358fs*+/- Cre+ vs Cre- organoids. DEGS, differentially expressed genes. (n=2 biological replicates, GSEA enrichment test). G) Representative manual brightfield images of FOXA1 *P358fs*+/- Cre+ and Cre- organoids cultured in a limited dilution assay. Cells seeded per well are denoted on the left. H) Representative IncuCyte brightfield images of FOXA1 *P358fs*+/- Cre+ and Cre- organoids cultured in a limited dilution assay. Cells seeded per well are denoted on the left. I) Top: Representative H&E and AR, FOXA1, and Ki67 IHC stained sections in subcutaneous FOXA1 *P358fs*+/- Cre+ organoid masses formed in *CB17/SCID* mice. Scale =50µm. Bottom: Mass volumes of Cre- and Cre+ organoids as measured at Day 40 (unpaired t-test).


A
10x Single-cell ATAC-seq
FOXA1 Class 2 (C2-iLumStem vs Lum clusters)
Significant:
Up
Down
n = 2161
n = 8884
p-val (adj) = 0.01
-Log10 (p-adj value)
Log2 Fold-change (Average)
B
Klf5 (Prostate tissues)
p<2.2e-16
mRNA expression
Lum
C2-iLumStem
C
Percentage(%)
Class 2-specific
WT-shared
Feature
Distal Intergenic
Downstream
Other Intron
1st Intron
Other Exon
1st Exon
5' UTR
3' UTR
Promoter
D
Zfp385b
Wild-type FOXA1
Wild-type FOXA1
FOXA1 Class 2
Cre-
Cre+
Cre+
Mouse prostate organoids
Cistromic dominance
Class 2 neo-sites
(Genomic sites bound only by the Class 2 mutant)
E
Mouse prostate organoids
Wild-type shared sites
Class2-mutant neo-enhancers
-Log(p-value) x 100
Fold enrichment (Target/Background)
Target %
Color by Motif Family
Others
NR
FOS:JUN
Fox
F
Cre-
Cre+
Expression (Fold change):
Klf family genes
FOXA1 Class 2- prost. organoids (Line A)
FOXA1 Class 2- prost. organoids (Line B)
G
Ms prostate organoids
Klf5
FOXA1 Class 2 mutant
LAPC4 (human PCa)
KLF5
H
FOXA1 Class 2 mouse prostate organoids
Junb
Maff
Jund
FOXA1 Class 2
KLF5
I
Control
CRE
AP1 family genes
RNA-seq (FPKM)
Class 2 mouse prostate organoids
Fos
Jun
Maf

Supplementary Figure A3-10: FOXA1 Class 2 mutants pioneer neo-enhancers co-bound by KLF5 and AP-1 family proteins. A) Volcano plot showing differentially accessible (upregulated and downregulated) ATAC-peaks from 10x single-cell multiome ATAC seq of FOXA1 *P358fs+/-* control and case-specific clusters (logistic regression test). B) Violin plot of *Klf5* expression in Class 2 mutant tissues (single-cell multiome gene expression, Wilcoxon test). C) Genomic location of wild-type-shared and Class 2 mutant binding sites defined from the FOXA1 ChIP-seq in mouse prostate organoids. D) ChIP-seq read-density tracks of the wild-type and Class 2 mutant in Cre- and Cre+ FOXA1 *P358fs+/-* mouse prostate organoids. The black dotted box highlights the Class 2 neo-sites, which are genomic regions bound explicitly by FOXA1 Class 2 mutant. E) Fold change and significance of HOMER motifs enriched within the wild-type shared sites and Class 2 neo-sites from FOXA1 ChIP-seq in mouse prostate organoids (HOMER, hypergeometric test). F) Fold-change heatmap of the *Klf* family of genes in two independent FOXA1 *P358fs+/-* Cre- and Cre+ mouse prostate organoid lines. G) ChIP-seq read-density tracks of the FOXA1 Class 2 mutant within the *Klf5* locus in Cre+ Class 2 mouse prostate organoids (top) and the human LAPC4 PCa cell line (bottom). H) ChIP-seq read-density tracks of the FOXA1 Class 2 mutant and KLF5 within noted genes in FOXA1 *P358fs+/-* Cre- and Cre+ Class 2 organoids. I) Barplots showing RNA-seq expression of AP-1 family genes in Class 2 Cre- and Cre+ mouse prostate organoids (RNA-sequencing, n=2 replicates).



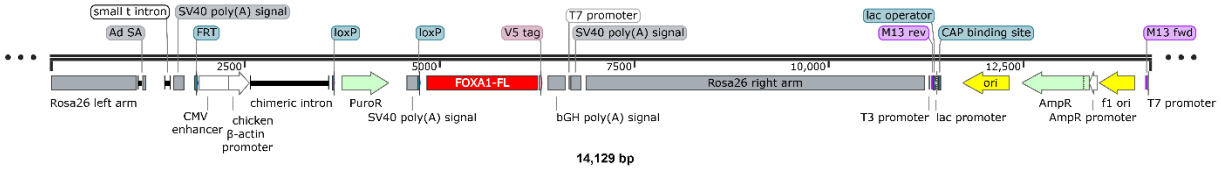
Supplementary Figure A3-11: FOXA1 Class 2 mutants activate KLF5 in clinical datasets. A) Boxplots of prostate weights from FOXA1 *P358fs*+/- control (n=4) and case (n=5) castrated prostate tissues (unpaired t-test). B) Correlation plots of *KLF5* transcript expression and two distinct KLF5 gene signatures (defined in Che et al.,2021) in FOXA1 Class 2 CRPC tissues (SU2C, n=371, pearson correlation test). C) Boxplots showing the KLF5 gene signature (defined in Che et al., 2021) in CRPC patient tissues (SU2C, n=371 patients) across distinct genomic driver groups (Wilcoxon rank-sum test). D) Correlation plots of *KLF5* transcript expression and a "Club Prostate Cancer score" in FOXA1 Class 2 CRPC tissues from the SU2C cohort (n=371 patients, pearson correlation test).

Appendix C

Supplementary notes for Chapter 3

Vector design of the FOXA1 transgenic mice

Detailed protocol for the generation of FOXA1 transgenic mice is provided in the methods section of Chapter 3. However, I would like to provide additional technical details, including vector design and mosaic transgene expression, in this section. To engineer the FOXA1 mutant mice, we knocked in a plasmid containing a human FOXA1 construct into the ROSA26 locus, a genomic "safe harbor" that is widely used for targeted transgene integration in mice. This strategy does not perturb the endogenous Foxa1 locus, which is important given FOXA1's essential role in normal prostate development. The detailed schematic of the construct used is shown in Appendix Figure C-1. The targeting vector contains a loxP-flanked STOP cassette upstream of the transgene that creates a translational block to prevent transgene expression prior to Cre recombination. The FOXA1 coding sequence is tagged with a V5 epitope, which is used in our immunohistochemical staining to detect transgene expression. Additionally, the construct also has a puromycin resistance (PuroR) gene for efficient selection in 2D cultures. The puroR gene is poly-adenylated, allowing us to confirm recombination in assays such as single-cell RNA-sequencing, where puroR-negative cells correspond to those expressing the human FOXA1-transgene.



Appendix Figure. C- 1: Schematic of the construct used to engineer the FOXA1 transgenic mouse lines.

Mosaic transgene expression

In this model, V5-tagged FOXA1 mutant transgene expression is driven by the composite CAG promoter, which contains the CMV enhancer and the chicken Beta-actin promoter. This is a widely used promoter that is known to provide strong, ubiquitous transgene expression *in vitro and in vivo*. Prostate-specific expression of FOXA1 mutants was achieved by breeding these animals with a Probasin promoter-driven CRE recombinase line. Given that FOXA1 mutations found in patients are often heterozygous, we initially generated mice harboring a single mutant allele of the human FOXA1 mutation (denoted as +/-). However, immunohistochemical staining using the V5 antibody revealed low and mosaic transgene expression in the prostate tissue. Hence, to enhance FOXA1 expression, we created mice with bi-allelic FOXA1 mutations (denoted as +/+). Although we were able to achieve increased overall expression, mosaic staining remained, likely suggesting incomplete or variable activation across the prostate epithelial cells. This pattern of mosaic expression has been previously reported for the CAG promoter, likely due to epigenetic silencing of the CAG promoter [1–4]. Such mosaicism presents a technical limitation that can delay or obscure phenotypes in transgenic animals and should be considered in future models. Alternative promoters such as EF-1 have been shown to have enhanced transgene expression, relative to CAG, and may need to be explored in the future. Nevertheless, despite the mosaicism, our FOXA1 transgenic mice presented robust phenotypes, highlighting the biological relevance of the FOXA1 mutations. Notably, in many instances, transformed regions were often adjacent to benign areas,

with strong FOXA1 expression seen only in the transformed regions and absent in the adjacent benign cells, highlighting the specificity of the phenotype.

References:

1. Moeller MJ, Soofi A, Sanden S, Floege J, Kriz W, Holzman LB. An efficient system for tissue-specific overexpression of transgenes in podocytes in vivo. Am J Physiol Renal Physiol. 2005;289: F481–8.

2. Furth PA, St Onge L, Böger H, Gruss P, Gossen M, Kistner A, et al. Temporal control of gene expression in transgenic mice by a tetracycline-responsive promoter. Proc Natl Acad Sci U S A. 1994;91: 9302–9306.

3. Schmidt EV, Christoph G, Zeller R, Leder P. The cytomegalovirus enhancer: a pan-active control element in transgenic mice. Mol Cell Biol. 1990;10: 4406–4411.

4. Dou Y, Lin Y, Wang T-Y, Wang X-Y, Jia Y-L, Zhao C-P. The CAG promoter maintains high-level transgene expression in HEK293 cells. FEBS Open Bio. 2021;11: 95–104.

Cell lines	**LLC0150 IC50 (nM)**	**Cell type**	**Lineage**	**NSD2 altered**
RS4;11	0.274	Cancer	Lymphoid	Yes
MOLT3	0.4198	Cancer	Lymphoid	No
RPMI-8402	0.43	Cancer	Lymphoid	Yes
MM1.S	3.273	Cancer	Lymphoid	Yes
H1048	4.053	Cancer	Lung	No
MM1.R	13.39	Cancer	Lymphoid	Yes
TC-205	18.45	Cancer	Primitive Neuro-Ectodermal Tumors (PNET)	No
H1836	24.21	Cancer	Lung	No
TC32	28.27	Cancer	Primitive Neuro-Ectodermal Tumors (PNET)	No
MOLT4	32.95	Cancer	Lymphoid	No
DoHH2	35.23	Cancer	Lymphoid	No
CHLA-9	35.91	Cancer	Primitive Neuro-Ectodermal Tumors (PNET)	No

CHLA-258	42.38	Cancer	Primitive Neuro-Ectodermal Tumors (PNET)	No
SUM185PE	44.68	Cancer	Breast	No
NB-1643	53.03	Cancer	Neuroblastoma	No
CB-AGPN	55.05	Cancer	Primitive Neuro-Ectodermal Tumors (PNET)	No
LASCPC-01	55.22	Cancer	Prostate	No
CAMA-1	58.19	Cancer	Breast	No
SEM	61.77	Cancer	Lymphoid	Yes
R-CHACV	69.68	Cancer	Lymphoid	Yes
T47D	78.05	Cancer	Breast	No
VCaP	84.84	Cancer	Prostate	No
MDA-PCa-2b	87.56	Cancer	Prostate	No
KARPAS-25	105.9	Cancer	Lymphoid	No
COG-N-561	116.1	Cancer	Neuroblastoma	No
CHLA-218	139.5	Cancer	Primitive Neuro-Ectodermal Tumors (PNET)	No

MDA-MB-330	191.8	Cancer	Breast	No
H446	218.1	Cancer	Lung	No
PA-1	232.9	Cancer	Ovary	No
MFM-223	237.8	Cancer	Breast	No
WA-72-PS	244.6	Cancer	Breast	No
22RV1	244.655	Cancer	Prostate	No
LNCaP	251.6	Cancer	Prostate	No
MDA-MB-468	273.9	Cancer	Breast	No
CWRR1	280.6	Cancer	Prostate	No
MDA-MB-453	294.2	Cancer	Breast	No
N87	304.4	Cancer	Stomach	No
IMR90	411	Normal	Lung	No
WA-72-As	443.5	Cancer	Breast	No
COG-E-352	546.6	Cancer	Primitive Neuro-Ectodermal Tumors (PNET)	No
H524	549	Cancer	Lung	No
NTERA-2	560.5	Cancer	Testis	No

LAPC4	631.2	Cancer	Prostate	No
ZR-75-1	673.1	Cancer	Breast	No
CHLA-99	813.7	Cancer	Ewing sarcoma	No
CHLA-25	907.5	Cancer	Primitive Neuro-Ectodermal Tumors (PNET)	No
Caov-3	966.4	Cancer	Ovary	No
MEG-01	1425	Cancer	Esophagus	No
BEAS-2B	1769	Normal	Lung	No
COLO205	1776	Cancer	Colon	No
MDA-MB-231	1856	Cancer	Breast	No
MDA-MB-436	2136	Cancer	Breast	No
5637	2168	Cancer	Bladder	No
H211	2200	Cancer	Lung	No
SAOS-2	2228	Cancer	Bone	No
HCC1143	2600	Cancer	Breast	No
HCC1146	3232	Cancer	Breast	No
TC-106	3295	Cancer	Ewing sarcoma	No

TC-138	3397	Cancer	Ewing sarcoma	No
SW626	3950	Cancer	Stomach	No
HeLa	4336	Cancer	Cervical	No
HCC1187	5183	Cancer	Breast	No
MDA-MB-415	5697	Cancer	Breast	No
957/hTERT	5942	Normal	Prostate	No
COG-N-557	7156	Cancer	Neuroblastoma	No
SNU1079	8446	Cancer	Liver/Bile duct	No
SNU387	8537	Cancer	Liver/Bile duct	No
BJ	8874	Normal	Skin	No
HK2	9840	Normal	Kidney	No
PC3	11680	Cancer	Prostate	No
RPB1293	11913	Cancer	Liver/Bile duct	No
H69	13753	Cancer	Lung	No
RWPE1	13807	Normal	Prostate	No
A673	21900	Cancer	Ewing sarcoma	No
TE6	22072	Cancer	Esophagus	No
SK-N-MC	24696	Cancer	Ewing sarcoma	No

HEK293FT	31102	Normal	Kidney	No
SCaBER	33131	Cancer	Bladder	No
UM-UC3	36186	Cancer	Bladder	No
PrECs	37608	Normal	Prostate	No
OC-8	38337	Cancer	Ovary	No
HCC1428	38766	Cancer	Breast	No
RPB1292	39080	Cancer	Liver/Bile duct	No
SNU16	42702	Cancer	Stomach	No
SNU423	47579	Cancer	Liver/Bile duct	No
HEPG2	52168	Cancer	Liver/Bile duct	No
PNT2	57996	Normal	Prostate	No
KATO III	59571	Cancer	Stomach	No
SK-OV-3	62327	Cancer	Ovary	No
PLC/PRF/5	68569	Cancer	Liver/Bile duct	No
COG-N-529	72257	Cancer	Neuroblastoma	No
MCF7	77479	Cancer	Breast	No
RPMI8226	83089.5	Cancer	Lymphoid	No
K562	86631	Cancer	Myeloid	No

DU145	93345	Cancer	Prostate	No
786-O	97532	Cancer	Kidney	No
AGS	98345	Cancer	Stomach	No
SNU840	104019	Cancer	Ovary	No
MCF10A	107036	Normal	Breast	No
RWPE2-W99	121661	Normal	Prostate	No
HT115	134497	Cancer	Colon	No
CADO-ES-1	144497	Cancer	Ewing sarcoma	No
SNU-5	178969	Cancer	Stomach	No
DAN-G	185156	Cancer	Pancreas	No
SK-HEP-1	199830	Cancer	Liver/Bile duct	No
LHSAR	201625	Normal	Prostate	No
SK-MEL-5	219126	Cancer	Melanoma	No
WPMY1	226137	Normal	Prostate	No
ACHN	307547	Cancer	Kidney	No
U2OS	375673	Cancer	Bone	No
TC-71	400363	Cancer	Ewing sarcoma	No
H716	558132	Cancer	Colon	No

LAMA87	Undetermined	Cancer	Myeloid	No
A549	Undetermined	Cancer	Lung	No

Supplementary Table 1: Antiproliferative half-maximal inhibitory concentration (IC50) of LLC0150 across human-derived normal and cancer cell lines.

Primers	**Forward**	**Reverse**
GAPDH	TGCACCACCAACTGCTTAGC	GGCATGGACTGTGGTCATGAG
FKBP5	TCTCATGTCTCCCCAGTTCC	TTCTGGCTTTCACGTCTGTG
TMPRSS2	CAGGAGTGTACGGGAATGTGATGGT	GATTAGCCGTCTGCCCTCATTTGT
AR	CAGTGGATGGGCTGAAAAAT	GGAGCTTGGTGAGCTGGTAG
NKX3-1	ACGTCCTTCCTCATCCAGGACA	AGGGCGCCTGAAGTGTTTTC
SLC45A3	TCGTGGGCGAGGGGCTGTA	CATCCGAACGCCTTCATCATAGTGT
ZBTB16	CAGTTTTCGAAGGAGGATGC	CCCACACAGCAGACAGAAGA

KLK3	ACGCTGACAGGGGGCAAAAG	GGGCAGGGCACATGGTTCACT
KLK2	GGCTCTGGACAGGTGGTAAA	GCCCCATGATGTGATACC TT

Supplementary Table 2: Primer sequences used in the study from Chapter 2.

sgNC#1:5'-GTAGCGAACGTGTCCGGCGT-3';
sgNC#2:5'-GACCGGAACGATCTCGCGTA-3';
sgNSD2-SET:5’-CACCGAAAGTCCAGATCTACACAG-3’;
sgNSD2-Long:5’-CACCGCCCCGTTGCCTATCACAGCG-3’;
sgNSD2: 5’- ACAACGTTGGTGATTTGGTG-3’.
sgNSD1#1:5’-CACCGAGTGGCGTAGACTTTTTCTT-3’; sgNSD1#2:5’-CACCGCTATCGGCAGTACTACGTGG-3’;
sgNSD1#3:5’-CACCGTATGTGAGTTACTAATGAG-3’;
sgNSD1#4:5’-CACCGTAAGAGAGTTCCTTCAACT-3’;
sgNSD3#1: 5’-CACCGGATACCCATTTGTGAGTGG-3’;
sgNSD3#2: 5’-CACCGTTTGTGGTACCGAAGGAGG-3’

Supplementary Table 3: sgRNA sequences used in the study from Chapter 2.

Gene	**5-3' sequence (Pool of 4 siRNAs)**	**Catalog #**
Non-targeting control	UGGUUUACAUGUCGACUAA,UGGUUUACAUGUUGUGUGA,UGGUUUACAUGUUUUCUGA,UGGUUUACAUGUUUUCCUA	D-001810-10-05
NSD1	GGACGAGAAUUCUUUGAUU,GAACAGAAGUAGUACCAAU,GAUCAAAGCCUUCAUCCAA,GCCGAGAGCUGUUGAGAAA	L-007048-00-0005
NSD2	GGUCCAAAGUGUCGGGUUA,UGUCAGUGGAGGAGCGGAA,GGAGCAGGGCCUUGUCGAA,CCGGGUGUUUAAUGGAGAA	L-006571-01-0005
NSD3	GGUUGACACUGUAUCAGAA, GAACGUGCUCAGUGGGAUA, GCUUGAGGUUCAUACUAAA, GUCCACUGGUGUUAAGUUU	L-012875-00-0005

Supplementary Table 4: siRNA sequences used in the study from Chapter 2.

Adgrv1
Adh6a
Adh7
Adrb1
Aif1l
Ak4
Aldh1a7
Ankrd35
Anks1b
Antxr2
Arhgap26
Arhgef38
Aspa
Atoh8
Atp12a
Atp6v0e2
Bend5
Bend7
Best2
Bst1
C2cd4b
Ces2e
Chl1
Clic5

Cmpk2
Cxxc4
Cyp2s1
Cyp7b1
Dab2
Dach1
Dagla
Ddit4l
Ddx60
Dnah11
Dock4
Dpysl3
Dtna
Egln3
Egr3
Elobl
Emilin2
Enpp1
Epha7
Etv5
Exph5
Fbln1
Fhit
Flt1
Fmo3
Fn1
Gas7

Gata2
Gem
Gfra4
Glrp1
Gm43552
Gulp1
Hcn1
Hs3st1
Ido1
Ifi211
Ifit1
Ifit3b
Igfbp5
Igtp
Il20rb
Iqgap2
Irs2
Isg15
Isg20
Kcns3
Kctd12
Kif5c
Klhl38
Krt20
Krt75
Lars2
Lrrc8c

Nckap5
Ngfr
Npnt
Npr3
Ntng1
Nts
Nup210
Nxnl2
Nxpe2
Nxpe4
Nxph3
Oasl1
Oasl2
Olfr1314
Otop1
Pde1c
Pkia
Plscr4
Plxdc2
Pmaip1
Ppp1r3b
Pthlh
Rab27a
Rap1gap2
Rasl11b
Rbp1
Sema3e

Sh3gl2
Skint8
Slc16a12
Slc1a3
Slc24a3
Slc4a11
Slc7a11
Slc9a9
Slco3a1
Sntb1
Sorbs1
Sp100
Sptssb
Styk1
Sulf2
Synm
Syt7
Sytl3
Tenm2
Timp3
Tnfrsf8
Tnfsf4
Trib2
Ttn
Ugt8a
Vgll3
Vldlr

Vwa2
Wdr72
Wfdc2
Wif1
Zfhx4
Zfp365
Zfp608
Zfpm2

Supplementary Table 5: Putative FOXA1 Class 1 mutant gene targets in mouse prostate organoids.