12-2018

INTEGRATIVE BIOINFORMATICS APPROACHES TO ELUCIDATING PROSTATE CANCER CELL HETEROGENEITY, PLASTICITY, AND TREATMENT RESPONSE

Hsueh-Ping Chao

INTEGRATIVE BIOINFORMATICS APPROACHES TO ELUCIDATING PROSTATE CANCER CELL HETEROGENEITY, PLASTICITY, AND TREATMENT RESPONSE

A
Dissertation
Presented to the Faculty of
The University of Texas
MD Anderson Cancer Center UTHealth
Graduate School of Biomedical Sciences
in Partial Fulfillment of the Requirements for the Degree of
DOCTOR OF PHILOSOPHY

by
Hsueh-Ping Chao, Ph.D. Candidate
Houston, Texas
December 2018

INTEGRATIVE BIOINFORMATICS APPROACHES TO ELUCIDATING PROSTATE CANCER CELL HETEROGENEITY, PLASTICITY, AND TREATMENT RESPONSE

by
Hsueh-Ping Chao, Ph.D. Candidate

APPROVED:

Dean G. Tang, M.D., Ph.D.
Advisory Professor

Sharon Y.R. Dent, Ph.D.

Vishwanath Iyer, Ph.D.

Yue Lu, Ph.D.

Jianhua Hu, Ph.D.

APPROVED:

Dean, The University of Texas
MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences

ACKNOWLEDGEMENTS

As my Ph.D. journey is winding down, it's time to express my sincere gratitude to everyone who helped me throughout this joyful and arduous process.

First, I am indebted to my mentor, Dr. Dean Tang. I thank him for the continuous support during my Ph.D. thesis research and for his patience, motivation, and immense knowledge. His guidance has helped me throughout the time of research as well as the process of writing this thesis. I could not have imagined having a better advisor and mentor for my Ph.D. study.

I offer my profound respect and appreciation for Dr. Sharon Dent, for providing me an opportunity to work with her team and giving me all the support and guidance, which made the completion of my Ph.D. study smooth and successful. I could not have made it without her generous support.

I whole-heartedly thank Dr. Yue Lu, for coaching me step by step, supporting me at every turn, and believing in me, and without whom it was impossible for me to switch the field and appreciate the beauty of bioinformatics.

I would like to thank the rest of my thesis committee, Dr. Vishwanath Iyer and Dr. Jianhua Hu, for their patience, insightful comments and advices, and moral support. My thanks also go to my candidacy committee members Dr. Taiping Chen, Dr. Shawn Bratton, and Dr. Rick Finch.

A special gratitude is due to Dr. Jianjun Shen, who has been selflessly guiding me to conduct another independent bioinformatics research project, to Dr. Collene Jeter, who has taught me many things from the beginning to end of my Ph.D. research, commented on my dissertation, and encouraged me all the

way through the process, and to Sara Gaddis, who has guided me to catch up with the basic knowledge of computer science and coding. With great professional appreciation and personal gratitude, I shall acknowledge all my colleagues in the Tang's lab and the Dent's lab, all the collaborators, and all my friends, especially Xin Liu, Xin Chen, Quihui Li, Jason Kirk, Evangelia Koutelou, and Xianghong Kuang.

I would like to dedicate this work to my family, my parents Chia-Chu Chao and Lina Shiang, my grandparents, and my sister, for their unconditional love and moral support. Also, I am so blessed with the most amazing extended family, my uncles, Johnny, Leonard, Jimmy, and Philip, my aunts, Selina, Sumei, Daina, and Sally, my cousins, Sean, Jennifer, Vivian, George, Judy, Ruth, Elaine, Ida, Lizzie, and Emily, my nephew, Joseph, and my niece, Scarlett. I can't image my life in US without you guys. Finally, I owe a special thanks to my best teammate, greatest support, and biggest comfort, Daniel. You are my hero!

Thank you all for being part of this adventure!

PUBLICATIONS

Publications contributed by Hsueh-Ping (Eva) Chao during Ph.D. research

1. Jeter C, Yang T, Wang J, **Chao HP**, and Tang DG. NANOG in cancer stem cells and tumor development: An update and outstanding questions. ***Stem Cells*** 33(8):2381-90, 2015.
2. Liu X, Chen X, Rycaj K, **Chao HP**, Deng Q, Jeter C, Liu C, Honorio S, Li H, Davis T, Suraneni M, Laffin B, Qin J, Li Q, Yang T, Whitney P, Shen J, Huang J, and Tang DG. Systematic dissection of phenotypic, functional, and tumorigenic heterogeneity of human prostate cancer cells. ***Oncotarget*** 6(27):23959-86, 2015.
3. Rycaj K, Cho EJ, Liu X, **Chao HP**, Liu B, Li Q, Devkota K.A, Zhang D, Chen X, Moore J, Dalby NK, and Tang DG. Longitudinal tracking of subpopulation dynamics and molecular changes during LNCaP cell castration and identification of inhibitors that could target the $PSA^{-/lo}$ castration-resistant cells. ***Oncotarget*** 7(12):14220-40, 2016.
4. Chen X, Li Q, Liu X, Liu C, Liu R, Rycaj K, Zhang D, Liu B, Jeter C, Calhoun-Davis T, Lin K, Lu Y, **Chao HP**, Shen J, and Tang DG. Defining a population of stem-like human prostate cancer cells that can generate and propagate castration-resistant prostate cancer. ***Clin Cancer Res***. *22* (17):4505-16, 2016.
5. Zhang D, Park D, Zhong Y, Lu Y, Rycaj K, Gong S, Chen X, Liu X, **Chao HP**, Whitney P, Calhoun-Davis T, Takata Y, Shen J, Iyer RV, and Tang DG. Stem cell and neurogenic gene-expression profiles link prostate basal cells to

aggressive prostate cancer. ***Nat Commun***. 7:10798, 2016.

6. Jeter C, Liu B, Lu Y, **Chao HP**, Zhang D, Liu X, Chen X, Li Q, Rycaj K, Calhoun-Davis T, Yan L, Hu Q, Wang J, Shen J, Liu S, and Tang DG. NANOG reprograms prostate cancer cells to castration resistance via dynamically repressing and engaging the AR/FOXA1 signaling axis. ***Cell Discov***. 2:16041, 2016.
7. Liu C, Liu R, Zhang D, Deng Q, Liu B, **Chao HP**, Rycaj K, Takata Y, Lin K, Lu Y, Zhong Y, Krolewski J, Shen J, and Tang DG. MicroRNA-141 suppresses prostate cancer stem cells and metastasis by targeting a cohort of pro-metastasis genes including CD44, Rho GTPases and EZH2. ***Nat Commun***. 8:14270, 2017.
8. Zhang D, Lin K, Lu Y, Rycaj K, Zhong Y, **Chao HP**, Calhoun-Davis T, Shen J, and Tang DG. Developing a Novel Two-Dimensional Culture System to Enrich Human Prostate Luminal Progenitors That Can Function as a Cell of Origin for Prostate Cancer. ***Stem Cells Transl Med***. 6(3):748-760, 2017.
9. Tomida J, Takata K.I, Bhetawal S, Person D.M, **Chao HP**, Tang DG, and Wood DR. FAM35A associates with REV7 and modulates DNA damage responses of normal and BRCA1-defective cells. ***EMBO J.*** 37(12), 2018.
10. Li Q, Deng Q, **Chao HP**, Liu X, Lu Y, Lin K, Liu B, Tang GW, Zhang D, Tracz A, Jeter C, Rycaj K, Calhoun-Davis T, Huang J, Rubin MA, Beltran H, Shen J, Chatta G, Puzanov I, Mohler J, Wang J, Zhao R, Kirk J, Chen X, and Tang DG. Linking prostate cancer cell AR heterogeneity to distinct castration and Enzalutamide responses. ***Nat Commun***. 9(1):3600, 2018.

11. Wible D, **Chao HP**, Chen MD, Tang DG, and Bratton B.S. ATG5 cancer mutations and alternative mRNA splicing reveal a conjugation switch that regulates ATG12-ATG5-ATG16L1 complex assembly and autophagy. *Under revision*.
12. **Chao HP**, Chen Y, Takata Y, Walker M, Lin K, Kirk J, Simper SM, Mikulec CD, Rundhaug E.J, Fischer M.S, Chen T, Tang DG, Lu Y, and Shen J. Assessment of the impact of RNA-seq protocols on transcriptome complexity and biological representation. *Submitted.*
13. Li Q, Liu B, **Chao HP**, Ji Y, Mehmood R, Jeter C, Lu Y, Chen T, Moore J, Li W, Liu C, Rycaj K, Tracz A, Kirk J, Calhoun-Davis T, Xiong J, Deng Q, Huang J, Coffey R, Chen X, and Tang DG. LRIG1 Is an Androgen Receptor-Regulated Pleiotropic Feedback Tumor Suppressor in Prostate Cancer. *Submitted*.
14. Koutelou E, Wang L, Schibler A, Chao HP, Kuang X, Lin K, Lu Y, Shen J, Jeter C, Salinger A, Wilson M, Chen YC, Atanassov B, Tang DG, Dent SYR. Usp22 controls multiple signaling pathways that are essential for vasculature formation in the mouse placenta. *Submitted*.
15. Liu X, Rycaj K, **Chao HP**, Wu M, Goodell M, Xi Y, Li W, Tang DG. An intrinsic neurogenesis program linked to distinct epigenetic profiles confers undifferentiated ($PSA^{-/lo}$) prostate cancer cells cancer stem cell properties. *Manuscript in preparation*.

INTEGRATIVE BIOINFORMATICS APPROACHES TO ELUCIDATING PROSTATE CANCER CELL HETEROGENEITY, PLASTICITY, AND TREATMENT RESPONSE

Hsueh-Ping Chao, M.S.

Advisory Professor: Dean G. Tang, Ph.D.

Background and Significance:

Prostate cancer (PCa) is the most common non-cutaneous tumor in American men, and the second leading cause of cancer-related deaths. PCa-related deaths can be attributed to heterogeneous tumors containing metastatic and therapy-resistant cancer cells. Cancer stem cells (CSC) are an important contributor to this tumor heterogeneity, which are present in primary tumors and become enriched in castration-resistant PCa (CRPC). Our lab has demonstrated that the prostate cancer stem cells (PCSCs) are enriched in the phenotypically undifferentiated PCa cell population that lacks the expression of differentiation marker prostate-specific antigen (PSA). Our work has also demonstrated that PCa cells manifest significant plasticity such that phenotypically differentiated PSA^+ PCa cells can be reprogrammed to the castration-resistant, stem-like state by chronic castration or overexpression of the stemness factor NANOG. Therefore, my overarching hypothesis is **that *$PSA^{-/lo}$ cells possess intrinsic molecular and epigenetic features that regulate their aggressiveness and stemness and contribute to tumor progression and therapy resistance, and that these properties can be gained through epigenetic reprogramming of more differentiated PSA^+ PCa cell population.*** Throughout my Ph.D. thesis research, I

employed an integrative bioinformatics approach to test this hypothesis with the following Specific Aims.

Specific Aim 1. Determine the intrinsic transcriptomic and epigenetic identities of $PSA^{-/lo}$ and PSA^{+} PCa cells using an integrative bioinformatics approach (Chpater 3)

Previous work demonstrated that the $PSA^{-/lo}$ cell population pre-exists in untreated patient tumors, becomes enriched in CRPCs, and is negatively associated patient overall survival. My hypothesis herein is that *$PSA^{-/lo}$ cells possess unique intrinsic transcriptomic and epigenetic features that causally regulate their stemness and aggressiveness.* To test this hypothesis, we purified $PSA^{-/lo}$ and PSA^{+} cell populations from untreated PCa cell culture models and performed deep RNA-Seq and H3K4me3 and H3K27me3 ChIP-Seq analyses. I have performed integrative bioinformatics analyses on these newly generated large-scale data, and the results reveal an interesting neurogenesis program in $PSA^{-/lo}$ PCa cells that is linked to their unique histone-association profiles.

Specific Aim 2. Use integrative bioinformatics approaches to elucidate how NANOG drives epigenetic reprogramming in PSA^{+} PCa cells (Chapter 4)

CSCs are well known to contribute to tumor heterogeneity. However, it is still poorly understood what factors drive the formation of CSCs in tumor development and progression. There are a number of ways that a more differentiated cancer cell may undergo reprogramming to become a CSC, including genetic deletions and therapy induced reprogramming. We have previously shown that the ES cell pluripotency factor NANOG is required for CSC properties and that inducible NANOG expression in bulk $AR^{+}PSA^{+}$ LNCaP cells can reprogram them to a castration-resistant stem-like state.

Here in this project, we followed up on this latter observation and performed both NANOG ChIP-Seq and RNA-Seq experiments in LNCaP cells expressing an inducible NANOG transgene. Our integrative bioinformatics-based analyses of the datasets reveals, surprisingly, that NANOG reprograms LNCaP cells not through reactivating the endogenous SOX2/OCT4/NANOG pluripotency network but instead via dynamically repressing and engaging the AR/FOXA1 signaling axis as well as MYC.

Specific Aim 3. Use integrative bioinformatics approaches to elucidate molecular mechanisms underlying treatment-induced CRPC reprogramming (Chapter 5)

Previously studies from our lab indicate that long-term castration of (AR^+)PSA^+ PCa cells *in vitro* and *in vivo* gradually 'reprograms' these cells and androgen-dependent tumors into androgen-independent cultures/tumors. In this project, we modeled CRPC development in the LNCaP and LAPC9 tumor systems and performed deep RNA-Seq experiments. Integrative bioinformatics analyses of the newly generated data revealed gene expression changes both common and unique to the two models. Importantly, my analysis helps pinpoint BCL-2 as a critical driver and therapeutic target in both models of CRPC.

TABLE OF CONTENTS

CHAPTER SIX

Conclusions and perspectives

LIST OF ILLUSTRATIONS

LIST OF TABLES

ABBREVIATIONS

ACD	Asymmetric Cell Division
AD	Androgen Dependent
ADT	Androgen Deprivation Therapy
AI	Androgen Independent
ALDH	Aldehyde dehydrogenase
ALL	Acute Lymphoblastic Leukemia
AML	Acute Myeloid Leukemia
AR	Androgen Receptor
ARE	AR Responsive Elements
CARN	Castration-Resistant Nkx3.1-expressing cells
ChIP-Seq	Chromatin Immunoprecipitation Sequencing
CNA	Copy Number Alteration
CRPC	Castration-Resistant Prostate Cancer
CSC	Cancer Stem Cell
DEG	Differentially Expressed Gene
DHT	Dihydrotestosterone
EMT	Epithelial-to-Mesenchymal Transition
ES	Enrichment Score
ESC	Embryonic Stem Cell
FC	Fold change
FDR	False Discovery Rate

GR	Glucocorticoid Receptor
GnRH	Gonadotropin-Releasing Hormone
GO	Gene Ontology
GSEA	Gene Set Enrichment Analysis
H3K4me1	Histone 3 Lysine 4 Monomethylation
H3K4me3	Histone 3 Lysine 4 Trimethylation
H3K9me2	Histone 3 Lysine 9 Dimethylation
H3K9me3	Histone 3 Lysine 9 Trimethylation
H3K27ac	Histone 3 Lysine 27 Acetylation
H3K27me3	Histone 3 Lysine 27 Trimethylation
H3K36me3	Histone 3 Lysine 36 Trimethylation
IHC	Immunohistochemistry
IGV	Integrative Genomics Viewer
IPA	Ingenuity Pathway Analysis
iPSC	Induced Pluripotent Stem Cells
KO	Knock Out
MSigDB	Molecular Signatures Database
NEPC	Neuroendocrine Prostate Cancer
NGS	Next Generation Sequencing
NOD/SCID	Nonobese Diabetic/Severe Combined Immunodeficiency mouse
NSG	NOD/SCID/IL2Rγnull (NSG) mouse
OSKM	OCT3/4, SOX2, KLF4, and c-MYC
PCA	Principal Component Analysis

PCa	Prostate Cancer
PCSC	Prostate Cancer Stem Cell
PRAD	Prostate Adenocarcinoma
PSA	Prostate Specific Antigen
RNA-Seq	RNA Sequencing
SC	Stem Cell
SHH	Sonic Hedgehog
shRNA	Short Hairpin RNA
siRNA	Small Interfering RNA
TF	Transcription Factor
TNBC	Triple-Negative Breast Cancer
TSS	Transcription Start Site
UCSC	University of California, Santa Cruz

CHAPTER ONE

Background and Introduction

The overarching goal of my Ph.D. thesis research is to employ various bioinformatics and related approaches and tools to help understand large datasets derived from experimental prostate cancer (PCa) models and clinical samples. Specifically, by analyzing these large –omics datasets, I aim to help my (1) cancer biology colleagues to generate testable hypotheses. In this starting chapter of my Ph.D. thesis, I shall first introduce a few general cancer biology concepts followed by brief discussions of PCa biology and treatment. I shall conclude this chapter by presenting my overall hypothesis.

1.1 Tumor cell heterogeneity and its clinical relevance

Cellular heterogeneity is an omnipresent phenomenon in solid tumor masses. Intratumoral heterogeneity describes the co-existence of distinct cancer cell subpopulations within a malignancy. The phenotypic and functional diversity of cell populations can be attributed to both genetic and non-genetic factors (2). Genetic influence on phenotypic diversity is driven by cancer cell intrinsic genomic instability that in combination with cell proliferation gives rise to clonal variants of the original neoplastic cell (3). On the other hand, non-genetic sources include distinct alterations in the epigenetic landscape and transcriptome, both of which may be partly attributable to the evolutionary gauntlet of the harsh tumor microenvironment. The epigenetic landscape, i.e., chromatin modifications, principally including DNA methylation, histone methylation and histone acetylation, often underlies cancer cell

phenotypic instability, permanently or transiently, by impacting expression of transcription factors (TF) and gene regulatory networks governing gene expression programs (4). Considering that the evolution of malignant neoplasms is fundamental to disease progression and the emergence of therapeutic resistance, a comprehensive understanding of the molecular circuitry underlying tumor cell heterogeneity and its relationship to the tumor microenvironment and epigenetic landscape is imperative (5).

Tumor cell heterogeneity provides the fuel for the development of clinically advanced cancer including metastatic and therapy-resistant diseases. For instance, the degree of tumor heterogeneity has been shown to negatively correlate with overall survival in ovarian cancer (6), lung cancer (7), and acute myeloid leukemia (AML) (8), as well as worse prognosis in breast cancer (9). The reciprocal correlation between tumor cell heterogeneity and cancer patient prognosis may implicate numerous dynamic populations driving tumor evolution and disease progression.

Highly tumorigenic and stem-like cancer cells termed cancer stem cells (CSCs) have been implicated as engines driving tumor heterogeneity and manifestation of aggressive cancer cell phenotypes including invasion, metastasis and therapy resistance (10). Phenotypically distinct from their more differentiated and non-tumorigenic counterparts, a higher frequency of CSCs - characterized by a broad spectrum of tissue and tumor type-specific markers - has been shown to correlate with poor clinical outcome in glioblastoma (11), breast cancer (12), pancreatic cancer (13), among other tumors. Tumorigenic $CD44^{+}CD24^{-/lo}$ breast CSCs implicated in invasion and metastasis (14, 15) were used to derive a 186-gene invasiveness signature to predict overall and metastasis-free survival of breast cancer patients as well as predict

poor clinical outcomes of patients afflicted with other types of cancer such as brain, lung, and prostate (16). Regardless of the presence of defined “cancer stem cells”, malignant cell plasticity, i.e., cancer cell ‘de-differentiation’ or phenotype switching, has been implicated in the evolution of patient tumor cells, such as the emergence of chemotherapy-resistant cells in triple-negative breast cancer or TNBC (17), and, in melanoma (18).

In light of the emerging importance of tumor cell heterogeneity and plasticity in the reprogramming of cancer cells to lethal therapy-resistant and metastatic disease, crucial outstanding questions include: what are the distinct molecular circuitries underlying such malignant cancer cells and how can we use that knowledge to target those cells in the clinic?

1.2 Generation of tumor cell heterogeneity: Clonal evolution and CSCs

Clonal evolution is a process of clonal expansion, diversification and selection, following the rules of Darwinian evolution (19) (Figure 1-1). The CSC model proposes a hierarchal organization of cancer cells in a tumor clone, with CSCs undergoing different levels of differentiation (maturation) thus generating intra-clonal cellular heterogeneity (Figure 1-1). CSCs are defined as unique cancer cells possessing certain biological properties of normal stem cells (SC) such as the capacities of self-renewal and phenotypic/functional differentiation, manifested as the abilities for tumor initiation, long-term clonal propagation, and anchorage independence (20). Among a wide variety of cell surface markers, CD24, CD44, CD133, CD166 and ALDH have been routinely applied to select for, assess and characterize CSCs from multiple solid

tumors including breast, prostate, colon, brain and many others (21, 22). The “gold standard” assay to demonstrate definitive CSC traits of prospectively purified candidate CSC populations is tumor regeneration upon transplantation. For example, primary human colorectal cancer specimens contain relatively rare $CD133^+$ cells that, when xenografted in limiting numbers into immunocompromised recipient mice, generated tumors comprised of both rare $CD133^+$ and a majority of $CD133^-$ cells, whereas $CD133^-$ cells lacked similar tumor-regenerating capabilities (23). Similar recapitulation of heterogeneous tumor cell subpopulations has been observed in $CD133^+$ ovarian CSC-derived xenograft tumors (23, 24). Conversely, expression of the prostate differentiation marker prostate specific antigen (PSA) can distinguish PCa cells with reduced tumorigenic potential relative to more primitive $PSA^{-/lo}$ cells and, importantly, xenografted $PSA^{-/lo}$ PCa cells can give rise to heterogeneous tumors harboring both $PSA^{-/lo}$ and PSA^+ cells (25). These general phenomena have also been observed upon transplantation of stem versus non-stem cells derived from other cancers, including those in the brain (26), breast (27), colon (23, 28, 29), ovary (30), and pancreas (31, 32). However, CSCs are not a static entity - rather, the CSC population may yet be heterogeneous containing subsets of tumorigenic cells and CSC themselves, just like normal SCs, may possess significant plasticity (33) (Figure 1-1). Furthermore, not all CSC subsets may possess sustaining tumorigenic potential.

1.2.1 Cell-of-origin of cancer

The cancer cell-of-origin is defined as the normal cell in which the first cancer driver mutation occurs and that is responsible for tumor initiation but may not

necessarily be identical to CSCs (Figure 1-1). Whereas the cell-of-origin is responsible for the initiation of tumorigenesis, the CSC is responsible for driving tumor growth and long-term tumor maintenance and disease progression. The cancer cell-of-origin may arise from mutations in tissue SCs; alternatively, the cancer cell-of-origin may be more mature progenitor cells, or even differentiated cells that have undergone de-differentiation and re-acquired self-renewal properties (34). Cancer cell longevity supports the accumulation of additional oncogenic mutations that gradually confer the hallmarks of cancer (35), thereby facilitating the expansion of tumorigenic populations.

1.2.2 Self-renewal

Self-renewal refers to a mode of cell division that generates at least one daughter cell with long-term repopulating potential and multipotency (36). Self-renewing divisions may be asymmetric giving rise to one SC and one lineage-restricted progeny, or symmetric yielding two equipotent SCs or two differentiated daughter cells. During normal organogenesis, asymmetric cell division (ACD) is fundamental for generating the myriad of functional cells in a tissue or organ. Tumors, like abnormally developed organs, harbor a spectrum of phenotypically and histologically distinct cancer cells ranging from renewing stem-like cells to more mature and differentiated cells (Figure 1-2A) (20). Although aberrantly differentiated or present in skewed ratios relative to normal tissues, pre-existent stem-like cancer cells displaying high self-renewal and differentiation potential (often corresponding with a lack of terminal differentiation markers) have been found to be tumorigenic upon transplantation. For example, limiting numbers of self-renewing $CD133^+$ human colon

cancer-initiating cells recapitulated the parental tumor phenotype upon transplantation, forming heterogeneous tumors composed of relatively rare $CD133^+$ and a prevalence of $CD133^-$ tumor cells; $CD133^+$ cells could also uniquely reform secondary and tertiary tumors upon serial transplantation (23). Similarly, a single self-renewing $CD34^-p75^-$ murine melanoma cell reestablished a heterogeneous tumor mass upon transplantation whereas $CD34^+p75^-$ cells were tumorigenic but not capable of reestablishing tumor heterogeneity (37). Thus, self-renewal is a cardinal property of CSCs to generate cellular heterogeneity observed in tumors.

CSCs may acquire self-renewal property by "hijacking" normal SC self-renewal signaling mechanisms, such as WNT/β-catenin, Sonic Hedgehog (SHH) and Notch (38). For example, the WNT/β-catenin signaling has been implicated in the maintenance of normal and cancer stem cell self-renewal. In mouse embryonic SCs (ESCs), paracrine and autocrine WNT signaling promotes self-renewal preventing differentiation to (less primitive) epiblast stem cells (39). Furthermore, inhibition of the β-catenin negative regulator glycogen synthase kinase-3 (GSK3) stabilizes β-catenin (40), which, in turn, interacts with and inhibits Tcf3 repression of the ESC pluripotency network Oct3/4, Sox2, and Nanog (41); alternatively, β-catenin may independently interact with Oct4 to regulate the SC properties in ESCs (42). In the cancer context, aberrant Wnt signaling has been implicated in a wide variety of malignancies, including breast, colon, skin, and brain cancers, as well as leukemia (43-45). Stem-like colorectal cancer cells display higher Wnt activity, and hepatocyte growth factor (HGF) has been shown to activate β-catenin activity and promote the CSC phenotype (46).

WNT also sustains breast CSCs, induces infiltration, and augments metastatic colonization (47).

TFs regulating normal SC self-renewal have also been implicated in cancer. For example, ectopic expression of the pluripotency and reprogramming factor Oct4 in a doxycycline-inducible transgenic mouse model induces rapid hyperplasia and dysplasia in the skin and colon by expansion of progenitor cells and blocking their differentiation (48). These studies illustrate that CSCs may deregulate normal SC self-renewal signaling networks to drive malignant transformation and maintain tumor growth (43).

1.2.3 Plasticity and evolution

Cellular plasticity is the ability of cells to reversibly change the directions of their development, by dedifferentiation, transdifferentiation, and reprogramming to adapt to a unique microenvironment (Figure 1-2B). Cell plasticity is tightly constrained and rarely occurs in fully differentiated somatic cells under normal homeostatic conditions but may be important for regeneration of damaged tissues and organs (Figure 1-2A). Dedifferentiation is a restoration process of reverting terminally specialized cells back to a more primitive state, whereas transdifferentiation is a process of breaking the lineage barrier and reverting terminally specialized cells into an alternative lineage. In normal cells, transdifferentiation can be a two-step process: cell dedifferentiation followed by reactivation of another differentiation program. Experimentally, both dedifferentiation and transdifferentiation can be reprogrammed via epigenetic mechanisms. *In vitro*, cell reprogramming can be engineered via exogenous growth

factors and directed expression of TFs specifying cell fate. For example, overexpression of a cocktail of TFs, OCT3/4, SOX2, KLF4, and c-MYC (OSKM), is sufficient to reprogram mouse fibroblasts to a pluripotent state generating so-called induced pluripotent stem (iPS) cells (Figure 1-3) (49). Reprogramming of differentiated cells to iPS cells requires transcriptome and chromatin changes to erase the lineage-specific program, followed by increased expression of genes crucial to reestablish and maintain pluripotency, then by changes in the epigenetic landscape to lock the cells in the reprogrammed state (Figure 1-3) (50). This breakthrough study (51) and many following studies reveal that critical TFs orchestrate dynamic reorganization of the epigenome and intimately interplay with the transcriptome during cell reprogramming (52).

Cancer cells in general, due to their unstable genome and more 'fluid' epigenome, display higher plasticity than their normal, differentiated counterparts. CSCs also possess significant plasticity to adapt to changes during disease progression. For example, the $CD133^{+}CD144^{+}$ CSC subpopulation derived from patient glioblastoma is endowed with characteristics similar to endothelial progenitors with expression of vascular endothelial-cadherin (53, 54). Under drug perturbation, CSCs and non-CSCs are inter-convertible, such as stem-like $ABCG2^{+}$ and more differentiated $ABCG2^{-}$ cells in breast cancer and PCa, and stem-like $CD44^{+}$ and more differentiated $CD44^{-}$ cells in PCa (55, 56). Besides treatment, other elements in the tumor microenvironment, such as inflammatory molecules (e.g., TNF-α and reactive oxygen species) and hypoxia, can also promote cancer cell dedifferentiation together with increased expression of stemness genes and CSC properties (57). The

acquisition of epithelial-mesenchymal transition (EMT), a phenotype often correlated with CSCs, facilitates cancer progression, especially invasion and metastasis (58). Furthermore, EMT-related TFs such as Zeb1 have been implicated as a driving force of cellular plasticity and metabolic changes by globally switching the cancer cell gene expression profiles and promoting metastasis in pancreatic cancer (59). Thus, specifically targeting emerging and evolving CSCs is essential to successfully treat patients in the clinic.

In Chapters 4 and 5, I shall present examples of PCa cell plasticity induced by a 'stemness' gene NANOG and by therapeutic treatment, castration, respectively.

1.2.4 Therapeutic targeting of CSCs

Clinical studies indicate that the percentage of $CD44^{+}CD24^{-}$ breast CSCs in patient tumors after neoadjuvant chemotherapy is increased (60) and CSC-like molecular features have been detected in the transcriptome of breast tumors following endocrine therapy or chemotherapy as compared to treatment-naïve tumors (61). Additionally, a drug screen targeting mesenchymally transdifferentiated CSCs revealed a novel compound (salinomycin) with a select toxicity for CSCs that could potentially be combined with standard chemotherapy (paclitaxel) to benefit patient prognosis in breast cancer (62). Targeting CSCs may also be applied to treat PCa, particularly castration-resistant PCa (CRPC). For example, Wnt/β-catenin signaling regulates expression of the stem cell marker CD44 and pharmacological inhibition of Wnt in combination with the conventional chemodrug docetaxel synergistically inhibits CRPC xenograft growth, presumably by targeting both CSCs and non-CSCs,

respectively (63). Inhibition of the SHH self-renewal pathway by siRNAs directly against pathway components has also been shown to reduce tumorigenic potential and metastasis in multiple cancer types in xenograft models (64, 65). Such evidence offers a new therapeutic strategy that combination of agents aimed at both CSC and non-CSC may produce synergistic anticancer efficacy. Encouragingly, various experimental therapeutics and therapeutic strategies targeting different aspects of CSC properties to selectively eradicate CSCs have entered or are being translated into clinical trials (66).

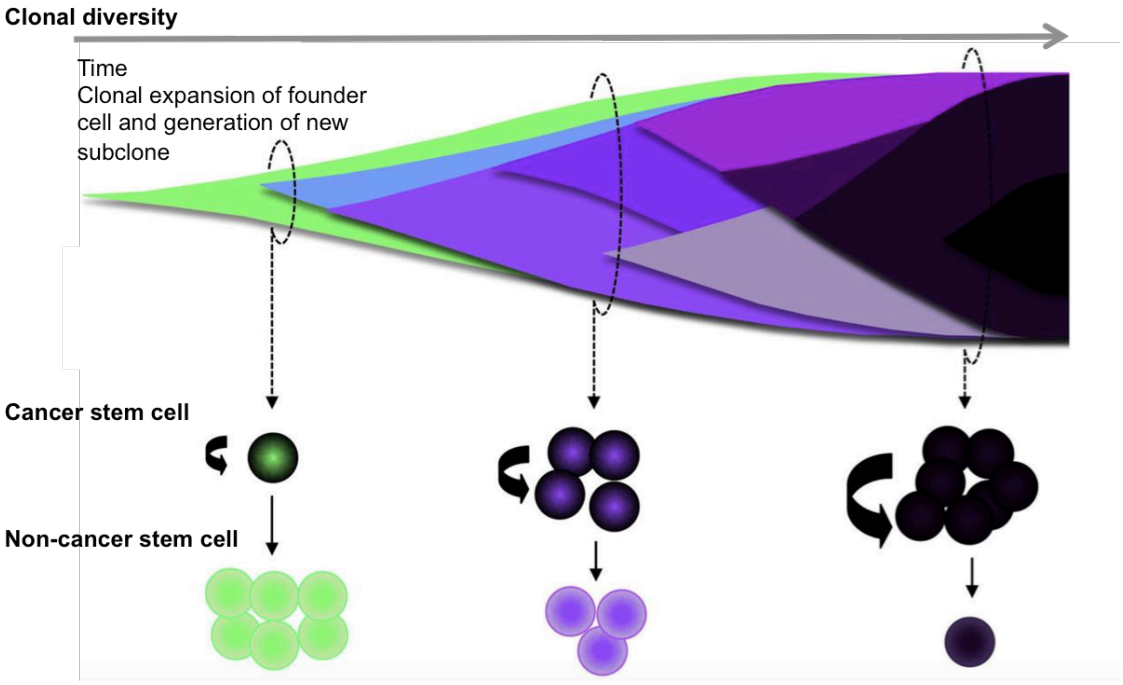


Figure 1-1. Clonal evolution and cancer stem cells (adapted from (67) Kreso, A., and J. E. Dick. 2014. Evolution of the cancer stem cell model. *Cell stem cell* 14: 275-291).

In clonal evolution, the founder tumor cell can give rise to subclones that have gained novel mutations (or epigenetic changes) imparting an oncogenic evolutionary advantage to particular cell subsets making up the heterogeneous tumor mass. Dynamic CSC subpopulations may emerge during tumor evolution accounting for subclonal cellular heterogeneity over time.

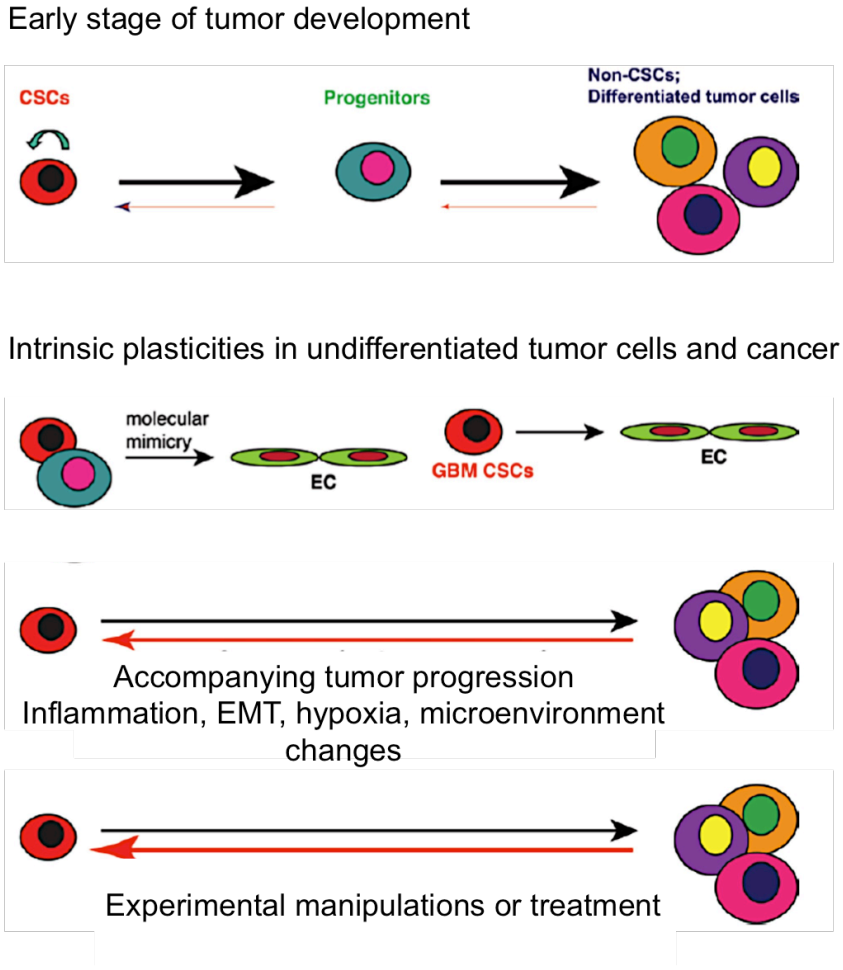


Figure 1-2. Cell plasticity in cancer (adapted from (20) Tang, D. G. 2012. Understanding cancer stem cell heterogeneity and plasticity. *Cell research* 22: 457-472).

In untreated tumors, CSCs may self-renew and generate lineage-restricted progenitor cells that then give rise to (phenotypically) differentiated tumor cells. Only very rarely can the differentiated tumor cell 'spontaneously' dedifferentiate to regain CSC properties (back arrow). On the other hand, CSCs may manifest intrinsic plasticity, such as glioblastoma multiforme (GBM) CSCs transdifferentiating into endothelial cells (EC) to promote tumor development. During tumor progression, environmental stresses (e.g., inflammation, EMT, hypoxia etc) may facilitate non-CSC dedifferentiation to CSCs. Experimental perturbations and therapeutic treatments also contribute to dedifferentiation of non-CSCs.

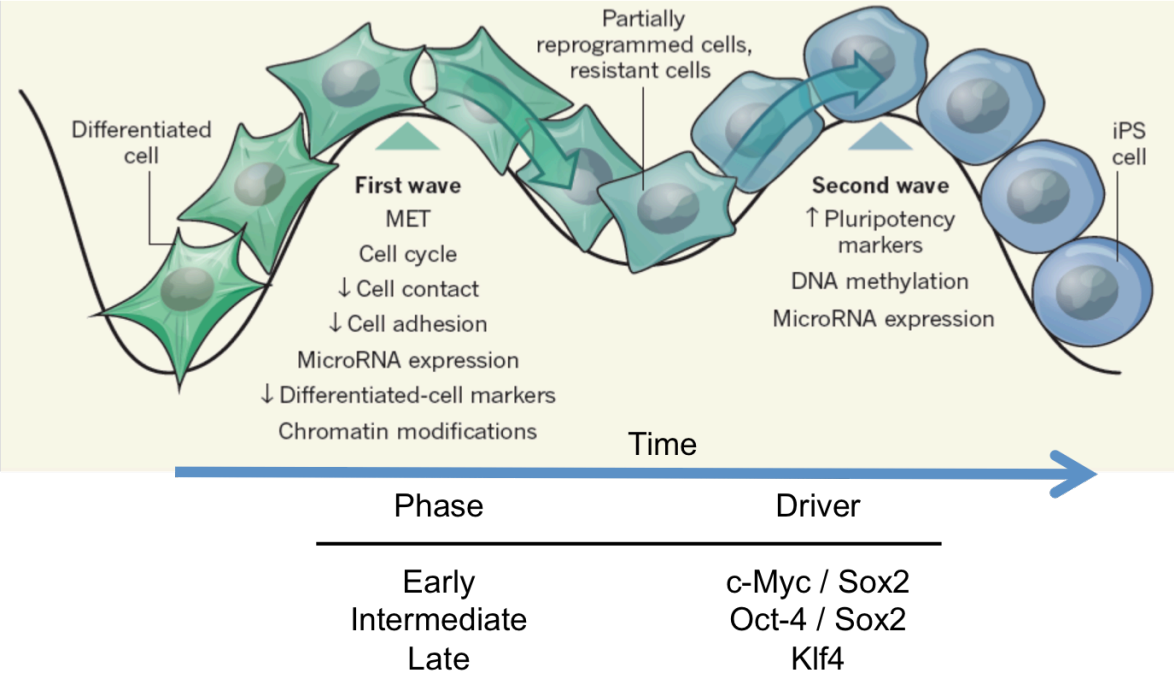


Figure 1-3. Fibroblast reprogramming via iPS cell factors (adapted from (50) Sancho-Martinez, I., and J. C. Izpisua Belmonte. 2013. Stem cells: Surf the waves of reprogramming. *Nature* 493: 310-311).

Adult fibroblasts and mouse embryonic fibroblasts subject to forced dedifferentiation and acquisition of ESC traits by engineered expression of the OSKM (Oct4, Sox2, Klf4 and c-Myc) TFs. In the early phase of reprogramming, differentiated cells undergo transcriptome and chromatin landscape changes, which may be driven by c-Myc to trigger mesenchymal epithelial transition (MET) and erase cell identity, such as differentiation markers. The intermediate phase, regulated by Oct4 and Sox2, leads to partially reprogrammed cells giving rise to a heterogeneous dynamic cell population. In the late phase of reprogramming, Klf4 promotes partially reprogrammed cells to gradually express endogenous genes involved in re-establishing pluripotency and to alter DNA methylation to 'seal' the cell in reprogrammed state.

1.3 Epigenetics

A recent study investigated the chemoresistance evolution in TNBC treated with neoadjuvant chemotherapy by integrative single-cell DNA- and RNA- sequencing (17). The results revealed that chemoresistance-associated mutations existed in pre-treatment tumors and mutation frequencies of post-treatment tumors were either unchanged or decreased. However, transcriptome profiles were reprogrammed by acquiring new chemoresistant transcriptional changes and converged on a few common pathways related to extracellular matrix degradation, CDH1 targets, hypoxia, EMT, and angiogenesis (17). These observations provide direct evidence that mechanisms other than genetic alterations critically underlie adaptive phenotype emergence.

Canonical non-genetic factors regulating developmental processes are epigenetic regulatory mechanisms, which include DNA methylation, crosstalk of histone posttranslational modifications, incorporation of histone variants, chromatin remodelers managing chromatin structure, and non-coding RNAs that modulate and fine-tune gene expression (reviewed in (68)). These machineries play versatile roles in cell fate decision-making processes (Figure 1-4).

1.3.1 Chromatin landscape affects transcriptome and cell fate

Chromatin is relatively more accessible in ES cells than in terminally differentiated cells and is subject to dynamic changes during cell differentiation (68). In ES cells, pluripotency genes are transcriptionally active with enrichment of active marks on enhancer (e.g., H3K4me1, H3K27ac, and histone acetyltransferase p300),

promoter (H3K4me3), and gene body (H3K36me3) regions, as well as less inhibitory marks on heterochromatin (e.g., H3K9me2, H3K9me3, and HP1) and lower levels of methylated DNA (5'-methyl cytosine). On the other hand, lineage-restricted genes are considered "poised" for expression manifested by the presence of bivalent (both transcriptionally activating and repressing) marks on enhancer (H3K4me1, p300, and H3K27ac) and promoter (H3K4me3 and H3K27me3) regions and low level of H3K27me3 on intergenic regions (68). In differentiated cells, pluripotency genes are turned off by removing active histone marks on enhancers, promoters, and gene bodies as well as the addition of suppressive marks (H3K9me2 and HP1) and heavy DNA methylation on promoters and heterochromatin. Lineage genes are correspondingly activated by H3K4me1, H3K27ac, and p300 on enhancers, H3K4me3 on promoters, and H3K36me3 on gene bodies (68).

Histone modifications are more than simply markers of the gene transcription state - they not only adjust chromatin accessibility and recruitment or removal of effector proteins regulating transcription, DNA replication, and/or DNA repair (69) but also delineate a unique cellular identity (68). For example, ES cells possess bivalent chromatin domains, which contain both H3K4me3 and H3K27me3 marks, regulated by trithorax group and polycomb group, respectively (70-72). The development genes enriched with bivalent marks on promoters are poised for exquisite and rapid response of gene expression. In contrast, those genes that lose bivalency preserve the cell in differentiated fate (73). The other way around, chromatin remodelers affecting histone modifications may help in the establishment of cell identity. For example, the ATP-dependent chromatin remodeler, Tip60-p400, co-localizes with H3K4me3 around the

gene transcription start site (TSS), and, RNAi-mediated knockdown of Tip60-p400 in mouse ES cells causes upregulation of genes responsible for early differentiation (74). Interestingly, this study also suggests that knocking down the pluripotency TF Nanog reduces p400 binding to target genes thus impacting gene expression in ES cells (74). These studies, together, demonstrate that epigenetic regulatory mechanisms that alter the chromatin landscape and the TFs cross talk with chromatin remodelers to coordinately regulate cell fate decisions.

1.3.2 TFs and cell state

TFs such as Oct4, Sox2, and Nanog are crucial in maintaining pluripotency and self-renewal (75, 76). Introducing the OSKM factors can reprogram fibroblasts into iPS cells (Figure 1-3). Oct4 is critical to maintain ESCs in a pluripotent state and keep their stem-cell identity (48). Functional studies using cultured ESCs demonstrate that overexpressing Nanog promotes ESC self-renewal; conversely, targeted disruption of Nanog in mouse embryos causes endoderm-like cell differentiation and Nanog-null embryos fail to form a pluripotent inner cell mass. (77). In more lineage-restricted cells later in development, and, in homeostasis in the adult, a panoply of tissue-specific master TFs regulate gene expression programs via occupancy of super-enhancers dictating and maintaining cell specification (78).

Among the TFs involved in either developmental programming (cell differentiation) or induced reprogramming, the so-called ‘pioneering’ TFs can uniquely access closed chromatin and further recruit other TFs, cofactors, chromatin-modifiers or modelers to activate target gene expression. During iPS cell reprogramming, Sox2

occupies chromatin first followed by Oct4, which, together, assemble an enhanceosome to launch the reprogramming process to a pluripotent state (79). Pioneering TFs are also indispensable for development, such as Forkhead box protein A1 (FoxA1), a member of the winged-helix FoxA family crucial for organ development and a key mediator of postnatal development and homeostasis of epithelial tissues (80). For example, FoxA1 can bind heterochromatin, and recruit chromatin remodelers and other TFs to remodel the chromatin landscape and transcriptionally activate target genes, such as Muc2 (involved in the differentiation of goblet cells in the gut epithelium (81)) and alb1 (involved in liver development (82)). Additionally, FoxA1 is known to alter the activity of nuclear hormone receptors estrogen receptor (ER) and androgen receptor (AR), thus regulating breast and prostate development, respectively (80, 83). Of interest, FoxA1 has also been implicated in breast and prostate cancer development (80, 83).

1.3.3 Oncogenic reprogramming

In cancer development, oncogenic transformation mirroring cellular reprogramming suggests that pluripotency and self-renewal TFs may play critical roles in the emergence of various malevolent properties, including tumorigenicity, clonogenic growth, invasion, metastasis, and therapeutic resistance. Master TF-mediated reprogramming in cancer cells may thus drive the formation and plasticity of CSCs (84). For instance, inducing NANOG expression in a wide variety of cancer cell types (e.g., breast, prostate, liver and ovary) promotes cell reprogramming with expression of stem cell markers, (e.g., CD44 and CD133) and increased tumor-

initiating capacity (85-93). In our lab, loss- and gain-of-function studies have demonstrated that NANOG loss inhibits sphere formation, clonal growth, and tumor development in several tumor systems (87) whereas NANOG overexpression enhances CSC phenotypes and functional reprogramming including CRPC development (93). Suppressing NANOG specifically in $PSA^{-/lo}$ prostate CSCs inhibits xenograft tumor regeneration in castrated hosts (25). In addition to NANOG, other pluripotency molecules including SOX2 and OCT4 have also been implicated in regulating CSC properties and tumor development (94).

Deregulated epigenetic mechanisms have also been implicated in tumorigenesis. Hypermethylation of CpG islands and global DNA hypomethylation are commonly observed in the cancer epigenome (95). DNA hypomethylation at repeat sequence or retrotransponsons result in gene instability by increasing chromosomal rearrangement and gene translocation events and DNA hypomethylation activate oncogenes, such as RAS and IGF2 (due to loss of imprinting), potentially contributing to cancer development and disease progression (96). Finally, epigenetic switching is altered in cancer cells such that developmental genes silenced by H3K27me3 in ES cells and only activated during cell development may be switched from a flexible repressive polycomb mark to stable DNA methylation that permanently silences these genes. Therefore, precisely regulating epigenetic machinery is critical for normal cell differentiation and organ development, and aberrant regulation of these pivotal control mechanisms may drive oncogenic reprogramming including cancer initiation and malignant progression.

In Chapter 4, I shall discuss LNCaP cell reprogramming by inducible expression of NANOG, a process that resembles oncogenic reprogramming.

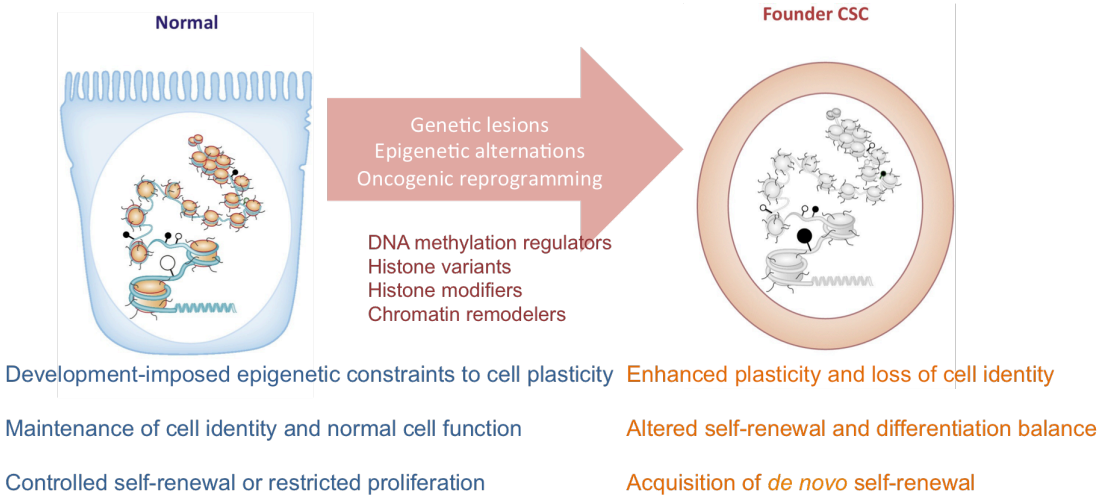


Figure 1-4. CSCs induced by genetic lesions, epigenetic alterations, and oncogenic reprogramming (adapted from (97) Wainwright, E. N., and P. Scaffidi. 2017. Epigenetics and Cancer Stem Cells: Unleashing, Hijacking, and Restricting Cellular Plasticity. *Trends in cancer* 3: 372-386).

Normal SCs or committed cells can be transformed to CSCs by genetic lesions, epigenetic alterations, and oncogenic reprogramming, which disrupt the normally tight control of cell identity, plasticity, and self-renewal. The 'reprogrammed' CSCs may acquire self-renewal properties and manifest increased plasticity.

1.4 Prostate and PCa

The prostate is a muscular, androgen-regulated and androgen-dependent secretory gland of the male reproductive system involved in producing seminal fluids. The prostate is composed of three distinct cell types, luminal cells, basal cells, and neuroendocrine (NE) cells. Luminal and basal cells are structured into two stratified epithelial layers, with the luminal layer facing the lumen and the basal layer adjacent to the basement membrane with interspersed NE cells. Luminal cells, the differentiated, androgen-dependent secretory cells of the gland, express abundant AR, PSA, and the cytokeratins CK8 and CK18. Basal cells, on the other hand, express p63, CK5, and CK14, and low levels of AR. NE cells are rare and express chromogranin A, synaptophysin, and serotonin, but do not express AR.

1.4.1 Prostate development and prostate stem cells (PSCs)

The prostate is of endodermal origin, arising from the urogenital sinus epithelium upon stimulation by AR-expressing, ectodermally derived urogenital sinus mesenchyme (98). In humans, branching morphogenesis primarily occurs embryonically whereas in mouse the process mainly occurs postnatally, with prostate growth and maturation occurring upon pubescence stimulated by rising androgen levels (98). Prostate morphogenesis and homeostasis are initiated and maintained by steroid hormones that regulate prostate development especially androgens. The androgen receptor, AR, is a steroid hormone dependent nuclear TF and plays a pivotal role in prostate development (99). AR binding to 5α-dihydrotestosterone (DHT),

a metabolite of testosterone, causes a conformational change in the receptor, leading to the formation of an AR dimer that then translocates into the nucleus. Dimeric AR binds to androgen response elements (AREs) on AR-target genes, triggering downstream gene expression such as PSA and transmembrane protease serine 2 (TMPRSS2) driving normal male sexual development and differentiation. Autocrine and paracrine growth and differentiation factors under mesenchymal AR control induce epithelial cell proliferation and differentiation, whereas epithelial AR regulates mesenchymal cell differentiation and basal cell proliferation (100).

In addition to AR, several other TFs also help coordinate prostate organogenesis. For instance, FOXA1 is expressed in the developing prostate epithelium and regulates cell maturation and ductal morphogenesis. FOXA1 physically interacts with AR and a FOXA1 responsive element is critical for PSA expression, which indicates functional maturation of luminal epithelial cells (101). Of note, the roles of FoxA1 in epithelial differentiation and maturation can be both AR-dependent and independent (100). Similarly, the homeobox domain TF Nkx3.1 is expressed in budding tips of the urogenital sinus epithelium during fate determination and maintained continuously during branching morphogenesis and throughout life, indicating that this TF is crucial for both prostate morphogenesis and homeostasis. Other signaling pathways and molecules such as Notch, SHH, FGF-10, TGFβ, and WNT also spatially and temporally regulate prostate gland development.

Unlike the mammary gland that undergoes cyclic changes in sync with menstrual cycles and in response to pregnancy, the adult prostate is a relatively dormant organ with slow cell turnovers. However, experimentally castrating the rodent

prostate will induce dramatic organ atrophy and re-supplementation of androgens restores organ growth to the original size (102), suggesting the presence of PSCs that can survive androgen deprivation and regenerate the entire organ. Tissue recombination studies using fractionated cells suggested that a subset of rare basal cells ($CK5^+CK14^+CK8^+CK18^+CK19^+GSTpi^+p63^+$) might be tripotent and able to generate basal, luminal, and NE cells (98), Similarly, human PSCs have been prospectively isolated by virtue of their high expression of $\alpha2\beta1^+$ integrin and $CD133^+$, with FACS purified $\alpha2\beta1^+CD133^+$ basal cells displaying higher capacities forming colonies *in vitro* and prostatic glandular structures *in vivo* (103). More recently, lineage tracing studies in mice have definitely identified a subset of basal cells as the primitive PSCs being able to generate terminally differentiated basal cells, luminal cells and NE cells (104). In the adult prostate, however, both basal and luminal cell compartments seem to be largely self-sustained by lineage-restricted progenitor cells (105).

1.4.2 PCa cell of origin and prostate CSCs (PCSCs)

The majority of human PCa present as adenocarcinomas with an AR^+PSA^+ luminal phenotype, making most oncologists believe that the cell of origin of human PCa is the luminal cell. However, tissue recombination (*ex vivo*) assays using basal vs. luminal (stem/progenitor) cells that express activated oncogenes suggested basal cells as the preferred cell of origin for PCa. For example, transplantation experiments, using tissue recombinants containing AKT-transduced murine prostate epithelial cells combined with rat embryonic UGS mesenchyme and transplanted into immunocompromised mice, demonstrate that activation of AKT in $Sca1^{hi}$ basal cells

but not luminal cells was sufficient for PCa initiation (106, 107). Similarly, human prostate-derived primary basal cells transduced to express activated AKT, ERG, and AR gave rise to adenocarcinomas with histological features reminiscent of human tumors whereas similarly transduced luminal cells were ineffective in forming tumors (107).

On the other hand, in contrast to the results obtained from these transplantation-based experiments using manipulated human prostate cells, lineage-tracing studies in genetic mouse models have arrived at opposite conclusions, i.e., luminal cells (or cells in the luminal cell compartment) actually represent the preferred cell of origin for PCa. For instance, conditional deletion of *Pten* in luminal cells induced luminal-type of adenocarcinomas (108-110). Indeed, *Pten* deletion in $K8^+$ luminal cells generated adenocarcinomas with shorter latency and higher incidence than *Pten* deletion in $K5^+$ basal cells (111). Thus, PCa derived from transformed luminal cells appears to be more aggressive (than basal cell-derived tumors) and harbors a molecular signature associated with worse prognosis in PCa patients (111).

Recent RNA-Seq studies using purified human prostate basal and luminal epithelial cells (112, 113), 2D luminal progenitor cell cultures (114), and 3D organoid cultures (115) have reconciled the above-discussed discrepancies observed in transplantation vs. lineage tracing studies with respect to the PCa cell-of-origin. The consensus from these new studies is that both basal and luminal cells can be tumorigenically transformed and, depending on the transforming events, the two cell types may give rise to different histological subtypes of PCa. Specifically, transformed luminal cells or luminal progenitor cells generate mostly adenocarcinomas whereas

transformed basal cells preferentially generate basal cell and squamous type of carcinomas.

Regardless of the potential cell of origin of PCa, work from our lab and many others' has provided strong evidence for the presence and clinical relevance of PCSCs, which have been prospectively identified, purified and characterized from a variety of human PCa cell lines, xenografts and patient tumor samples. Our early study revealed that the multi-drug resistant side population (SP) from the LAPC9 xenograft model manifested 500-fold higher tumorigenicity than the non-SP or bulk LACP9 cells (55). Subsequently, we demonstrated that PCa cells expressing high levels of cell surface CD44 are enriched in clonogenic, tumorigenic and metastatic stem/progenitor cells (56). In a separate study, we showed that PCa cell holoclones could regenerate all three types of clones in culture (i.e., holoclones, meroclones, and paraclones), and, importantly, harbored self-renewing PCSCs that could long-term propagate xenograft tumors (116). Combinatorial markers have also been employed to enrich PCa stem/progenitor cells. For example, $CD44^{+}\alpha2\beta1^{hi}CD133^{+}$ cells from human primary PCa have been reported to possess high proliferative, self-renewal, and differentiation capacities (117). Similarly, in cell line and xenograft models, $CD44^{+}CD133^{+}$ PC3 and DU145 cells have been shown to be more tumorigenic than marker-negative cells (118). Likewise, $CD44^{+}\alpha2\beta1^{hi}CD133^{+}$ DU145 cells (119) and $CD44^{+}CD24^{+}$ LNCaP (120) cells express higher levels of certain stemness genes (e.g., Oct4, BMI1, β-catenin) and display higher clonogenic and tumorigenic potential than $CD44^{+}\alpha2\beta1^{lo}CD133^{-}$ DU145 and $CD44^{+}CD24^{-}$ LNCaP cells, respectively. Interestingly, docetaxel-resistant DU145 and 22RV1 cells have been shown to lack

differentiation markers (e.g., CK19, CK18, HLA-I antigen, PSA and AR), display high tumor-initiating capacity, and overexpress some stemness genes such as NOTCH2 (121).

Relying on PSA as one of the most prostate-specific differentiation markers and by developing a lentiviral-based tracking system using the *PSA* promoter (*PSAP*) to drive the expression of GFP or dsRed, our lab has demonstrated PCSCs in the $PSA^{-/lo}$ PCa cell population, which preferentially express some stem cell-associated genes, are relatively quiescent (slow cycling), can undergo ACD generating differentiated $PSA^{+/hi}$ daughter cells, and possess long-term tumor-propagating capabilities in castrated hosts (25). Of clinical relevance, the $PSA^{-/lo}$ PCa cells are intrinsically resistant to androgen deprivation therapy (ADT) and become enriched upon castration (25). Interestingly, castration of bulk $AR^{+}PSA^{+}$ PCa cells (e.g., LNCaP) *in vitro* leads to the generation of homogenous population of $PSA^{-/lo}$ cells that become completely refractory to antiandrogens such as enzalutamide (122). It should be noted that true PCSCs in the $PSA^{-/lo}$ PCa cell population, assessed by their abilities to undergo ACD, are generally rare, ranging from 3-18% depending on models (25). Moreover, the $PSA^{-/lo}$ PCa cell pool is yet heterogeneous containing subsets of tumorigenic cells (123) (Figure 1-5). For instance, we have recently demonstrated that a subset of $PSA^{-/lo}$ cells marked by the $ALDH^{hi}CD44^{+}\alpha2\beta1^{+}$ phenotype [also called triple $marker^{+}$ (TM^{+})] are highly castration-resistant and can initiate and long-term propagate CRPC upon serial transplantations in castrated NSG mice (124). Altogether, these studies (25, 122-124) establish the $PSA^{-/lo}$ PCa cells as a clinically relevant cell population highly enriched in PCSCs that fulfill the most stringent defining criteria of CSCs. These

studies also establish the PSA$^{-/lo}$ PCa cell population as a therapeutic target in treating PCa and CRPC.

In Chapter 3, we further define the transcriptomic features and several histone mark profiles in PSA$^{-/lo}$ PCa cells.

1.4.3 PCa treatment

AR is not only important for normal prostate development. It is frequently overexpressed and/or mutated in treatment-failed PCa and the vast majority of early stage prostate tumors are androgen-dependent (99). Due to AR's role in PCa, the first-line treatment for PCa patients in the clinic is ADT. ADT includes different types of hormonal interventions, including bilateral orchiectomy, gonadotropin-releasing hormone (GnRH) antagonist, GnRH agonist, anti-androgens (e.g., bicalutamide and enzalutamide), adrenal androgen suppression (abiraterone) and combined androgen blockade (Figure 1-6). ADT, as a standard therapeutic strategy, is used to treat advanced PCa patients (i.e., those who are ineligible for surgical resection due to high-grade tumors and/or the presence of metastasis) and patients who experience biochemical recurrence (i.e., rising serum PSA levels after primary therapy) (125, 126). ADT regimens, such as GnRH, bicalutamide, enzalutamide, or abiraterone, effectively shrink primary tumors and decrease the serum PSA levels, at least initially (Figure 1-6) (127-129). Enzalutamide is a novel second-generation AR antagonist and functions via obstructing nuclear localization and chromatin binding of AR (130). Although enzalutamide increases patient overall survival by 4.8 months in clinical trials

for metastatic CRPC patients, enzalutamide-resistant disease occurs in < 2 years and progression is inevitably terminal (128, 129, 131).

Clinically, PCa patients are put on androgen blockade and/or antiandrogen treatment for many months and years until treatment failure, evidenced by rising serum PSA levels and radiographic progression. Such long-term treatment is frequently called chronic and persistent castration. In the laboratory setting, PCa cells are often acutely treated with various ADT protocols for hours to 1-2 days, which will not model the persistent castration in the clinic. In our lab, we modeled the clinical treatment by continuously exposing LNCaP PCa cells to castration regimens (dextran-stripped serum (CDSS), enzalutaminde, and CDSS with bicalutamide) and longitudinally tracked the dynamic changes in LNCaP cell subpopulations for up to 21 months (122). Under such 'chronic castration' conditions, we found that PCa cells underwent dynamic and cyclic molecular changes (122).

The therapeutic target of ADT and enzalutamide is AR. AR expression is heterogeneous in both primary tumors and CRPC, and changes dynamically overtime during tumor progression and treatment (132-134). In CRPC, AR expression is highly variable and not correlated with grade and stage (123, 135-138). Mechanisms that underlie development of resistance to AR-targeting therapies may include AR amplification increasing its sensitivity to its agonists, AR mutation leading to its activation by alternative hormone ligands (e.g., glucocorticoids), constitutively activated AR (ligand independent), upregulation of AR-independent pathways provoking downstream gene expression bypassing AR, atypical activation of AR by post-translational modifications, and aberrant expression and activity of AR co-factors

(99, 139). Of interest, AR regulates a distinct transcription program in androgen-independent PCa (compared to in hormone-naïve tumors), with distinct target genes such as cell cycle related UBE2C (140). Relevant to our work, castration enriches the undifferentiated $PSA^{-/lo}$ cell population with CSC properties (25, 123). Moreover, lower PSA mRNA levels in high-grade tumors are correlated with worse biochemical recurrence (BCR)-free and overall survival (123). Thus, the inherently castration-resistant PCSCs, whether pre-existent in untreated primary tumors or enriched in treatment-failed tumors, represent a clinically relevant therapeutic targets that require novel therapeutic strategies.

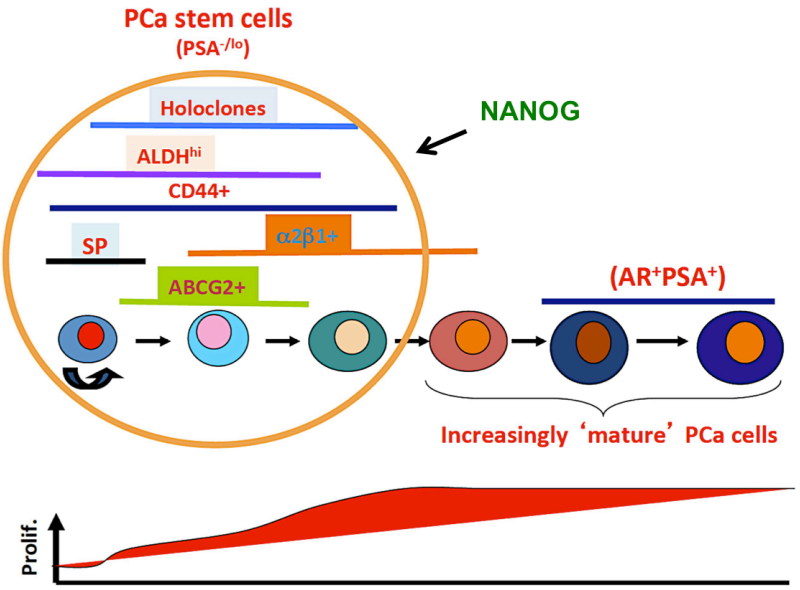


Figure 1-5. A hypothetical model of tumorigenic heterogeneity of human PCa cells (adapted from (123) Liu, X., X. Chen, K. Rycaj, H. P. Chao, Q. Deng, C. Jeter, C. Liu, S. Honorio, H. Li, T. Davis, M. Suraneni, B. Laffin, J. Qin, Q. Li, T. Yang, P. Whitney, J. Shen, J. Huang, and D. G. Tang. 2015. Systematic dissection of phenotypic, functional, and tumorigenic heterogeneity of human prostate cancer cells. *Oncotarget* 6: 23959-23986).

The $PSA^{-/lo}$ stem/progenitor cells pre-exist in untreated primary tumors. The $PSA^{-/lo}$ population is heterogeneous containing subsets of tumorigenic cells that express a spectrum of surface markers (e.g., CD44 and CD133), phenotypes (e.g., $ALDH^{hi}$ and side population (SP)) and stemness-promoting factors (e.g., NANOG). The $PSA^{-/lo}$ PCa cell is relatively quiescent (bottom panel) and can undergo asymmetric self-renewing cell divisions (arrow) generating both $PSA^{-/lo}$ and more differentiated, less tumorigenic PSA^{+} progeny (top). In experimental models, $PSA^{-/lo}$ cells possess long-term tumor-propagating activities and can preferentially survive castration. Also, the minor $PSA^{-/lo}$ population in primary PCa frequently becomes the predominant cell population in androgen-independent xenografts and clinical CRPC. NANOG has been shown to positively regulate certain CSC subsets (e.g., $CD44^{+}$ cells) in the $PSA^{-/lo}$ cell population.

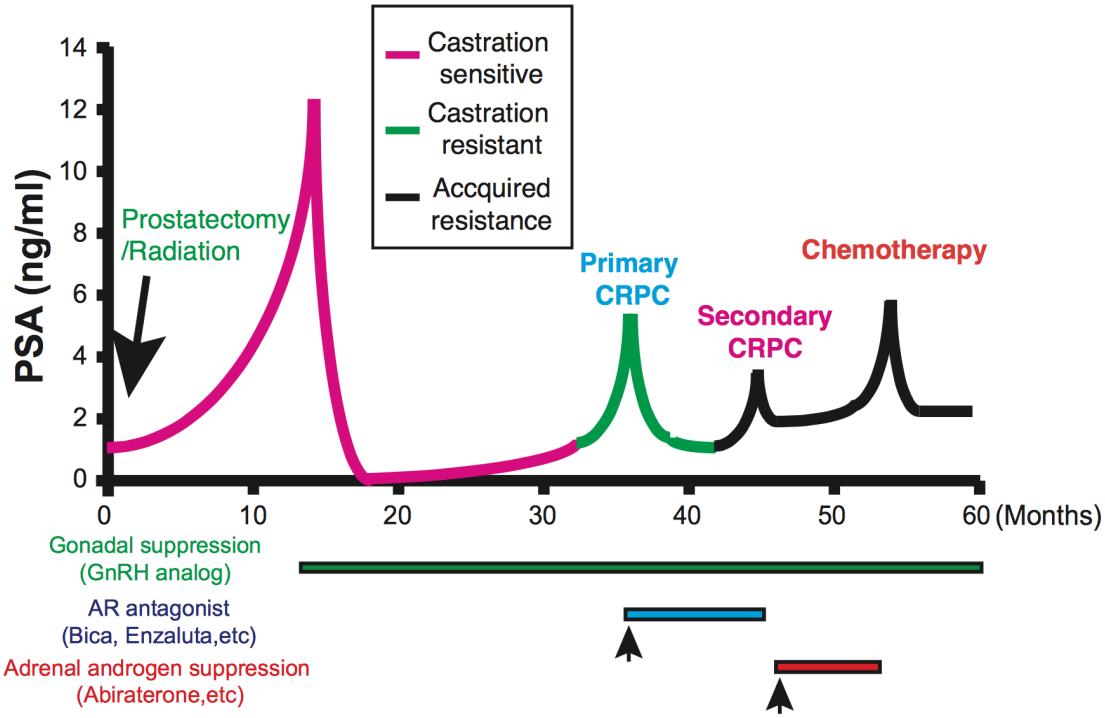


Figure 1-6. Schema of current clinical treatment for advanced and metastatic PCa.

Shown is a clinical treatment course of PCa patients by longitudinal tracking of serum PSA levels after patients are first treated by prostatectomy and/or radiation. The first-line hormonal therapy GnRH analogs with or without short-term AR antagonist are given when post-treatment serum PSA levels rise. The treatment is effective for about a year and the PSA levels rise again as recurrence of primary CRPC. The second-line castration regimen aims to suppress both AR function by AR antagonist and adrenal androgen production by abiraterone. The secondary CRPC appears after shorter intervals.

1.5 Bioinformatics

Bioinformatics, the application of computer algorithms to globally assess high-volume biological datasets, can be employed to help illuminate features of the subject being investigated. To dissect molecular characteristics and cell subtypes, analyses typically focus on the transcriptome but may also include the genome, metabolome, proteome, cistrome (Chromatin Immunoprecipitation Sequencing, ChIP-Seq), and epigenome (chromatin modifications) along with biochemical and molecular biology investigations. Recent advances in RNA-Sequencing (RNA-Seq) have provided researchers a powerful and revolutionary method to globally identify and quantify transcripts in particular cell subsets (or even single cells) to connect molecular phenotypes and biological properties (141). We have been applying transcriptome profiling to interrogate tumor cell heterogeneity and the molecular and functional underpinnings of CSC such as chemodrug or castration resistance (142). In Chapter 2, I shall describe in detail several major bioinformatics approaches and pipelines I have adopted in conducting my Ph.D. thesis research. Below I provide a very brief description of these approaches and pipelines.

RNA-Seq is a methodology of identification and quantification of relative transcript abundance between samples. Transcriptome profiles can be deciphered by comparison to known gene ontology (functional classifications) using bioinformatics tools, such as Ingenuity Pathway Analysis (IPA) and Genomic Regions Enrichment of Annotations Tool (GREAT). IPA interprets transcriptome profiles based on comprehensive knowledge of gene biological functions, canonical pathways and protein networks from published literature. GREAT similarly applies functional

categorization of TF-occupied cis-regulatory regions incorporating biological knowledge from public ontologies (143). In contrast, Gene Set Enrichment Analysis (GSEA) provides statistical modeling for evaluating concordance and distinctions in transcript profiles marking a particular biological state based on comparisons with previously published transcriptomes and gene signatures (144). Public gene expression profile databases, such as The Cancer Genome Atlas (TCGA), Oncomine and Gene Expression Omnibus (GEO) comprise patient transcriptional profiles, facilitating the discoveries of clinical relevance. TCGA collects high-quality and sufficient sample size for statistical interrogations. Oncomine contains various sample types, including normal tissues, primary tumors, metastases, and recurrent tumors, often associated with patient diagnosis or disease status. GEO is a repository with thousands of transcriptome datasets from many different studies and containing technical information such as experimental conditions, cell types, etc. In my studies I have utilized these bioinformatics tools widely to help decipher the biological meaning buried in the cancer cell transcriptomes and dissect tumor cell heterogeneity.

1.5.1 Transcriptome analysis used in dissecting cancer cell heterogeneity and in de-convoluting tumor complexity

Annotation of transcriptomic profiles help delineate cancer cell heterogeneity and predict prognosis in cancer types, such as acute myeloid leukemia (145), colorectal cancer (146), and glioma (147). For example, a recent study (148) applied multi-dimensional scaling analysis to dissect single-cell RNA-Seq from 430 cells in primary glioblastomas and showed that each individual tumor contains a spectrum of

cell populations with gene expression profiles of four glioblastoma subtypes, proneural, neural, classical and mesenchymal. Also, RNA-Seq analysis of single cells from each individual tumor revealed a gradient of association with a glioblastoma stem cell (GSC) stemness signature derived from differentially expressed gene (DEG) profiles of cultured stem-like cells (cultured spheroids that were enriched in GSCs) and serum differentiated neural-like cells (148). Thus, the transcriptome profiles represent and decrypt various molecular phenotypes of cancer cell subpopulations, including CSCs that co-exist in a given patient tumor.

Transcriptome profiling assists translating molecular knowledge to clinical relevance, including diagnosis, prognosis, and prediction, and facilitate designing and tailoring the therapeutic strategies. A recent study investigated the transcriptome profiles of hematopoietic stem cells versus leukemia stem cells to discover a commonly shared transcriptional program associated with canonical stem cell functions, including ABCB1, MEIS1, ERG, HLF, and MECOM (145). Strikingly, these stemness gene signatures also successfully predicted patient survival. Another recent study compared transcriptome changes in TNBC before and after neoadjuvant chemotherapy and uncovered two chemoresistance gene signatures (AKT1 signaling and hypoxia) associated with poor patient survival (17). These studies suggest that the patient transcriptomic profile can ‘divulge’ transcriptional reprogramming by therapies.

Whole-genome, whole-exome, and transcriptome analyses have been studied in association with clinical significance providing unbiased insights into cancer molecular mechanisms and revealing novel therapeutic targets in diverse cancers

(141, 149). Ultimately, such molecular dissections may lead to the discovery of new prognostic biomarkers and therapeutic strategies for personalized medicine in the clinic.

1.6 Hypothesis

Tumor cell heterogeneity and plasticity represent critical features of most clinical tumors and important mechanisms of therapy resistance. PCa is a highly heterogeneous malignancy that contains many phenotypically and functionally distinct cell subpopulations, which also dynamically change during tumor progression and in response to treatments. Our work has demonstrated that the phenotypically undifferentiated $PSA^{-/lo}$ PCa cell population harbors authentic PCSCs that are intrinsically resistant to castration and antiandrogens (25). Our work has also demonstrated that PCa cells manifest significant plasticity such that phenotypically differentiated PSA^{+} PCa cells can be reprogrammed to the castration-resistant, stem-like state by chronic castration (150) or overexpression of the stemness factor NANOG (151). These observations lead to my overarching hypothesis that *PCSCs possess intrinsic molecular and epigenetic features that help regulate their aggressiveness and stemness and contribute to tumor progression and therapy resistance, and that these CSC properties can be gained through epigenetic reprogramming of more differentiated cancer cells. Dissecting their transcriptome profiles will provide novel insights on disease progression and designing novel therapeutic strategies.*

To test this hypothesis and to execute my overall Ph.D. research plan, I applied bioinformatics approaches to analyzing the -omics data in 3 experimental systems.

First, I helped analyze the RNA-Seq and histone modification ChIP-Seq data in highly purified $PSA^{-/lo}$ vs. PSA^{+} PCa cell subpopulations, aiming to elucidate the intrinsic transcriptomic and epigenetic differences between the PCSC and non-CSC subpopulations (Chapter 3). This part of the work has not been published but we are in the process of preparing the manuscript. Second, I helped analyze both transcriptomic and ChIP-Seq data in NANOG-induced LNCaP cell reprograming, representing one aspect of cancer cell plasticity caused by stemness gene overexpression (Chapter 4). This part of my work was published during the course of my Ph.D. thesis research (151). Finally, I helped analyze RNA-Seq data in our models of treatment-induced PCa cell plasticity (Chapter 5). This work was recently published in *Nat. Commun* (152).

CHAPTER TWO

Materials and Methods

2.1 Databases

The Cancer Genome Atlas (TCGA)

The Cancer Genome Atlas (TCGA) (https://www.cancer.gov, https://portal.gdc.cancer.gov) has characterized multi-dimensional data types in 33 cancer types with hundreds of samples in each cancer type and 10 rare cancer types. The data types include DNA sequencing (e.g., whole exome sequencing, whole genome sequencing), miRNA sequencing (miRNA and isoform), mRNA sequencing (exon, gene, splice junctions, and isoform), total RNA sequencing (exon, gene, splice junctions, and isoform), array-based DNA methylation, and protein expression. TCGA database has also collected comprehensive clinical information, diagnostic image, and pathology reports. For prostate cancer (PCa), a total of 498 Prostate adenocarcinoma (PRAD) samples are available in TCGA, including 45 cases with GS6 (Gleason Score 6; Gleason Score is a pathology grading system the describes the overall degree of tumor differentiation and malignancy, the higher GS the less differentiated the tumoe), 246 with GS7, 62 with GS8, 131 with GS9 and 3 with GS10. There are also 52 pairs of matched PCa and benign/normal prostate tissues.

The TCGA is not only a database but also has extended online data analysis platforms, such as cBioPortal (http://www.cbioportal.org) and FireBrowse (http://firebrowse.org). To further analyze the detailed phenotypic features, the level 3 PRAD RNA-Seq and clinical data were downloaded in 2015. The transcription profile table was generated by merging all the *rsem.genes.normalized_results data with

FILE_SAMPLE_MAP.txt to get the TCGA barcodes and further determining the sample types. Of note, there are only 10 PCa patients expired in TCGA.

Oncomine™

Oncomine™ (Oncomine) (https://www.oncomine.com) is an online data-mining and visualization platform with database repository collecting 729 microarray datasets. The unique advantages of Oncomine are microarray platform integration and dataset diversification. For microarray platform integration, Oncomine standardizes processing, normalizing, and analyzing all datasets by unifying the analytical methods of cross-platform normalization and meta-analysis, which allow for gene centric analysis. For dataset diversification, Oncomine acquires datasets with various study designs and clinical information, which leads to different available analysis types, like cancer versus normal tissue, high-grade versus low-grade cancer, recurrent versus primary cancer, metastatic versus primary cancer, and different cancer subtypes. Also, Oncomine offers various analysis methods, like differential expression analysis, co-expression analysis, meta-analysis, and cancer outlier profile analysis. Differential expression analysis is the crux of Oncomine resource and applies *Student's t test* to determine the difference of two classes and *Pearson's correlation* to determine the difference of multiple ordinal classes. Co-expression analysis helps identify synchronous gene expression patterns, and it applies average linkage hierarchical clustering to identify sets of co-expressed genes. Meta-analysis is an extended function of differential expression analysis and the program orders the individual genes by the median rank across all selected datasets. The meta-analysis can be carried out with more than one gene, which is called concept analysis. Cancer outlier

profile analysis is used to identify potential oncogenes differentially expressed in a subset of samples due to the cancer heterogeneity. More importantly, the processed expression data and clinical information is available to download from Oncomine for detailed analysis. For PCa, there are 61 datasets with 3,820 samples including 18 datasets with cancer versus normal analysis (16 mRNA and 2 DNA), 16 datasets with metastasis versus primary analysis (13 mRNA and 3 DNA), 10 datasets with recurrence analysis, 22 datasets with grade type analysis, and 11 datasets with stage type analysis, and 5 datasets with survival analysis. Detailed information is organized in table 2-1.

Gene Expression Omnibus (GEO) (153)

Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo) is a public high-throughput functional genomics data repository, which archives both raw data and processed data from array and sequence platforms and associated descriptive metadata. The data in GEO follow the standard guidelines of Minimum Information About a Microarray Experiment (MIAME) (154) and Minimum Information about a high-throughput SEQuencing Experiment (MINSEQE) from Functional Genomics Data Society (http://www.fged.org), which in turn facilitate establishment of databases that could be interpreted and replicated for users. GEO was designed for high-throughput expression data but now is collecting data from studies on genome methylation, genome binding/occupancy, genome variation/copy number, chromosome conformation, and protein profiling. Moreover, GEO offers web-based tools and strategies for data analysis and visualization (GEO Profiles and Genome Data Viewer), advanced interactive analysis tool (GEO2R), and download tools. The most updated

statistics of GEO datasets is 108,705 series records with 2,819,723 samples in total and 1,428 series records with 20,609 samples in PCa. To simplify the processing of the datasets used in our studies, all the raw data of each dataset from GEO were manually downloaded. The information of all the datasets used in my studies was in table 2-2.

2.2 RNA-Seq analysis

RNA-Seq has replaced microarrays as the current 'standard' laboratory technique to routinely analyze global gene expression. Basic steps in an RNA-Seq experiment are presented in a flowchart (Figure 2-1A) and mainly consist of:

Library Preparation and Sequencing Platform: RNA-Seq libraries in our studies were constructed generally by using TruSeq Stranded mRNA following manufacturer's instructions (Illumina). The libraries were sequenced using 2x75 bases paired-end protocol on Illumina HiSeq 2000 instrument. In total, 28-39 million pairs of reads were generated per sample. Each pair of reads represents a cDNA fragment from the library.

Mapping: The total reads were mapped to mouse genome (mm10) or human genome (hg18, hg 19 or hg38) by TopHat (version 2.0.10). The TopHat prerequisite software tools included Bowtie2 (version 2.1.0), samtools (version 0.1.18), and Python (version 2.7.10). Both the expected inner distance and the standard deviation were set as 50 bp.

Differential Expression: The number of fragments mapped to each known gene from GENCODE Release M4 or 21 (https://www.gencodegenes.org/) was quantified using htseq-count from HTSeq package (version 0.6.1) (155). To avoid the transcriptional

noise, genes with fewer than 10 fragments in all the samples were removed before differential expression analysis. The differential expression between conditions was statistically assessed by R/Bioconductor package edgeR (version 3.0.8) (156) or DESeq (version 1.14.0) (157). Genes with false discovery rate (FDR) ≤ 0.05 and length > 200 bp were called as differentially expressed.

Sample Clustering: Hierarchical clustering of samples was performed using the cpm (count per million fragments) values of all the genes by hclust function in R. Euclidean distance and complete clustering method were used.

Pairwise Correlation: *Pearson's* correlation coefficient was computed by cor function in R using the cpm values with the top 5% most abundant genes across all the samples removed**.**

Principal Component Analysis (PCA): The function in stats package prcomp in R was applied. The cpm values for each gene across all the samples were centered to zero and scaled to have unit variance before the analysis took place.

Gene Landscape Profile: For the fragments that have both ends mapped, the first reads were kept. Together with the reads from the fragments that have only one end mapped, every read was extended to its 3' end by 200 bp in exon regions. For each read, a weight of 1/n was assigned, where n is the number of positions the read was mapped to. The sum of weights for all the reads that cover each genomic position was rescaled to normalize the total number of fragments to 1M and averaged over 10 bp resolution. The normalized values of each single gene were displayed using web-based University of California, Santa Cruz (UCSC) genome browser

(https://genome.ucsc.edu) or local Integrative Genomics Viewer (IGV, http://software.broadinstitute.org/software/igv/).

Functional Annotations: Ingenuity Pathway Analysis (IPA, https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis/) is based on Ingenuity Knowledge Base and serves to uncover and interpret the underlying meanings of the experimental context. The DEGs (Differentially Expressed Genes) were uploaded for Core analysis with Fisher's exact test to test the statistical significance of gene function and pathway enrichment. Canonical pathway analysis and upstream regulator analysis were applied in this study to predict the pathways changed based on gene expression changes and to predict the regulators causing changes in gene expression, respectively (Figure 2-1B). GSEA (http://software.broadinstitute.org/gsea/index.jsp, (144)) is a computational method to examine the statistic concordance of functionally unknown gene expression change profile and prior defined gene sets with biological information. It helps to interpret the genome-wide expression change profile and is linked to known biological functions and phenotypes (Figure 2-1C). To examine the gene signature enrichment, the normalized counts estimated by edge R (156) or DESeq (157) of all genes and replicates were included to compare to the known biological phenotypes as customized gene sets. To examine the GO term enrichment, only the DEGs were ranked and included to interrogate gene signatures in Molecular Signature Database (MSigDB, http://software.broadinstitute.org/gsea/msigdb/index.jsp, version6.0) of GSEA (Figure 2-1D). For the statistical test in GSEA, the permutation type was set as "gene_set" and all the other parameters were set as default.

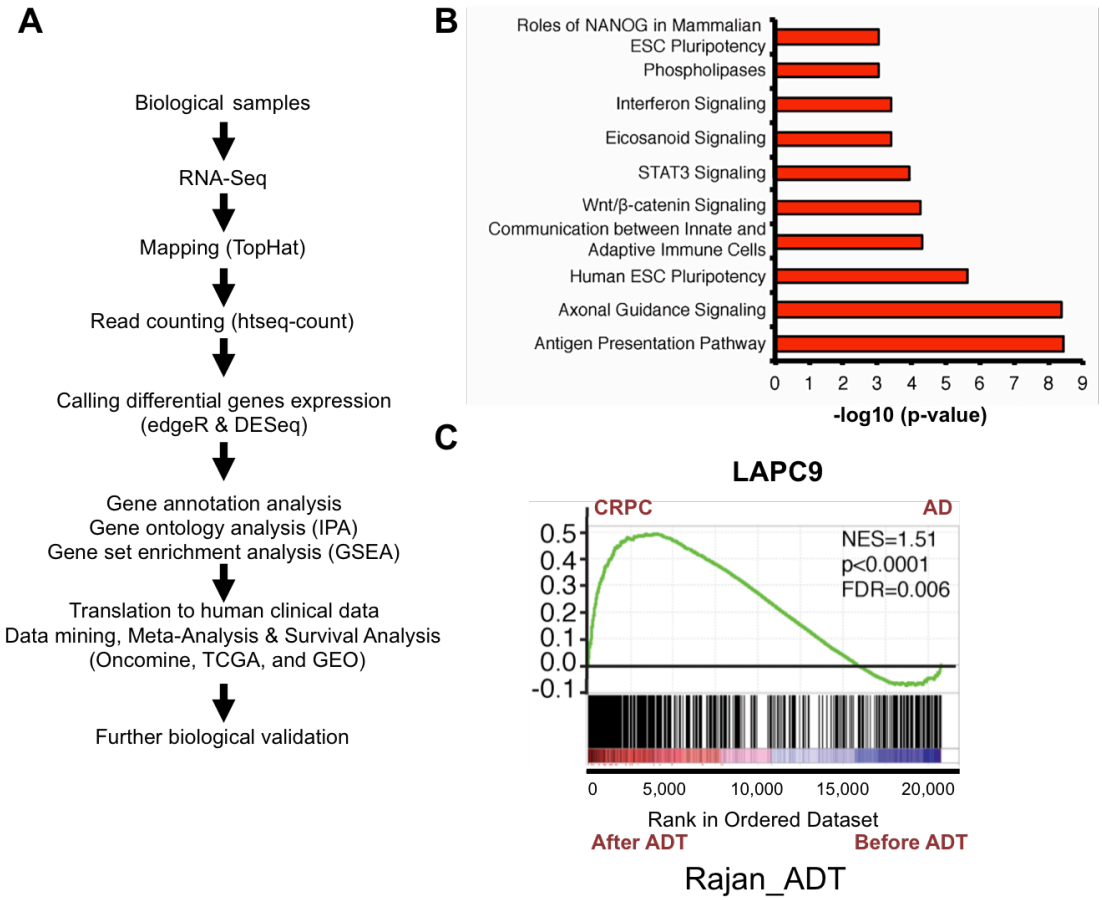


Figure 2-1. RNA-Seq analysis pipeline and gene annotation analysis

A. RNA-Seq pipeline. The RNA-Seq pipeline was developed by Next-Generation Sequencing core of MDACC at Science Park. TopHat and htseq-count were applied for read mapping and read counting, respectively. Both edgeR and DESeq were employed to identify the DEGs. Next, gene ontology analysis (Ingenuity® Pathway Analysis, IPA and gene set enrichment analysis, GSEA) is performed to assist gene annotation or interpretation of expression profiles into biological functions. Potential clinical significance of gene expression changes is investigated via data-mining in public databases, such as Oncomine, TCGA, and GEO.

B. An example of canonical pathway analysis by IPA. IPA identifies particular signaling pathways enriched in input DEGs. The height of the bar indicates -log10 p-value of enrichment for each pathway.

C. An example of GSEA plot. GSEA is designed to statistically test the concordance between expression change profile and the previously defined gene sets. The gene set is the one with defined biological phenotype (e.g., after and before ADT from Rajan dataset), and the expression change profile is the whole-DEG profile to be tested (e.g., CRPC and AD tumor lines derived from LAPC9). The results reveal that gene expression changes in LAPC9 CRPC (over AD tumors) are significantly associated with the gene expression changes of patient tumors after ADT (compared to the same patient tumors before ADT).

2.3 2D global GO enrichment analysis

The enrichment scores (ES) against all GO gene sets in Molecular Signatures Database (MSigDB) were computed by GSEA using the whole-transcriptome profile from each experiment. Then enrichment scores of the GO categories that were enriched in both experiments are presented as a scatter plot in gray dots (Figure 2-2A). The red dots indicate the GO term associated with particular GO terms (e.g., neuron/neuronal development and SC development).

2.4 DEG profile comparison

The log2 fold changes of all genes in the two experimental conditions are plotted as the gray dots in the scatter plot (Figure 2-2B). The red dots represent the genes of interest GO categories. For example, here we marked all the neurogenesis genes in red (Figure 2-2B). The violin plots displayed on the margins of the scatter plot are the distributions of log2 change values of the genes highlighted in red. Scatter plots and violin plots were drawn by ggplot2 and grid packages in R.

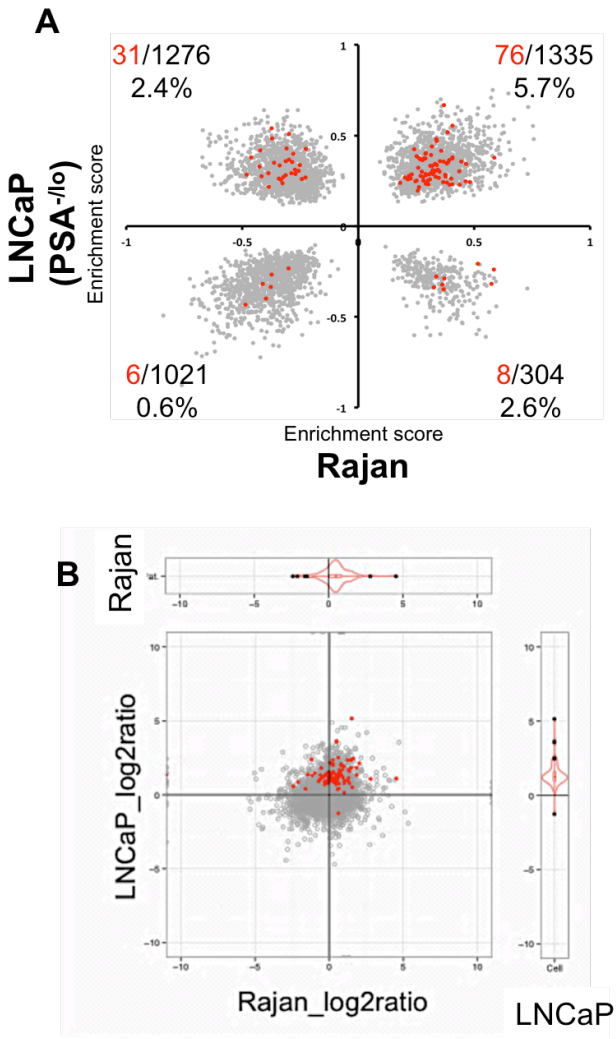


Figure 2-2. 2D global GO analysis and DEG profile comparison

A. 2D global GO analysis. The gray dots indicate all the GO terms enriched in both experimental conditons (i.e., LNCaP PSA$^{-/lo}$ and Rajan after ADT, Table 2-2), and the red dots are the GO terms associated with specific biological functions. The numbers on the figures show the number of specific terms (in red; i.e., neurogenesis related GO terms), the number of total GO terms (in black), and their ratios in each quadrant.

B. DEG profile comparison. The gray dots indicate the log2 fold change of each gene in the two experiments, and the red dots mark the interested genes. The violin plots on top and right panels are the distribution of log2 fold change values for the red dots.

2.5 Survival analysis

Single gene survival analysis

Survival analysis was performed and Kaplan-Meier survival plot was generated using the survfit function in R package survival. In brief, the normalized gene expression values of each individual gene from patient samples were imported with both survival time and survival status from Oncomine. The patient samples were ranked according to the gene expression. Each of the ranked samples from the first quartile to the third quartile was taken as the cutoff to assign the samples into high or low expression group and the one with the smallest p-value was chosen as the final cutoff.

Gene signature survival analysis

There were four PCa datasets (Nakagawa, Setlur, Taylor, and Yu) with overall survival information and six PCa datasets (Barwick, Glinsky, Holzbeierlein, Lapointe, Nakagawa, and Taylor) with biochemical recurrence (BCR) survival information in Oncomine. Among these datasets, Barwick and Nakagawa had very limited number of genes (487 genes), and Glinsky, Holzbeierlein, Lapointe, and Yu had relatively limited number of patients (79, 54, 112, 112 patients, respectively). The remaining datasets were Setlur with 363 samples and Taylor with 140 samples, and were used in our gene signature survival analysis. The gene signature survival analysis was conducted in two steps. First, the gene signatures were selected by meta-analysis comparison between primary and metastatic tumors in Oncomine. All the candidate genes with overall median rank < 2000 and p-value < 0.05 were included in the gene signature. Second, the largest cohort (Setlur) was used as the training set for the program to

decide the parameters of the predictor to assign the samples into binary risk classes (high or low risk). To create a predictor to classify subjects, the numeric score was computed by pairwise, ridge, LASSO linear regression using LMC package in R. The risk coefficients (i.e., the weight for each gene) were estimated by the pairwise linear coefficient (158). To define a threshold to separate high and low risk group, the optimized risk coefficients were determined as the ones with the best discrimination power computed by ROC (measured by AUC, Area under Curve of the combined predictor) in the training set. The Taylor dataset was used as an independent testing set to further examine the model.

2.6 In-house gene function tool

AmiGO (159) and in-house non-redundant GO analysis tool

AmiGO2 (http://amigo.geneontology.org/amigo) is one of the web-based tools under Gene Ontology Consortium (GOC) (http://www.geneontology.org). AmiGO2 has two parts: client software tools and data backend. The client interface is mainly written in JavaScript and scripting language, Perl, is applied to handle synchronous operations. Because of the increasing complexity of GO terms and improving efficiency of data queries, AmiGO2 has been moved from MySQL database backend to Solr search platform.

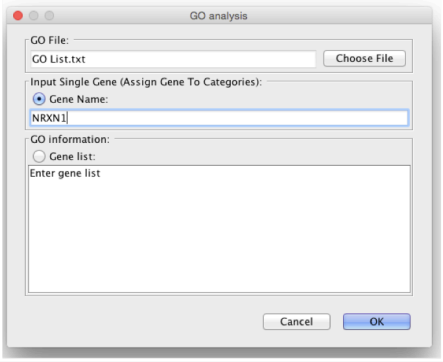
GO analysis sometimes is more complicated due to the redundancy of GO terms and different structure of GO interdependency. There are a few programs that are designed to reduce the redundancy (160) but can only minimize the redundant terms. To tailor a GO database for our research goal, we applied literature-based

manual curation to classify gene functions into non-redundant functional categories. To assist establishing the non-redundant GO database (Figure 2-3A-D) and data visualization (Figure 2-3E), I wrote a small program with graphical user interface (GUI) in Java. Because our definition of non-redundant GO is that one gene can only be signed into one category and user can decide categories, I applied HashMap to store the genes as keys and GO terms as values and ArrayList of HashMap to store the GO terms as keys and multiple genes as values. The program has two major functions. First, it can take user's input to establish the database. Second, it can make an inquiry of single gene or gene list and output the inquiry result of single gene or the pie chart of the gene list. Also, the program can connect to MySQL database of AmiGO by Object/Relational Map (ORM), Hibernate. In the future, the program will be updated to connect to Solr by SolrJ because of the update in AmiGO2 system.

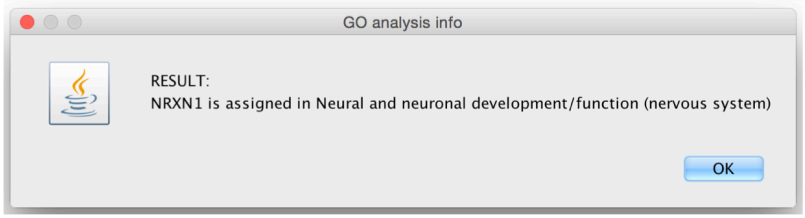
A


JAR
GO.jar

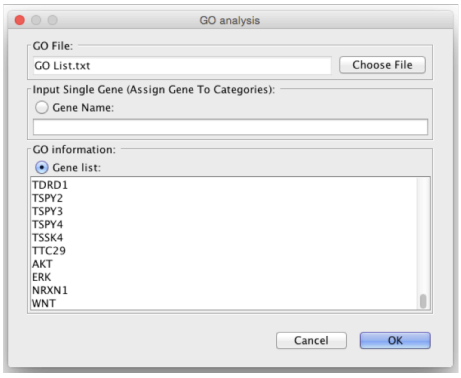
B


GO analysis
GO File:
GO List.txt
Choose File
Input Single Gene (Assign Gene To Categories):
Gene Name:
NRXN1
GO information:
Gene list:
Enter gene list
Cancel
OK

C


GO analysis info
RESULT:
NRXN1 is assigned in Neural and neuronal development/function (nervous system)
OK

D


GO analysis
GO File:
GO List.txt
Choose File
Input Single Gene (Assign Gene To Categories):
Gene Name:
GO information:
Gene list:
TDRD1
TSPY2
TSPY3
TSPY4
TSSK4
TTC29
AKT
ERK
NRXN1
WNT
Cancel
OK

E

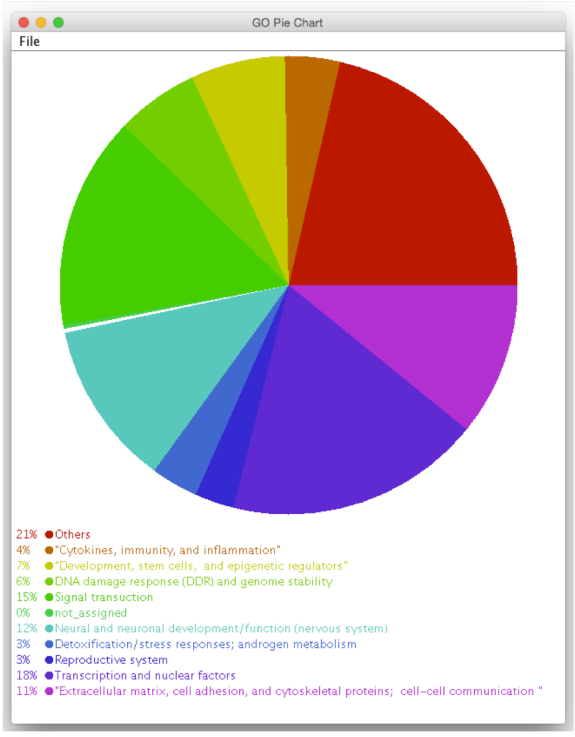

GO Pie Chart
File
21% Others
4% "Cytokines, immunity, and inflammation"
7% "Development, stem cells, and epigenetic regulators"
6% DNA damage response (DDR) and genome stability
15% Signal transuction
0% not_assigned
12% Neural and neuronal development/function (nervous system)
3% Detoxification/stress responses; androgen metabolism
3% Reproductive system
18% Transcription and nuclear factors
11% "Extracellular matrix, cell adhesion, and cytoskeletal proteins; cell-cell communication "

Figure 2-3. Non-redundant GO database tool

A. Java icon. Double click the JAVA icon to start the program

B and D. Inquiry GUI. The GO file function allows user to upload the GO annotation file. The GUI also includes two inquiry features for single gene searching and whole gene list searching.

C and E. The result windows. C is the result of single gene and D is the result of a gene list, showing as the pie chart and percentage of genes assigned in each categories.

Table 2-1. Prostate cancer datasets in Oncomine

Data	year	Name	# genes	# sample	Normal/Tumor (total 16)	Metastasis (pri/metastasis) (total 13)	Gleason Score (5/6/7/8/9/10)	patient death/Survival	recurrence (BCR free survival) (N/Y, 1,3,5 year)	Horm Refrac (N/Y)	ERG rearrangment (N/Y)	platform	publication
1	2009	Arredouani	19,574	21	8/13	NA	NA	NA	NA	NA	7/6	Human Gneome U133 Plus 2.0 Array	Clin Cancer Res 2009/09/15
2	2010	Barwick	487	139	NA/139	NA	2/36/90/5/6/0	NA	88/13, 51/23, 21/27	NA	70/69	Illumina DASL Human Cancer Panel	Br J Cancer 2010/02/02
3	2005	Best2	12,624	20	NA	NA	0/0/3/5/0/0	2 year: 3/5	NA	10/10	NA	Human Genome U133A Array	Clin Cancer Res 2005/10/01
4	2005	Bittner	19,574	60	NA	59/1	NA	NA	NA	NA	NA	Human Genome U133 Plus 2.0 Array	Not Published 2005/01/15
5	2007	Chandran	14,738	31	NA	10/21	0/0/10/0/0/0	NA	NA	NA	NA	CodeLink UniSet Human 20K I Bioarray	BMC Cancer 2007/04/12
6	2004	Glinsky	12,624	79	NA/79	None:17,Focal:6, Invasive: 18, Established: 38	2/15/44/10/8/0	NA	NA, 53/26, 42/32	NA	NA	Human Genome U133A Array	J Clin Invest 2004/03/01
7	2012	Grasso	19,189	122	28/59	59/35	0/0/2/1/0/0 (2 neuroendocrine differntiation sample)	31/NA, 28/3, 18/12, (include followup time, and dead: 35)	NA	NA	32/27	Agilent Human Genome 44K	Nature 2012/05/20
8	2004	Holzbeierlein	8,603	54 (1 cell line)	4/40	40/9	0/0/2/4/1/0	41/NA, 33/1, 13/3	41/2, 33/4, 13/10	6/3	NA	Human Genome U95A-Av2 Array	Am J Pathol 2004/01/01
9	2004	Lapointe	10,166	112	41/62	62/9	0/24/22/10/5/0	NA	14/3,NA,NA (at 27th month 22/7)	NA	NA	Platform not pre-defined in Oncomine	Proc Natl Acad Sci U S A 2004/01/20
10	2002	LaTulippe	8,603	35	3/23	23/9	0/2/14/2/5	NA	14/9	NA	NA	Human Genome U95A-Av2 Array	Cancer Res 2002/08/01
11	2006	Liu	12,624	57	13/44	NA	0/13/16/10/0/0	NA	NA	NA	NA	Human Genome U133A Array	Cancer Res 2006/04/15
12	2001	Luo	5,064	25	NA (9/16, precursor and PCa)	NA	NA	NA	NA	NA	NA	Platform not pre-defined in Oncomine	Cancer Res 2001/06/15
13	2002	Luo2	15,302	30	15/15	NA	2/3/7/1/2/0	NA	NA	NA	NA	Hu35KsubD Array, Hu35KsubC Array, Human Genome U95A-Av2 Array, Hu35KsubB Array, Hu35KsubA Array	Mol Carcinog 2002/01/01
14	2001	Magee	5,338	15	4/8	8/3	NA	NA	NA	NA	NA	HumanGeneFL Array	Cancer Res 2001/08/01
15	2008	Nakagawa	487	596	NA	NA	4/79/290/69/145/9	594/2, 579/15, 522/55	581/13, 543/51, 476/116	NA	NA	Illumina DASL Human Cancer Panel	PLoS ONE 2008/05/28
16	2008	Nakagawa2	449	596	NA	NA	4/79/290/69/145/9	594/2, 579/15, 522/55	581/13, 543/51, 476/116	NA	NA	Platform not pre-defined in Oncomine	PLoS ONE 2008/05/28
17	2006	Nanni	12,624	26 pri cultures, 4 cell lines	4/23 (3 precusor)	NA	0/4/17/0/1/0	NA	17/5, 1/7, NA	NA	NA	Human Genome U133A Array	Mol Cancer Res 2006/02/01
18	2001	Ramaswamy	9,885	288	NA	10/4	NA	NA	NA	NA	NA	HumanGeneFL Array, Hu35KsubA Array	Proc Natl Acad Sci U S A 2001/12/18
19	2003	Ramaswamy2	9,885	76	NA	10/3	NA	NA	NA	NA	NA	HumanGeneFL Array, Hu35KsubA Array	Nat Genet 2003/01/01
20	2008	Setlur	6,084	363	NA	NA	0/110/153/41/54/0	357/6, 317/41, 277/86	NA	NA	292/62	Illumina DASL Transcriptionally Informative Gene Panel	J Natl Cancer Inst 2008/06/04
21	2008	Setlur2	6,084	109	NA	NA	NA	NA	NA	NA	60/41	Illumina DASL Transcriptionally Informative Gene Panel	J Natl Cancer Inst 2008/06/04
22	2002	Singh	8,603	102	50/52	NA	4/15/29/2/2/0	NA	42/8	NA	NA	Human Genome U95A-Av2 Array	Cancer Cell 2002/03/01
23	2007	Tamura	16,459	35	NA	23/12	NA	NA	NA	10/25	NA	Platform not pre-defined in Oncomine	Cancer Res 2007/06/01
24	2010	Taylor3	22,238	185 (6 cell lines)	29/131	131/19	0/41/74/8/7/0	131/NA, 110/1, 59/2 (131/19)	115/9, 86/21, 37/24	NA	NA	Platform not pre-defined in Oncomine	Cancer Cell 2010/07/13
25	2007	Tomlins	10,656	101	23/30 (precursor:17)	NA	0/10/8/5/7/0	NA	NA	3/16	NA	Platform not pre-defined in Oncomine	Nat Genet 2007/01/01
26	2006	True	3,432	31	NA	NA	0/4/21/1/5/0	NA	NA	NA	NA	Platform not pre-defined in Oncomine	Proc Natl Acad Sci U S A 2006/07/18
27	2003	Vanaja	17,779	40	8/27	27/5	0/12/0/0/15/0	NA	NA	NA	NA	Human Genome U133A Array, Human Genome U133B Array	Cancer Res 2003/07/15
28	2005	Varambally	19,574	19	6/7	7/6	NA	NA	NA	NA/6	NA	Human Genome U133 Plus 2.0 Array	Cancer Cell 2005/11/01
29	2008	Wallace	12,603	89	20/69	NA	1/17/48/1/2/0	NA	NA	NA	NA	Human Genome U133A 2.0 Array	Cancer Res 2008/02/01
30	2001	Welsh	8,603	34	9/25	NA	0/7/9/5/2/0	NA	NA	NA	NA	Human Genome U95A-Av2 Array	Cancer Res 2001/08/15
31	2004	Yu	8,603	112	23/65	65/24	3/15/27/7/12/0	60/NA, 34/2, 9/3	NA	NA	NA	Human Genome U95A-Av2 Array	J Clin Oncol 2004/07/15

Table 2-2. Overall information of datasets used in the studies.

Dataset name	Accession no.	Data type	Platform	Sample type	Experimental design	Publication
Beltran_NEPC	N/A	RNA-Seq	Illumina GA II Sequencer	benign + PCa + NEPC	6 benign prostate tissue, 30 PCa and 7 NEPC tumors	Cancer Discov. 2011
Liu_PSA^{-lo}/PSA^{+}	N/A	RNA-Seq	Illumina HiSeq2000 system	PCa cell lines	PSA-/lo and PSA+ cells derived from LNCaP and LAPC9 PCa cell lines	Chapter 3, manuscript in preparation
Rajan_ADT	GSE48403	RNA-Seq	HiSeq 2000 sequencer	PCa before + after ADT	14-paired locally advanced or metastatic PCa samples with before and after ADT from 7 patients	Eur Urol. 2014
Sun_recurrence	GSE25136	microarray	Affymetrix U133A human gene array	PCa + PCa-recur	40 non-recurrent primary prostate tumors and 39 recurrent tumors	Prostate. 2009
Varambally_Normal/Tumor and Primary/Metastasis	GSE3325	microarray	Affymetrix U133 2.0 Plus arrays	benign + PCa + PCa-meta	6 pooled samples from benign prostate, primary and metastatic PCa and 13 individual benign prostate, primary and metastatic PCa tissues	Cancer Cell. 2005
Wang_AD/AI	SRA053575	RNA-Seq	Illumina Genome Analyzer II system	AD +AI PCa cell lines	AD LNCaP cells and AI LNCaP cells cultured in medium with antiandrogen flutamide and collected the cells at passage 110 generation	Cancer Lett. 2014
Zhang_Basal/luminal	GSE67070	RNA-Seq	Illumina HiSeq2000 system	normal prostate	30 patient basal and luminal cells derived from benign prostate tissues	Nature Communication. 2016

AD: androgen dependent
ADT: androgen deprevation therapy
AI: androgen independent
NEPC: neuroendocrine prostate cancer
PCa: prostate cancer
PCa-recur: recurrent prostate cancer
PCa-meta: metastatic prostate cancer

CHAPTER THREE

Bioinformatics analyses reveal intrinsic transcriptomic and epigenetic differences in PSA$^{-/lo}$ and PSA^{+} PCa cells

3.1 Introduction

Prostate specific antigen (PSA) is a 34 KDa glycoprotein enzyme belonging to the kallikrein (*KLK*) gene family. A secretory component of prostatic seminal fluid, the role of PSA in prostate adenocarcinoma is obscure, although PSA is a known PCa biomarker. Immunohistochemistry studies have provided evidence that the degree of the patient's prostate tumor differentiation is positively correlated with the expression levels of PSA protein in PCa cells (161), and undifferentiated PCa cells express low or undetectable PSA (PSA$^{-/lo}$). In clinical specimens, PSA$^{-/lo}$ cells appear rarely in early-stage tumors but become more abundant in late-stage and locally advanced PCa and become the predominant cell population in treatment-failed castration-resistant PCa or CRPC (25, 123). Of clinical significance, PCa patients with tumors harboring more than 50% (phenotypically differentiated) PSA^{+} cells survive longer (138, 162), and higher intra-tumoral PSA (*KLK3*) mRNA levels have been shown to predict better patient survival (123).

Like most cancers, PCa exhibits significant heterogeneities in cellular morphologies and histopathological structures manifested as a mixture of multiple tumor foci displaying various degrees of differentiation (25, 134). Intriguingly, even seemingly 'homogeneous' PCa cell lines still exhibit heterogeneous subpopulations. For example, our previous study demonstrated ~2-6% of PSA$^{-/lo}$ cells that pre-exist in

unperturbed LNCaP human PCa cultures, and the percentage of $PSA^{-/lo}$ cells markedly increases in response to castration treatments (25, 123). Interestingly, despite PSA being a known AR-regulated target gene, discordant expression of AR and PSA has been observed in patient tumors and xenografts, discernable as $AR^{+}PSA^{+}$, $AR^{+}PSA^{-}$, $AR^{-}PSA^{+}$, and $AR^{-}PSA^{-}$ phenotypes (123).

Through a spectrum of *in vitro* biological and *in vivo* tumor regeneration assays, we have observed that a fraction of the cells in the $PSA^{-/lo}$ cell population possesses properties attributable to normal stem cells (SCs) and cancer stem cells (CSCs), including the ability to undergo asymmetric renewing cell division to generate both $PSA^{-/lo}$ and PSA^{+} cells, preferential expression of SC genes, relative dormancy, and long-term tumor-propagating capacity (25). The $PSA^{-/lo}$ cell population is androgen-independent and refractory to androgen-deprivation therapy (ADT), and harbors highly tumorigenic CRPC cells (123). These functional studies (25, 123) suggest that $PSA^{-/lo}$ cells not only have CSC properties but may also be drivers of CRPC, and thus may represent therapeutic targets to treat CRPC.

In both SCs and CSCs, bivalent chromatin regulates cell-fate specification and self-renewal capacity (163). Bivalency describes simultaneous presence of the active (H3K4me3) and repressive (H3K27me3) histone occupancy of the gene promoters. Genes associated with the bivalent histone marks and a lack of DNA methylation on CpG island in their promoters are thought to be 'poised' to regulate cell differentiation (73, 164, 165). Cancer cells in general and CSCs in particular may acquire a certain degree of plasticity by disrupting the "double-lock" mechanism that silences the gene via both DNA methylation and histone modifications, which together establishes an

epigenetic barrier to prevent normal differentiated cells from moving back to the undifferentiated state (166). Plasticity of cancer cells has been surmised to contribute to tumor cell heterogeneity in breast (163, 167), colon (168, 169), and brain (170-172) cancers by generating cells with a range of phenotypic differentiation states and distinct clonal subsets. Transient changes in the epigenetic landscape have also been linked to therapy resistance in CSCs, such as the reversible acquisition of a drug-tolerant CSC subset via IGF-1R signaling and an altered chromatin state that requires the KDM5/Jarid1A histone demethylase (173). That these drug-tolerant 'persistors' are sensitive to the histone deacetylase inhibitor trichostatin A and an IGF-1R inhibitor (173) provides the proof of principle that subsets of CSCs may be targetable by epigenetic modulations.

3.2 Hypothesis

Previous work in our lab has demonstrated that the $PSA^{-/lo}$ cell population pre-exits in untreated patient tumors, becomes enriched in CRPC, and is negatively associated with patient overall survival (25, 123). Of clinical importance, the $PSA^{-/lo}$ cell population possesses CSC properties, including long-term tumor propagating capacity and the ability to regenerate and propagate CRPC in androgen-ablated hosts. The study has also revealed bivalent histone associations in the promoters of several select SC genes preferentially in the $PSA^{-/lo}$ cell population (123). Unknown is whether the $PSA^{-/lo}$ cell population is governed by regulation of several specific SC genes or whether the entire transcriptome is distinct from lineage-related but more differentiated PSA^{+} cells. Therefore, my hypothesis herein is that *$PSA^{-/lo}$ cells possess*

unique intrinsic transcriptomic and epigenetic features that causally regulate their stemness and aggressiveness. To test this hypothesis, we purified $PSA^{-/lo}$ and PSA^{+} cell populations from untreated LNCaP cell cultures and performed RNA-Seq and H3K4me3 and H3K27me3 ChIP-Seq experiments. Also, I have performed integrative bioinformatics analysis to further dissect the transcriptome profile of $PSA^{-/lo}$ cell population and its potential association with bivalently marked promoters.

3.3 Materials and Methods

Biological model

The lentiviral vector construct with a green fluorescent protein (GFP) reporter driven by the PSA promoter was packaged into lentiviral particles produced in 293FT packaging cells (Figure 3-1A). Regularly cultured bulk LNCaP cells were infected with the lentivirus harboring the reporter plasmid (Figure 3-1B). The infected cells were purified by flow cytometry into GFP^{+} (PSA^{+}) and $GFP^{-/lo}$ ($PSA^{-/lo}$) subsets 48-72 hours after infection (Figure 3-1C).

RNA extraction, RNA-Seq, and RNA-Seq analysis

Total RNA was extracted from purified PSA^{+} and $PSA^{-/lo}$ LNCaP cells. RNA-Seq library was prepared using the TruSeq Stranded Total RNA Library Preparation Kit with Ribo-Zero (#20020599, Illumina) following manufacturer's protocol, and 50 bp pair-end sequencing was performed on an Illumina HiSeq2000 system.

ChIP-Seq

PSA^{+} and $PSA^{-/lo}$ LNCaP cells were sorted and 100,000 purified cells were pooled together (from multiple sorts, as needed) for each ChIP assay. ChIP-Seq was

performed using antibodies specific for H3K4me3 (04-745, Millipore) and H3K27me3 (07-449, Millipore). ChIP-Seq libraries were generated using NEBNext ChIP-Seq Library Prep Master Mix following the manufacturer's protocol (New England Biolabs), and single-end 38 bp sequencing was performed on an Illumina HiSeq2000 system.

Bioinformatics analysis

For RNA-Seq, the reads were aligned to human genome (hg19) by TopHat (version 2.0.3) with parameter "-r 50 --mate-std-dev 50 --library-type fr-unstranded -G". The transcript assembly was performed by using cufflink (version 2.0.1) with input bam file from TopHat and parameters "-u -g gene.gtf". In total, there were around 88M-92M reads with 80% mapping rates. Then, the cuffdiff was performed to test DEG analysis with parameters "--library-type fr-firststrand", resulting in 1,075 genes upregulated in $PSA^{-/lo}$ cells and 285 genes upregulated in PSA^{+} cells with cutoff fold change ≥ 2.

Gene ontology (GO) analysis was performed by manually assigning the genes into non-redundant functional categories with the assistance of an in-house java program. This graphical user interface (GUI) program helped establish the non-redundant GO database by virtue of an extension function to search for GO terms of genes in AmiGO, an online gene ontology database ((174), http://amigo.geneontology.org/amigo) and facilitate data visualization. Global 2D GO analysis was applied by Molecular Signatures Database (MSigDB) of GSEA to perform unbiased whole-transcriptome analysis in comparison with primary prostate tumor (TCGA) dataset or with ADT failed prostate tumor (Rajan_after_ADT) dataset (175). GSEA was also applied for exploring the associated biological phenotypes. Whole-transcriptome profile analyses were also performed to compare the $PSA^{-/lo}$ cell

profile and TCGA dataset or the Rajan_after_ADT dataset. Also, the data used for gene expression heatmaps and survival analysis were from Oncomine. The details of GO and data-mining analyses were described in Chapter 2.

For ChIP-Seq analysis, raw reads were aligned to the hg19 genome assembly using BWA (version 0.6.2) with sub-command "aln" and default parameters. There were a total of 40-48M reads and 81%-84% mapping rate for H3K4me3 and 41-51M and 77%-79% mapping rate for H3K27me3. Only tags that uniquely mapped to the genome were used for further analysis. ChIP-Seq peaks were identified using MACS (version 1.3.7.1). The scanning window size was set for 1,000 bp and the cutoff of was $1e^{-5}$. The background was the input DNA controls from $PSA^{-/lo}$ and PSA^{+} cells. To summarize, in the promoter region (defined as ±1kb from transcription start site (TSS)), 21,089 and 20,347 H3K4me3 peaks and 24,411 and 25,352 H3K27me3 peaks were identified from $PSA^{-/lo}$ and PSA^{+} chromatin, respectively. For heatmap distribution, the peaks were centered by middle of the peak with a range of ±5kb. Whole regions were divided into 250 bins, and the average reads per kilobase transcript per million reads (RPKM) were calculated in each bin and plotted by heatmap.2 function in R. GO analysis was performed using GREAT (version 2.0, http://great.stanford.edu/public/html/).

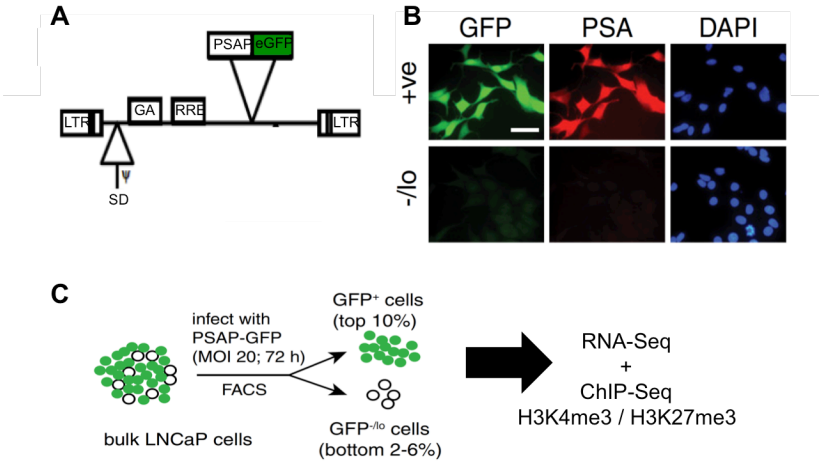


Figure 3-1. Strategy to purify and characterize subpopulations of LNCaP cells on the basis of PSA expression (adapted from (25) Qin, J., X. Liu, B. Laffin, X. Chen, G. Choy, C. R. Jeter, T. Calhoun-Davis, H. Li, G. S. Palapattu, S. Pang, K. Lin, J. Huang, I. Ivanov, W. Li, M. V. Suraneni, and D. G. Tang. 2012. The PSA(-/lo) prostate cancer cell population harbors self-renewing long-term tumor-propagating cells that resist castration. *Cell stem cell* 10: 556-569).

A. The PSAP-GFP lentivector reporter used in this study. GA. gag gene; eGFP, enhanced green fluorescence protein; LTR, long-terminal repeat; PSAP, prostate specific antigen promoter; RRE, Rev responsive element; SD, splice donor.

B. Representative fluorescence images of LNCaP cells transduced with the *PSAP*-GFP reporter and immunostained for PSA in both PSA^{+} and $PSA^{-/lo}$ cells, showing that the PSA reporter system accurately reports PSA expression in cells.

C. Schema of flow-based purification of PSA^{+} and $PSA^{-/lo}$ LNCaP cells and molecular analysis by RNA-Seq and ChIP-Seq. LNCaP cells with strong PSA expression (i.e., the top 10% of the $GFP^{+/bright}$ population) and with low or no PSA expression (i.e., the bottom 2-6% of $GFP^{-/lo}$ cells) were collected by flow cytometry and the transcriptome assessed by RNA-Seq and the epigenetic landscape by H3K4me3 and H3K27me3 ChIP-Seq.

3.4 Results

To understand the potential biological differences between the PSA^+ and $PSA^{-/lo}$ cell populations, I performed manually curated non-redundant gene ontology (GO) analysis to assign each DEG into the most relevant gene category with the assistance of an in-house program developed specifically for GO annotation (25). Analysis of the DEG profiles revealed significant upregulation of genes involved in development/stemness, epigenetic regulation, DNA damage response and genome stability, detoxification and stress responses, extracellular matrix and cell-cell communications, as well as neural and neuronal development/functions in the $PSA^{-/lo}$ cell subset. In contrast, genes that participate in metabolic processes, signal transduction, and cellular differentiation and many AR target genes, were enriched in the PSA^+ cell population (Figure 3-2A). GSEA suggested that the transcriptome profile of $PSA^{-/lo}$ cells, but not PSA^+ cells, was associated with normal SCs and CRPC (Figure 3-2B). These unique DEG profiles in $PSA^{-/lo}$ and PSA^+ cell subpopulations portend that, despite their identical genetic background, the two cell subpopulations likely possess distinct biological functions.

Unbiased global GO enrichment analysis demonstrated that the $PSA^{-/lo}$ LNCaP cells had more GO terms (gray dots) with a positive enrichment score (the ratio of GO terms in the 1^{st} and the 2^{nd} quadrant to those in the 3^{rd} and the 4^{th} quadrant: 2,604:1,367 (1.91; left panel) and 2,611:1,325 (1.97; right panel)) (Figure 3-2C). On the other hand, fewer GO terms with a positive enrichment score were observed in primary patient samples (the ratio GO terms in the 1^{st} and the 4^{th} quadrant to the 2^{nd} and the 3^{rd} quadrant: 1,267:2,701). On the other hand, patient CRPC displayed an

enrichment pattern more similar to that derived from PSA$^{-/lo}$ cells, with more GO terms with a positive enrichment score than primary patient tumors (the ratio of GO terms in the 1st and the 4th quadrant to the 2nd and the 3rd quadrant: 1,639:2,297). These observations may indicate highly active transcriptome programs in CSCs and relatively 'inert' transcriptomes in primary PCa. Intriguingly, very much like the AR^{-} PSA^{-} normal human prostate basal/stem cells that preferentially express a neurogenesis 'program' (112), the PSA$^{-/lo}$ LNCaP cells also preferentially expressed neurogenesis (i.e., neural and neuronal development/function) genes (Figure 3-2C; red dots). The data is consistent with results from our non-redundant GO manual curation (Figure 3-2A) showing enrichment of neurogenesis genes in PSA$^{-/lo}$ cells. When comparing global GO enrichment patterns of DEGs between PSA$^{-/lo}$ cells versus patient primary tumor or patient CRPC, neural and neuronal development/function terms (red dots) preferentially showed in quadrant 1 in comparison of PSA$^{-/lo}$ cells versus CRPC (Figure 3-2C), meaning that the neurogenesis GO terms are enriched in both PSA$^{-/lo}$ LNCaP cells and patient CRPC samples, but not primary tumors.

To establish potential clinical relevance and biological meaning of neurogenesis gene enrichment, the global transcriptome (i.e., all genes) in PSA$^{-/lo}$ LNCaP cells was analyzed by scatter plots and compared to the transcriptome profiles of patient primary and CRPC samples. The overall transcriptomes were evenly distributed, and the neurogenesis genes were mostly upregulated in PSA$^{-/lo}$ LNCaP cells and CRPC (quadrant 1 in right panel) but not in primary patient tumors (left panel) (Figure 3-3A). Moreover, the Oncomine Concept analysis revealed that the expression levels of

many neurogenesis genes were increased in primary patient tumors over normal tissues and in metastatic tumors compared to primary tumors (Figure 3-3B). Strikingly, Kaplan-Meir survival analysis showed that the mRNA expression levels of several neurogenesis genes (e.g., NRXN1 and CHRM3) correlated with poor patient survival (Figure 3-3C). These results, collectively, demonstrate that the neurogenesis gene expression profile, preferentially expressed in $PSA^{-/lo}$ PCa cells, is associated with clinical manifestations of CRPC, and advanced and metastatic PCa, and predicts worse patient survival.

Because the $PSA^{-/lo}$ and PSA^{+} PCa cells are lineage-related (25) and some SC gene promoters are enriched in bivalent (dual activating and repressive) histone modifications (123), the global active (H3K4me3) and suppressive (H3K27me3) histone marks on promoter regions in the two cell subpopulations were investigated by ChIP-Seq. There were 2,088 peaks associated with 1,528 genes preferentially marked by H3K4me3 from analysis of $PSA^{-/lo}$ cell chromatin and 1,346 H3K4me3 peaks associated with 1,194 genes called from analysis of PSA^{+} cell chromatin. The majority of peaks (19,001 peaks associated with 12,993 genes) were commonly marked by H3K4me3 in both cell populations (Figure 3-4A), with 33.7% peaks from the promoter region, 7.5% in coding exons, 28.3% in introns, and 30.4% located at distal intergenic regions (Figure 3-4B). The heatmap revealed notably more differential H3K4me3 signal in comparison of the peaks preferentially enriched in $PSA^{-/lo}$ chromatin (peaks called only in $PSA^{-/lo}$ cells), whereas the ChIP-Seq peak intensity preferentially enriched in PSA^{+} cells only differed modestly (Figure 3-4D). GO analysis revealed that the gene functions associated with activate histone marks in $PSA^{-/lo}$ cells were related

to cell development, cell migration, and neuron migration, whereas gene functions associated with active histone marks in PSA^+ cells corresponded to cell differentiation, cell fate specification, and hormone stimulation (Figure 3-4C).

The suppressive histone mark, H3K27me3, was enriched in 1,010 peaks associated with 604 genes in $PSA^{-/lo}$ cells and 1,951 peaks (1,103 genes) in PSA^+ cells (Figure 3-4E). The majority of peaks (23,401 peaks associated with 5,702 genes) were commonly marked by H3K4me3 in both cell populations (Figure 3-4E), with 9.1% of the repressive H3K27me3 peaks were localized to promoter regions, 6.6% on coding exons, 26.9% on introns, and 57.4% were located in distal intergenic regions (Figure 3-4F). Biological annotation of the genes marked by suppressive histone modification in $PSA^{-/lo}$ cells revealed their association with GO terms such as programmed cell death, steroid biosynthesis and metabolism, and cell adhesion. On the other hand, biological functions associated with stem cell development, Notch pathway and neuronal maturation were detected more prominently in H3K27me3-associated genes in PSA^+ cells (Figure 3-4G). The heatmap revealed notably more differential H3K27me3 signal in comparison of the peaks preferentially enriched in PSA^+ chromatin (Figure 3-4H). Thus, $PSA^{-/lo}$ and PSA^+ cells purified from the same untreated bulk LNCaP culture manifest quite different epigenetic landscapes reflecting their distinct transcriptome profiles.

Bivalent domains, particularly those associated with promoters of developmental genes, are a pivotal feature of embryonic SCs (176). A previous study from our lab demonstrated that several development-associated genes, including NKX3.1, FGF5, BCL-2, and CDH2, have bivalent histone modifications preferentially

in PSA$^{-/lo}$ cells (123). Here, we examined the genome-wide bivalency association within the ± 1 Kb of TSS in PSA$^{-/lo}$ vs. PSA^{+} LNCaP cells. The majority (96%) of TSS marked by H3K4me3 on gene promoters in PSA$^{-/lo}$ cells were also marked with H3K4me3 in PSA^{+} cells, and roughly 75% of TSS marked by H3K27me3 in PSA$^{-/lo}$ cells were also marked with H3K27me3 in PSA^{+} cells. However, only 44% of the bivalent TSS in PSA$^{-/lo}$ cells maintained their bivalency in PSA^{+} cells, with 19% losing their H3K27 repressive mark (i.e., H3K4me3 only), 35% losing their H3K4me3 activating mark (H3K27me3 only), and 2% losing both activating and repressive histone modifications (unmarked) (Figure 3-5A). In total, there were 1,126 bivalent marked TSS in PSA$^{-/lo}$ cells and 894 in PSA^{+} cells. Transition of chromatin status, i.e., loss of bivalent domains in PSA^{+} cells may indicate that PSA^{+} cells specifically activate and repress certain genes during differentiation from PSA$^{-/lo}$ cells (Figure 3-5A). To determine the effect of repressive and activating chromatin changes on transcription, transcript levels for gene promoters with transition of bivalent domains were evaluated. The mRNA levels of genes with bivalent domains on their TSS in PSA$^{-/lo}$ cells slightly increased in PSA^{+} cells with H3K4me3 only and slightly decreased if the mark transitioned to H3K27me3 only (Figure 3-5B). However, the comparisons are not statistically significant, which might be because the baseline expression levels of genes are widely variable or because the PSA$^{-/lo}$ subpopulation is yet heterogeneous (123).

Next, we made efforts to correlate preferential neurogenesis gene expression in PSA$^{-/lo}$ LNCaP cells with certain histone mark association. The most highly enriched neurogenesis transcripts in PSA$^{-/lo}$ cells include NRXN1, and its interacting protein,

NLGN1. NRXN1 belongs to neurexin cell-surface receptor family and NLGN1 belongs to neuroligin family. Neuroligins bind to neurexins to form $CTGF^{+}$-dependent neurexin/neuroligin complexes at synapses in the central nervous system, which are essential for neurotransmission (177). The neuroligins expressed on dendritic cells and neurexin expressed on axons trigger the formation of synaptic contacts and presynaptic and postsynaptic differentiation, respectively (177). In gastric cancer, patient tumors with *TP53* mutations co-occurring with mutations in *NRXN1* are predicted by the drug-gene association tool, Connectivity Map (CMAP), to have different drug response to chemotherapy and molecular and clinical profiles versus *TP53* mutation alone, suggesting a unique role for NRXN1 in cancer (178). Another neurogenesis gene highly expressed in $PSA^{-/lo}$ LNCaP cells is CHRM3, one of the muscarinic cholinergic receptors belonging to a larger family of G protein-coupled receptors normally expressed in smooth muscle cells, endocrine and exocrine glands, lungs, pancreas and brain (179). CHRM3 has been reported to preferentially couple with G_q, predominantly activating phospholipase C (PLC) and assisting in the mobilization of intracellular $CTGF^{+}$ stores subsequently activating protein kinase C (PKC) (179). CHRM3 has also been implicated in PKC-independent pathways, such as activation of PI3K and MAP kinase pathways, possibly underlying CHRM3-mediated cancer phenotypes such as cell proliferation and migration in a variety of cancers, including prostate (179, 180). Of importance, CHRM3 is upregulated in human CRPC cell lines, and overexpression of CHRM3 augments PC3-AR^{+} xenograft tumor growth in castrated hosts via AKT activation (180).

In contrast to the differentiation marker gene *KLK3* (encoding PSA), which displayed strong H3K4me3 enrichment, low levels of H3K27me3 and higher PSA expression in PSA^{+} cells compared to those in $PSA^{-/lo}$ cells (Figure 3-5C), NRXN1 and CHRM3, which were preferentially expressed in $PSA^{-/lo}$ cells, exhibited high H3K4me3 enrichment at the promoter particularly in $PSA^{-/lo}$ cells. The H3K27me3 mark was present on NRXN1 and CHRM3 promoters and gene bodies in both $PSA^{-/lo}$ and PSA^{+} cells (Figure 3-5D). Interestingly, although the NRXN1 mRNA level was significantly higher in $PSA^{-/lo}$ cells compared to PSA^{+} cells, its overall expression level was low (the intensity scale was only 0.41 and FPKM was 0.32). Nevertheless, NRXN1 transcription was greatly reduced in PSA^{+} cells (FPKM was 0.0048). Taken together, these results imply that bivalent domains on NRXN1 and CHRM3 in undifferentiated $PSA^{-/lo}$ cells might poise the genes for low expression in $PSA^{-/lo}$ cells that is further suppressed in PSA^{+} cells, with removal of the activating H3K4me3 mark (Figure 3-5D). Overall, $PSA^{-/lo}$ cells possess more genes with bivalent domains on promoter regions and low levels of expression, and the majority of these genes change their chromatin status to activated or suppressed states in more differentiated PSA^{+} cells.

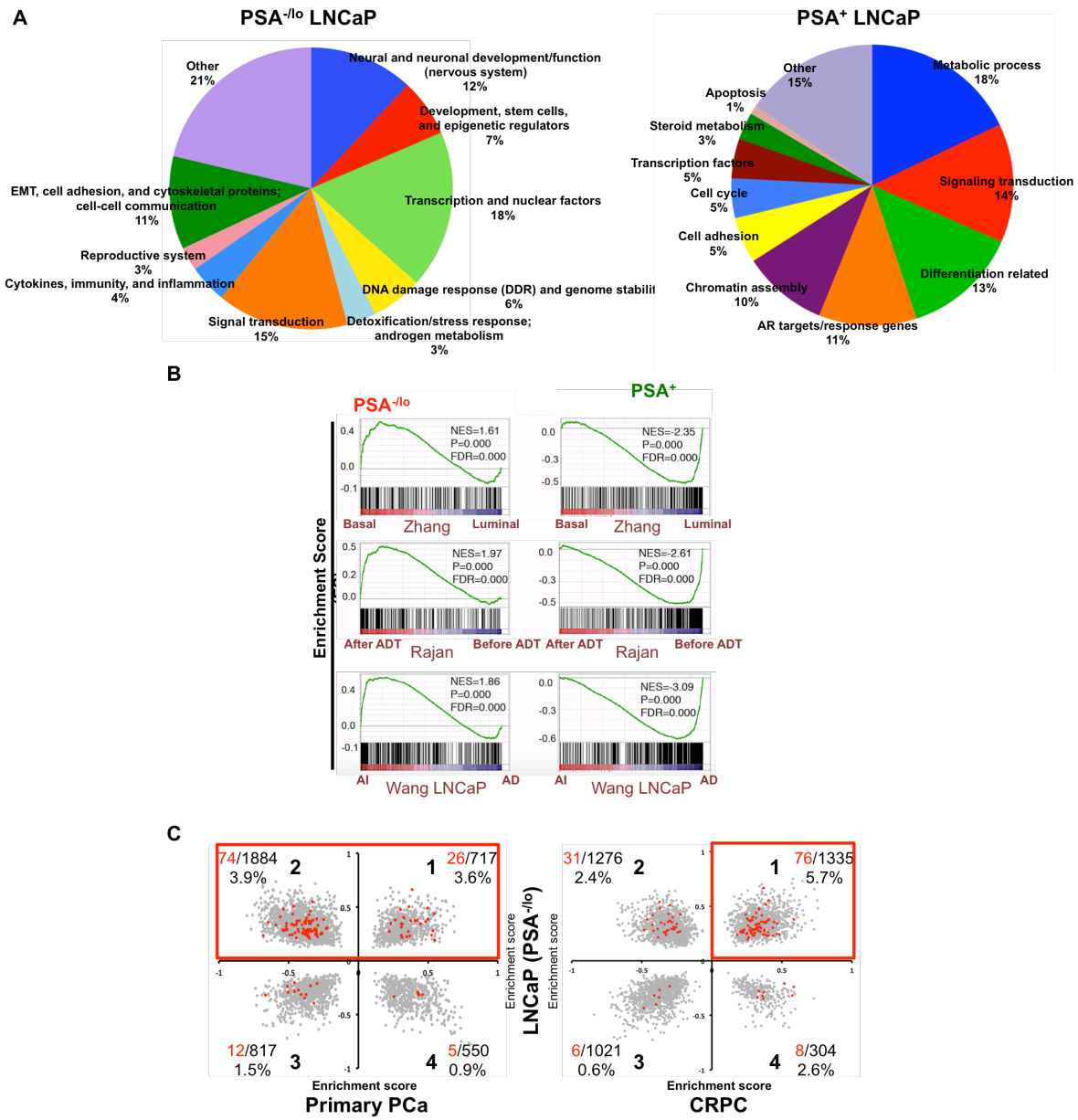

A
PSA-/lo LNCaP
Other 21%
Neural and neuronal development/function (nervous system) 12%
Development, stem cells, and epigenetic regulators 7%
Transcription and nuclear factors 18%
DNA damage response (DDR) and genome stabilit 6%
Detoxification/stress response; androgen metabolism 3%
Signal transduction 15%
Cytokines, immunity, and inflammation 4%
Reproductive system 3%
EMT, cell adhesion, and cytoskeletal proteins; cell-cell communication 11%
PSA+ LNCaP
Other 15%
Apoptosis 1%
Steroid metabolism 3%
Transcription factors 5%
Cell cycle 5%
Cell adhesion 5%
Chromatin assembly 10%
AR targets/response genes 11%
Differentiation related 13%
Signaling transduction 14%
Metabolic process 18%
B
PSA-/lo
PSA+
Enrichment Score (ES)
NES=1.61 P=0.000 FDR=0.000
NES=-2.35 P=0.000 FDR=0.000
Basal Zhang Luminal
NES=1.97 P=0.000 FDR=0.000
NES=-2.61 P=0.000 FDR=0.000
After ADT Rajan Before ADT
NES=1.86 P=0.000 FDR=0.000
NES=-3.09 P=0.000 FDR=0.000
AI Wang LNCaP AD
C
74/1884 3.9%
26/717 3.6%
12/817 1.5%
5/550 0.9%
Enrichment score
Primary PCa
LNCaP (PSA-/lo)
31/1276 2.4%
76/1335 5.7%
6/1021 0.6%
8/304 2.6%
CRPC

Figure 3-2. Functional annotation of the transcriptome profiles in $PSA^{-/lo}$ and PSA^{+} LNCaP cell subpopulations.

A. $PSA^{-/lo}$ and PSA^{+} LNCaP cells express functionally distinct genes. Non-redundant gene annotations of DEGs from $PSA^{-/lo}$ and PSA^{+} cells. Each DEG (fold change ≥ 2) was manually curated into a single functional gene category using the in-house non-redundant GO analysis tools.

B. The transcriptome profile of $PSA^{-/lo}$ cells tends to be positively associated with transcriptome profiles of normal human prostate basal cells (Zhang; (112)), patient CRPC post ADT (Rajan; (175)), and castration-selected LNCaP AI cells (Wang; (181)), whereas the transcriptome of PSA^{+} cells displays the reciprocal pattern and is associated with transcriptome profiles of normal differentiated ($AR^{+}PSA^{+}$) prostate luminal cells, AD patient tumors, and AD LNCaP cells, respectively. The DEGs of $PSA^{-/lo}$ and PSA^{+} cells were compared to transcriptome profiles of various known biological phenotypes by GSEA.

C. Global gene annotation of $PSA^{-/lo}$ cell transcriptome with enrichment of neural and neuronal development/function more similar to transcript profiles from patient CRPC (Rajan dataset) than those of primary tumors (TCGA dataset). Analysis was performed with all the GO gene sets in GSEA and NES (normalized enrichment scores) are displayed in scatter plots. The gray dots represent all the GO terms in GO gene sets and red dots the neurogenesis genes. The transcriptome profiles in $PSA^{-/lo}$ cells and patient CRPC, but not primary PCa, show positive enrichment of a neurogenesis gene signature.

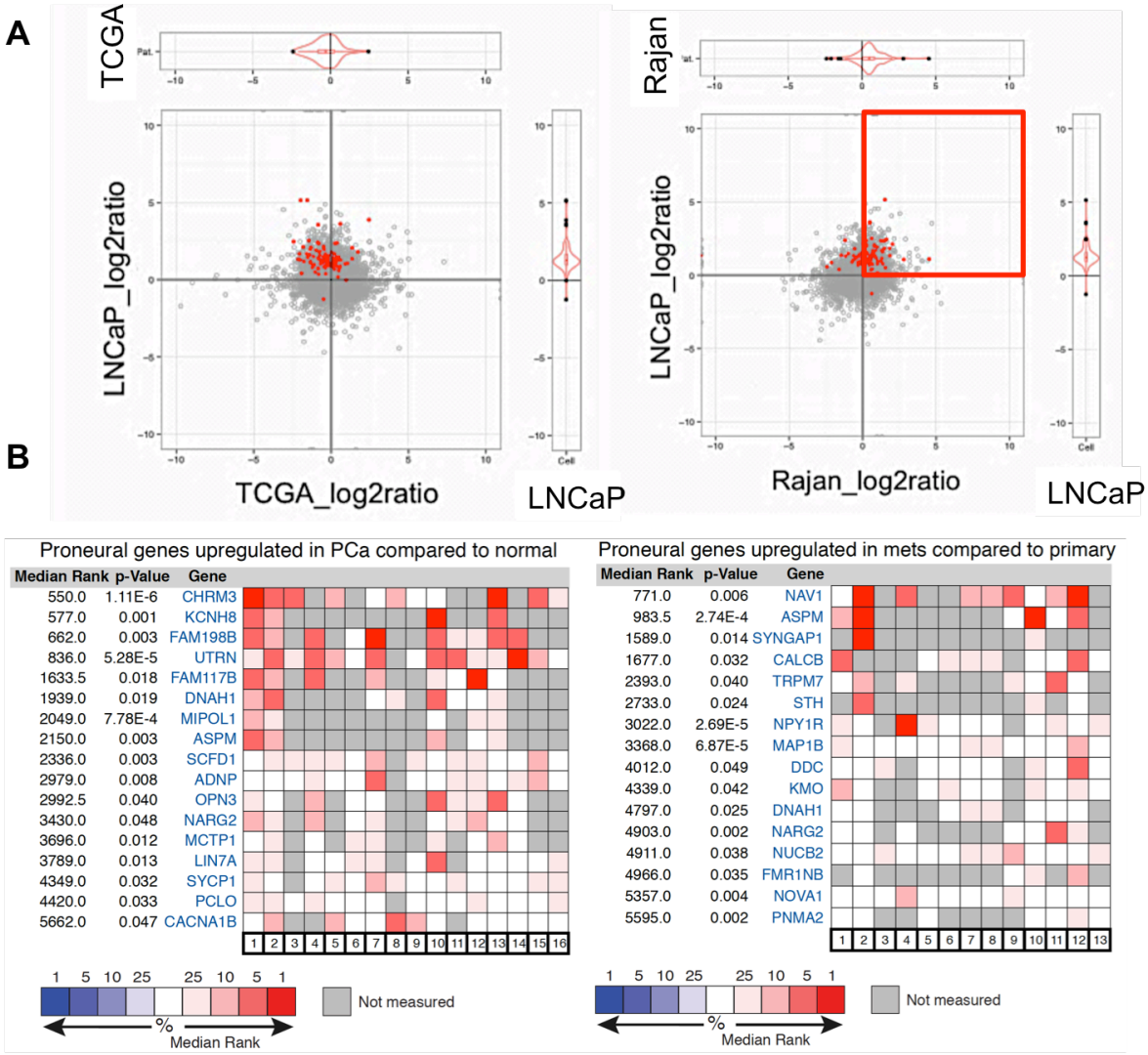

A
TCGA
LNCaP_log2ratio
TCGA_log2ratio
LNCaP
Rajan
LNCaP_log2ratio
Rajan_log2ratio
LNCaP
B
Proneural genes upregulated in PCa compared to normal
Median Rank p-Value Gene
550.0 1.11E-6 CHRM3
577.0 0.001 KCNH8
662.0 0.003 FAM198B
836.0 5.28E-5 UTRN
1633.5 0.018 FAM117B
1939.0 0.019 DNAH1
2049.0 7.78E-4 MIPOL1
2150.0 0.003 ASPM
2336.0 0.003 SCFD1
2979.0 0.008 ADNP
2992.5 0.040 OPN3
3430.0 0.048 NARG2
3696.0 0.012 MCTP1
3789.0 0.013 LIN7A
4349.0 0.032 SYCP1
4420.0 0.033 PCLO
5662.0 0.047 CACNA1B
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
1 5 10 25 25 10 5 1
%
Median Rank
Not measured
Proneural genes upregulated in mets compared to primary
Median Rank p-Value Gene
771.0 0.006 NAV1
983.5 2.74E-4 ASPM
1589.0 0.014 SYNGAP1
1677.0 0.032 CALCB
2393.0 0.040 TRPM7
2733.0 0.024 STH
3022.0 2.69E-5 NPY1R
3368.0 6.87E-5 MAP1B
4012.0 0.049 DDC
4339.0 0.042 KMO
4797.0 0.025 DNAH1
4903.0 0.002 NARG2
4911.0 0.038 NUCB2
4966.0 0.035 FMR1NB
5357.0 0.004 NOVA1
5595.0 0.002 PNMA2
1 2 3 4 5 6 7 8 9 10 11 12 13
1 5 10 25 25 10 5 1
%
Median Rank
Not measured
C
NRXN1 (Grasso)
Low (20)
High (15)
P=0.021
CHRM3 (Grasso)
Low (85)
High (37)
P=0.002
FAM198B (Glinsky)
P=0.003
Low (24)
High (55)
UTRN #1 (Glinsky)
P=0.0032
Low (22)
High (57)
ASPM (Grasso)
Low (86)
High (36)
P=0.023
NAV1 (Grasso)
P=0.013
Low (59)
High (63)
SYNGAP1 (Grasso)
P=0.0019
Low (12)
High (23)
NPY1R (Setlur)
Low (257)
High (106)
P=0.009
Overall survival
Time (years)

Figure 3-3. Clinical relevance of neurogenesis genes enriched in the PSA$^{-/lo}$ subpopulation.

A. Global transcriptome plot shows that neurogenesis (i.e., neural and neuronal development/function) genes are enriched in PSA$^{-/lo}$ LNCaP cells and patient CRPC relative to primary (AD) tumors. The gray dots represent individual mRNAs from the transcriptome profiles of PSA$^{-/lo}$ cells and primary patient tumors in TCGA (left panel) or PSA$^{-/lo}$ cells and patient CRPC (right panel) with neuronal development/function associated genes highlighted in red.

B. The Oncomine concept analysis was performed with cutoff at median rank < 6,000 and p-value < 0.05. The results show the upregulated neurogenesis genes when comparing primary tumors versus normal tissue (left) and when comparing metastases over primary tumors (right).

C. The expression of individual neurogenesis genes correlates with poor patient survival. The indicated neurogenesis genes upregulated in PSA$^{-/lo}$ LNCaP cells were used to stratify patients into high and low expressers; patient survival information was obtained from Oncomine datasets. Data presented were statistically significant with p-value cutoff < 0.05.

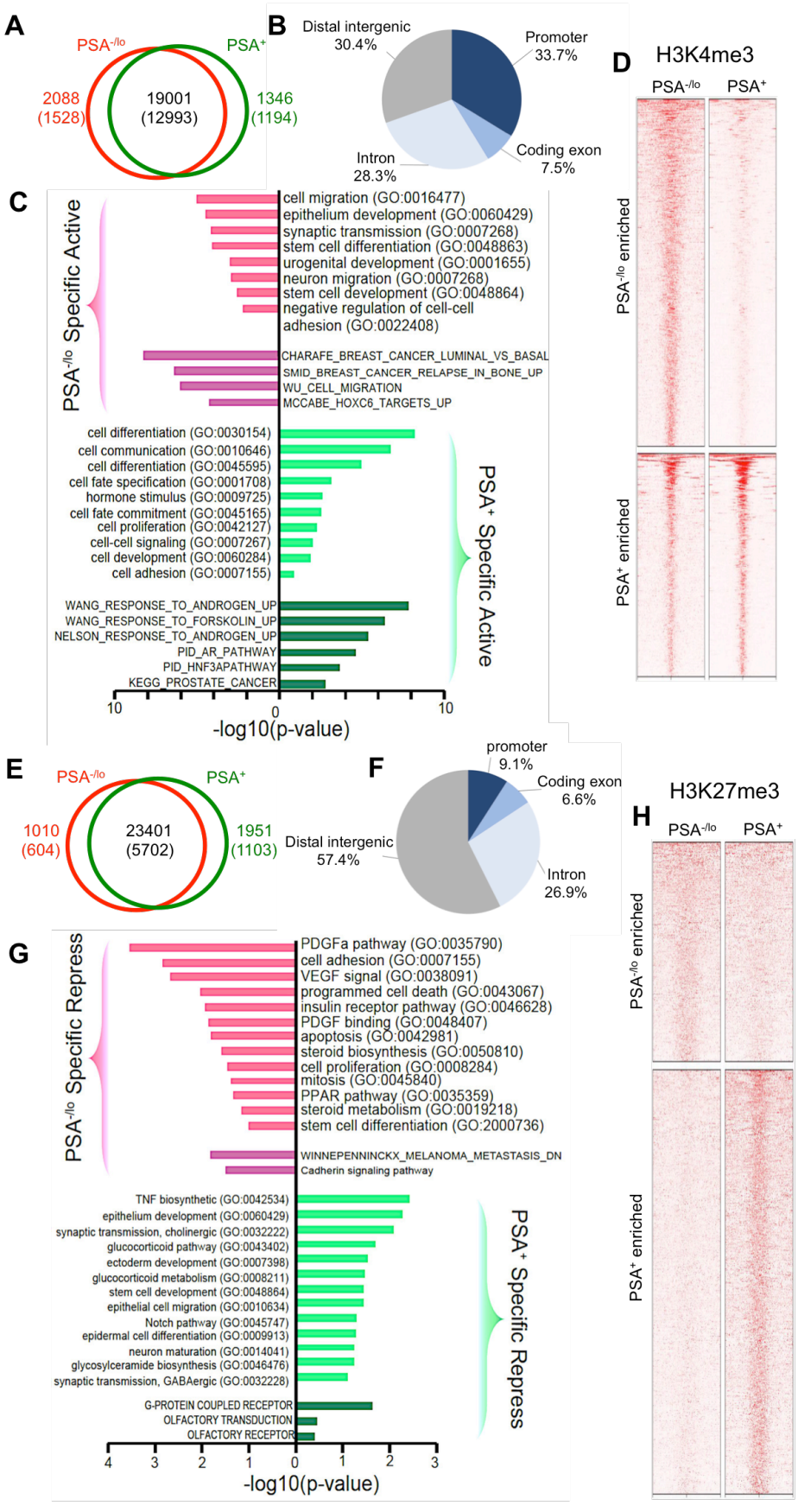

A
PSA-/lo
PSA+
2088
(1528)
19001
(12993)
1346
(1194)
B
Distal intergenic
30.4%
Promoter
33.7%
Intron
28.3%
Coding exon
7.5%
D
H3K4me3
PSA-/lo
PSA+
PSA-/lo enriched
PSA+ enriched
C
PSA-/lo Specific Active
cell migration (GO:0016477)
epithelium development (GO:0060429)
synaptic transmission (GO:0007268)
stem cell differentiation (GO:0048863)
urogenital development (GO:0001655)
neuron migration (GO:0007268)
stem cell development (GO:0048864)
negative regulation of cell-cell adhesion (GO:0022408)
CHARAFE_BREAST_CANCER_LUMINAL_VS_BASAL
SMID_BREAST_CANCER_RELAPSE_IN_BONE_UP
WU_CELL_MIGRATION
MCCABE_HOXC6_TARGETS_UP
cell differentiation (GO:0030154)
cell communication (GO:0010646)
cell differentiation (GO:0045595)
cell fate specification (GO:0001708)
hormone stimulus (GO:0009725)
cell fate commitment (GO:0045165)
cell proliferation (GO:0042127)
cell-cell signaling (GO:0007267)
cell development (GO:0060284)
cell adhesion (GO:0007155)
WANG_RESPONSE_TO_ANDROGEN_UP
WANG_RESPONSE_TO_FORSKOLIN_UP
NELSON_RESPONSE_TO_ANDROGEN_UP
PID_AR_PATHWAY
PID_HNF3APATHWAY
KEGG_PROSTATE_CANCER
PSA+ Specific Active
10
0
10
-log10(p-value)
E
PSA-/lo
PSA+
1010
(604)
23401
(5702)
1951
(1103)
F
promoter
9.1%
Coding exon
6.6%
Distal intergenic
57.4%
Intron
26.9%
H
H3K27me3
PSA-/lo
PSA+
PSA-/lo enriched
PSA+ enriched
G
PSA-/lo Specific Repress
PDGFa pathway (GO:0035790)
cell adhesion (GO:0007155)
VEGF signal (GO:0038091)
programmed cell death (GO:0043067)
insulin receptor pathway (GO:0046628)
PDGF binding (GO:0048407)
apoptosis (GO:0042981)
steroid biosynthesis (GO:0050810)
cell proliferation (GO:0008284)
mitosis (GO:0045840)
PPAR pathway (GO:0035359)
steroid metabolism (GO:0019218)
stem cell differentiation (GO:2000736)
WINNEPENNINCKX_MELANOMA_METASTASIS_DN
Cadherin signaling pathway
TNF biosynthetic (GO:0042534)
epithelium development (GO:0060429)
synaptic transmission, cholinergic (GO:0032222)
glucocorticoid pathway (GO:0043402)
ectoderm development (GO:0007398)
glucocorticoid metabolism (GO:0008211)
stem cell development (GO:0048864)
epithelial cell migration (GO:0010634)
Notch pathway (GO:0045747)
epidermal cell differentiation (GO:0009913)
neuron maturation (GO:0014041)
glycosylceramide biosynthesis (GO:0046476)
synaptic transmission, GABAergic (GO:0032228)
G-PROTEIN COUPLED RECEPTOR
OLFACTORY TRANSDUCTION
OLFACTORY RECEPTOR
PSA+ Specific Repress
4
3
2
1
0
1
2
3
-log10(p-value)

Figure 3-4. Functional annotation of histone binding profiles in $PSA^{-/lo}$ and PSA^{+} LNCaP cell subpopulations.

A-D. GO annotation by GREAT (Gene Region Enrichment Annotation Tool) of the genes marked by H3K4me3 on promoters in both LNCaP cell subpopulations. A. The Venn diagram presents the total numbers of peaks called and the degree of overlap between $PSA^{-/lo}$ and PSA^{+} LNCaP cells. There are 19,001 H3K4me3 peaks (12,993 genes) commonly detected in both cell populations, 2,088 peaks (1,528 genes) and 1,346 peaks (1,194 genes) preferentially appeared in $PSA^{-/lo}$ and PSA^{+} cells, respectively. B. Genomic distribution of H3K4me3 occupancy. Promoter region is defined as ±3 Kb from the TSS and distal intergenic region includes < -3Kb from TSS and outside the transcription end site (TES). C. GO analyses, performed by GREAT with GO terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) signaling pathways, revealed distinct categories of gene enrichment in the two cell subpopulations. D. Heatmaps of the H3K4me3 peaks (±5 Kb) preferentially enriched in $PSA^{-/lo}$ cells (top) or in PSA^{+} cells (bottom).

E-H. GO annotation by GREAT of the genes marked by H3K27me3 on promoters in both LNCaP cell subpopulations. E. In total, there are 23,401 peaks associated with 5,702 genes in both cell populations, and 1,010 peaks (604 genes) and 1,951 peaks (1103 genes) called in $PSA^{-/lo}$ and PSA^{+} cells, respectively. F. Genomic distribution of H3K4me3 occupancy. G. Differential H3K27me3 marked genes reveal divergent biological functions and phenotypes in GO analysis. H. Heatmaps showing the preferentially enriched H3K27me3 peaks in the two cell subpopulations.

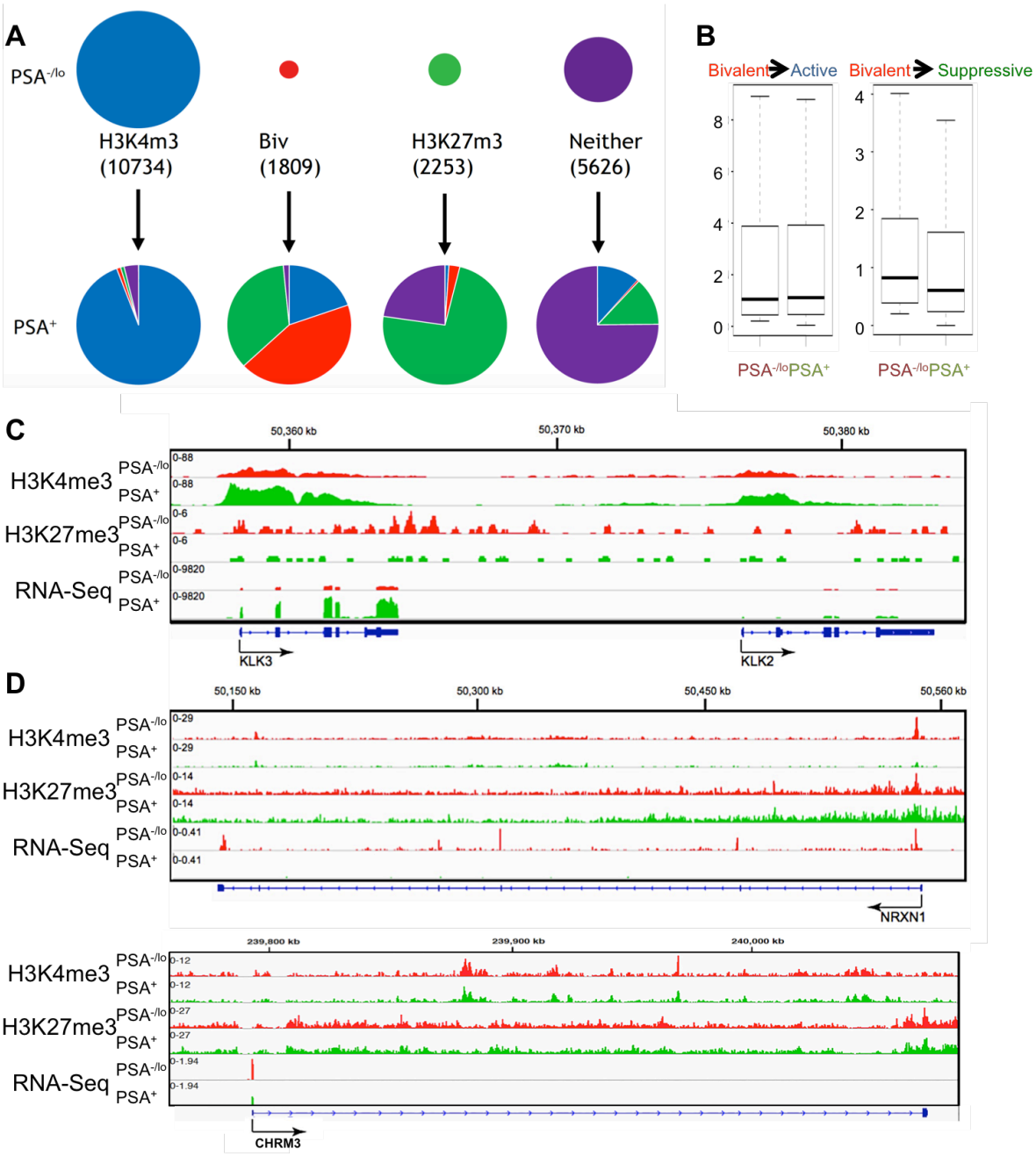


Figure 3-5. Chromatin state transitions in the two LNCaP cell subpopulations and their associated gene expression profiles.

A. Histone modification shift in comparison of peak type frequency between PSA$^{-/lo}$ and PSA^{+} cells. The sizes of the top circles represent the numbers of peak regions with H3K4me3, H3K27me3, both, or neither on promoter regions (± 1Kb from TSS) in PSA$^{-/lo}$

cells. The lower circles are the pie charts indicating the percentages of promoter regions that have changed status in PSA^+ cells. Blue indicates active (H3K4me3), red bivalent (both H3K4me3 and H3K27me3), green suppressive (H3K27me3), and purple neither mark.

B. Gene expression corresponding to chromatin status transition derived from profiling $PSA^{-/lo}$ cells versus PSA^+ cells. All genes with bivalently marked TSS in $PSA^{-/lo}$ cells with FPKM > 0.2 were separated into two groups according to a change to activating (H3K4me3) or suppressive (H3K27me3) histone status on promoters in PSA^+ cells. Box plots show average transcript abundance in $PSA^{-/lo}$ and PSA^+ cells for genes with the indicated chromatin mark transition.

C-D. Histone and mRNA landscapes of the indicated representative genes. Shown are UCSC tracks from H3K4me3, H3K27me3, and RNA-Seq landscape. Green peaks represent tracks of PSA^+ cells, and red peaks the tracks of $PSA^{-/lo}$ cells.

3.5 Discussion

In general, cultured cancer cell lines are considered to be relatively homogeneous in comparison to highly heterogeneous patient tumors. Nevertheless, distinct LNCaP subsets can be distinguished on the basis of the expression of the differentiation marker PSA. $PSA^{-/lo}$ and PSA^{+} subpopulations co-exist in the untreated cultures, and their gene expression profiles reveal molecular distinctions reflecting their biological features (Figure 3-2A). The bulk cells display a spectrum of intermediated cells with various PSA expression levels. Here, I focus on comparing the intrinsic differences of transcriptome differences and histone modifications between lineage-related cell subpopulations with top 6-10% (PSA^{+}) and lowest 2-3% ($PSA^{-/lo}$) PSA expression. This strategy aims to simply our understanding of the PCa cells at the two extreme spectra – i.e., the most differentiated ($PSA^{+/hi}$) and the least differentiated ($PSA^{-/lo}$) without considering the continuum in between. It should be noted that the current study is conducted in a single PCa cell system, LNCaP. However, the $PSA^{-/lo}$ cell population is highly clinical relevant. Firstly, $PSA^{-/lo}$ cells pre-exist in untreated patient tumors and become enriched in advanced tumors and the predominant cell population in CRPC (123). Secondly, higher PSA mRNA level is associated with better overall patient survival. Thirdly, biological relevance of the $PSA^{-/lo}$ population has been corroborated in multiple cancer models, including the LAPC4, LAPC9, and LNCaP as well as patient derived $PSA^{-/lo}$ cells (124). Therefore, although my current study focused on one cell model, I believe that data presented here and conclusions drawn should be applicable to PCa in the clinic.

Analysis of unperturbed $PSA^{-/lo}$ cell transcriptomes uncovers unique SC, development, epigenetic regulator, extracellular matrix, and cell-cell communication enriched gene-expression profiles. In contrast, the PSA^{+} cell transcriptome is enriched in genes related to metabolic processes, signal transduction, differentiation, AR target/responsive, and cell cycle. Also, the gene expression profile of $PSA^{-/lo}$ cells is highly associated with the transcriptome of stem-like normal prostate basal cells and treatment-failed CRPC (Figure 3-2B). These results are consistent with the biological phenotypes discovered previously in which $PSA^{-/lo}$ cells are found to be undifferentiated, stem-like cells, whereas PSA^{+} cells were found to be more differentiated and functional prostate cell types (25). These DEG profiles further support the unique capabilities of $PSA^{-/lo}$ PCa cells to survive androgen deprivation and drive CRPC progression and implicate these cells as potential therapeutic targets in the clinic.

The bivalent domains on the promoters of development genes poise them for timely responses to differentiation signals (176). Promoter bivalency also regulates cellular plasticity and phenotypic flexibility via modulating expression of genes involved in EMT signaling pathways, such as WNT and TGFβ (182). $PSA^{-/lo}$ cells appear to have more genes associated with bivalent domains along with upreguation of some epigenetic modifiers (e.g., ARID2, JMJDIC, KDM5A, KDM6A, KDM6B, SETD2, SETD5, TET1, and TET3), which might suggest that not only bivalency but also other epigenetic mechanisms may be in operation to sustain CSC properties of the $PSA^{-/lo}$ cells. Also, genes important to stemness and EMT are upregulated in $PSA^{-/lo}$ DEG profiles (e.g., WNT5A and TGFβ, MMP7) (Figure 3-2A). Together, the results

suggest that lineage-related $PSA^{-/lo}$ and PSA^{+} cells possess distinct epigenetic landscapes, which are crucial for cell identities.

$PSA^{-/lo}$ cells are enriched in genes normally associated with neural and neuronal development/function. Intriguingly, such neurogenesis genes are also preferentially expressed in $AR^{-}PSA^{-}$ normal prostate basal/stem cells and have been linked to cell plasticity and adverse clinical features of PCa (Figure 3-2B) (112). Detailed examination of the $PSA^{-/lo}$ transcriptome profile shows that at least some of these genes, like NRXN1 and CHRM3, are poised with bivalent marks with a loss of H3K4me3 at the TSS in more differentiated PSA^{+} cells (Figure 3-5D). Therefore, we reason that NRXN1 and CHRM3 may function as stem cell factors and regulate CSC properties of $PSA^{-/lo}$ cells. In support of this hypothesis, our pilot functional studies show that siRNA- or shRNA-mediated knockdown of NRXN1 expression or CHRM3 reduced the colony- and sphere-forming properties of $PSA^{-/lo}$ or bulk LNCaP cells *in vitro* (Figure 3-6A-C; data not shown). Importantly, knocking down NRXN1 in LAPC9 xenografts using shRNA lentiviral vectors inhibited tumorigenesis, as indicated by reduced tumor incidence and tumor weight upon transplantation of limiting cell numbers into immunocompromised hosts (Figure 3-6D). Interestingly, suppressing CHRM3 expression appeared to also inhibit $PSA^{-/lo}$ cell invasion and migration abilities (Figure 3-6E).

Of clinical relevance, CHRM3 has been implicated in CRPC (180) and the neurogenesis gene signature is highly enriched in patient CRPC transcriptome profiles (Figure 3-2C and 3-3A). In further support, we have observed that expression of NRXN1 and CHRM3 is upregulated in not only $PSA^{-/lo}$ LNCaP cells but also LNCaP

cells selected after chronic castration in vitro (Figure 3-6F). Moreover, expression of these genes adversely correlates with patient survival (Figure 3-3C). Although it remains to be determined whether the entire cohort of neurogenesis genes or a 'neurogenesis gene signature' could predict patient survival, our data has linked a few highly expressed neurogenesis genes, individually, to the poor patient survival (Figure 3-3C). In the re-analyzed Rajan dataset comparing patient samples before and after ADT, NRXN1 mRNA expression showed around 2.9 times higher with p-value < 0.012 after ADT (compared to before ADT), but, surprisingly, CHRM3 was actually downregulated 4.3 times in patient tumors after ADT. This latter observation is inconsistent with an earlier report (180) and the reason underlying this inconsistency remains unclear.

In summary, our data has shown that the expression levels of many individual neurogenesis genes are adversely associated with patient survival. Concept analysis from Oncomine also reveals that many genes are highly upregulated in prostate tumors compared to normal and in metastases compared to primary tumors. Moreover, previous studies have demonstrated that the neurogenesis gene profiles expressed in the normal human prostate basal/stem cells have been linked to aggressive PCa (112, 183). All such evidence suggests the clinical relevance of neurogenesis genes on the whole. Of interest, the 'default differentiation path in normal stem cells, when cultured under inappropriate conditions, is the neural differentiation. This indicates that the neural differentiation program is built in the intrinsic stem cell-associated differentiation program (112, 184). Also, neural stem and

progenitor cells are well known to be very ‘aggressive’ and are highly migratory and invasive undergoing EMT during development (112, 183).

To conclude, the CSC-enriched $PSA^{-/lo}$ PCa cell subpopulation exhibits intrinsic transcriptomic and epigenomic differences compared with lineage-related PSA^{+} differentiated PCa cells. The $PSA^{-/lo}$ CSC subset possesses SC properties discernible in both its transcriptome profiles (this study) and biological functions (25), and may serve as a cell-of-origin for CRPC (124). It is tempting to speculate that the neurogenesis genes enriched in $PSA^{-/lo}$ cells may be causally related to NEPC (neuroendocrine PCa) phenotype although the $PSA^{-/lo}$ cells purified from multiple PCa models and patient tumors lack the expression of mature NEPC markers such as chromogrannin A and synaptophysin (25, 123). Regardless, these interrogations demonstrate the molecular divergence of prostate CSCs and non-CSCs, potentially shedding light on CSC-targeting therapeutics to eradicate PCa, especially CRPC. Considering that the $PSA^{-/lo}$ cells pre-exist in the primary tumors (25, 123), combinatorial therapies targeting both CSCs (with novel CSC-directed therapies) and non-CSCs (e.g., using enzalutamide to target differentiated $AR^{+}PSA^{+}$ PCa cells) for early-stage PCa patients may prove to be valuable in the clinic.

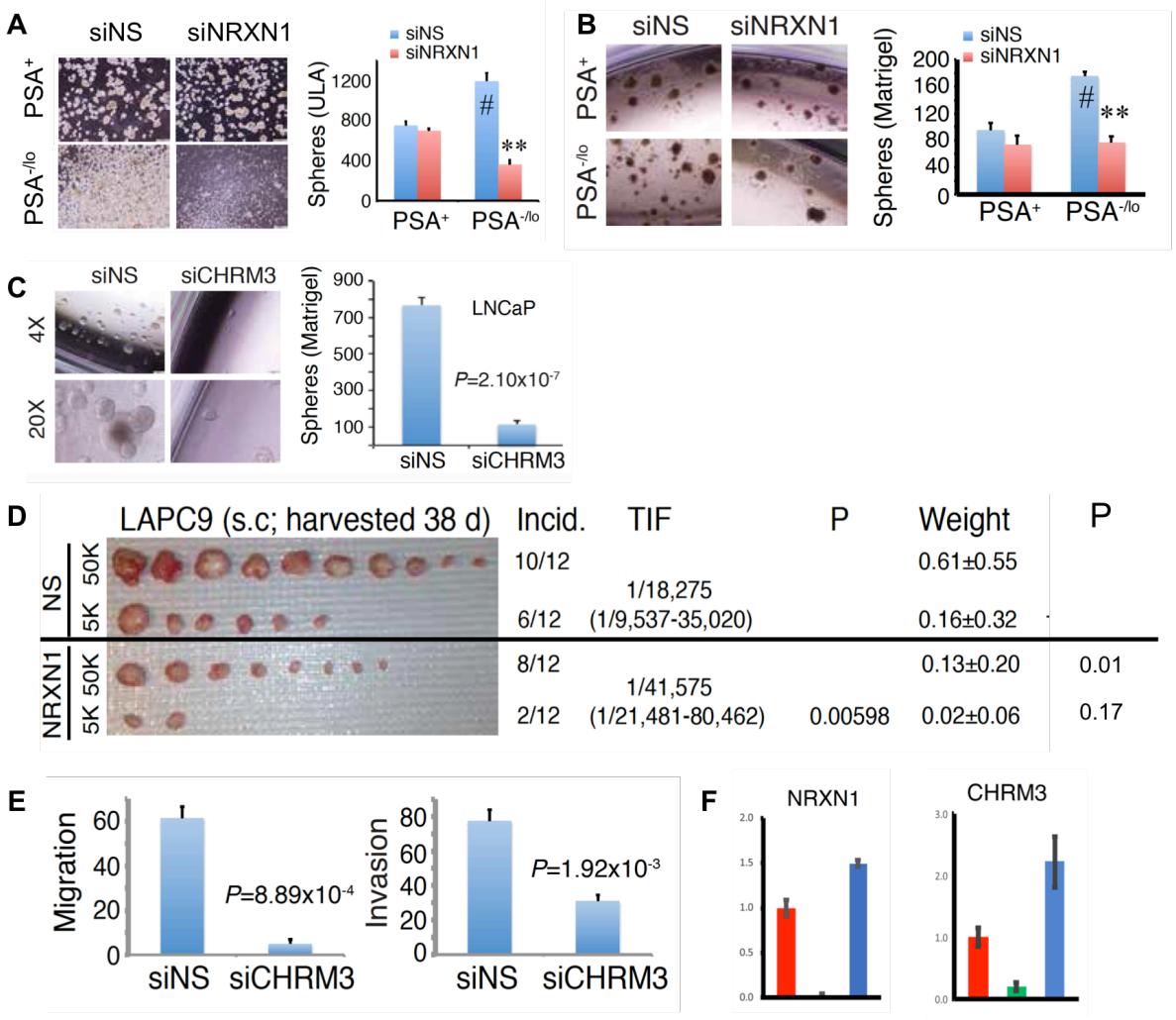


Figure 3-6. Characterization of NRXN1 and CHRM3 biological functions (performed by Dr. Xin Liu and Dr. Kiera Rycaj)

A-C. Knockdown down the expression of NRXN1 and CHRM3 inhibits the CSC properties of $PSA^{-/lo}$ cells. (A) The results display images of clongenic assay of siRNA repressing NRXN1 (versus non-silencing control (NS)) in both $PSA^{-/lo}$ cells and PSA^{+} cells. (B) The results are sphere formation assay and its quantification data of $PSA^{-/lo}$ cells and PSA^{+} cells with siRNA of NRXN1 or NS. (C) The results show sphere formation assay upon knocking down CHRM3 in $PSA^{-/lo}$ cells. # indicated p-value < 0.05 and ** is p-value < 0.01

D. Knockdown of NRXN1 inhibits tumorigenesis. The tumor incidence and tumor weights were measured at 38 days after xenografting 5K or 50K LAPC9 cells transduced with nonsilencing control shRNA or NRXN1 shRNA prior to subcutaneous transplantation into immunocompromised hosts.

E. Knocking down CHRM3 inhibits $PSA^{-/lo}$ cells migration and invasion. Transwell assays were performed without or with matrigel for migration and invasion assays, respectively. Crystal violet stained cells on the bottom side of well were counted.

F. Expression of NRXN1 and CHRM3 in $PSA^{-/lo}$, PSA^{+}, and long-term castrated LNCaP cells. Quantitative RT-PCR was applied to detect the mRNA expression in $PSA^{-/lo}$, PSA^{+}, and MDV3100 treated LNCaP cells. LNCaP cells were treated in MDV3100 containing medium longer than a year to select the long-term castration resistant LNCaP cells.

CHAPTER FOUR

Bioinformatics analyses reveal unexpected molecular mechanisms underlying NANOG-induced PCa cell reprogramming

4.1 Introduction

Self-renewal is a critical biological property for normal adult SCs, embryonic stem cells (ESCs), induced pluripotent stem cell (iPSCs), and CSCs. Nanog is a transcription factor involved in the self-renewal of ESCs by working coordinately with other stemness genes, including octamer-binding transcription factor 4 (OCT4) and the sex-determining region Y HMG-box2 (SOX2) (77, 185). The main role of Nanog is to maintain the ESCs in a stable undifferentiated state and establish the pluripotent ground state in the inner cell mass. NANOG is expressed in epiblast ES cells and undetectable in most of the somatic cells in normal adult organs, except in germ cells. Nanog, collaborating with other pluripotency transcription factors, is also critical in iPSCs reprogramming to facilitate adult somatic cells to regain the SC properties (186). The primary function of Nanog on iPSC reprogramming is transition of pre-iPSCs to fully induced ground state iPSCs but not during early stages of iPSC generation (187). Mechanistically, Nanog directly interacts with methylcytosine hydroxylase TET1/2, which globally demethylates 5'-mC to generate 5'-hydroxymethylcytosine (5hmC), resulting iPSC generation (188).

The NANOG expressed in ESCs, i.e., NANOG1, is encoded by the gene located on chromosome 12p13 (189). There are many retrotransposed NANOG

'homologs' in the human genome (189). These retrogene variants (pseudogenes) of NANOG, i.e., NANOGP2 to NANOGP11, are encoded by genes located on chromosome 12, 2, 6, 7, 9, 10, 14, 15, X, and X, respectively. NANOGP5 shows 85% homology to NANOG, whereas all other variants have more than 90% similarity. The human NANOGP8 (NP8) can encode full-length NANOG protein of 305 amino acid, whereas NANOGP2, P4, P5, P9, P10 are presumed to generate truncated protein products due to the presence of premature stop codons in the coding sequence (190, 191). NP8 is highly similar to NANOG with five nucleotide substitutions, only 1 of which, nucleotide 795, encodes a putative amino acid change at Q253H (87).

Although NANOG is not expressed in the majority of normal somatic cells, NANOG expression has been reported in many cancer types, including breast (192), bladder (193), ovary (194), and prostate (195) cancers. Interestingly, NP8, instead of NANOG1, has been reported to be the predominant NANOG 'isoform' expressed in somatic cancer cells, although in some cancers NANOG mRNA might potentially arise from the *NANOG1* locus (196). Functionally, NANOG in cancer cells has been implicated in regulation of migration and metastasis (197-200), apoptosis and cell cycle (201-204), angiogenesis (205, 206), multidrug resistance (89, 207), and CSC self-renewal (208).

Work from our laboratory has demonstrated that NP8 mRNA and NANOG protein are expressed in PCa (87, 93). Interestingly, NANOG positivity is negatively correlated with AR expression in patient PCa cells in vivo and enriched in $CD44^{hi}$ PCa stem/progenitor cells, which are frequently $AR^{-/lo}$ (87). Knocking down NANOG inhibits tumor development in xenograft models, and suppresses clonal and clonogenic

growth of PCa cell lines and primary human PCa cells *in vitro* (87). On the other hand, overexpressing NANOG and NP8 increases expression of CSC-associated molecules, CD133 and CD44, and decreases differentiation molecules, AR and PSA (93). These results indicate that NANOG promotes CSC phenotypes and NANOG-expressing cells might serve as cell-of-origin of these cancers. Furthermore, NANOG and NP8 overexpression facilitates the development of ADT resistance by stimulating survivability of LNCaP cells in experimental androgen deprivation conditions and, importantly, enhance the clonogenic ability and tumor regeneration of LNCaP cells in androgen-deprived host (93). Inducible overexpression of NANOG promotes acquisition of CSC properties and resistance to castration in PCa cells (93), and knocking down NANOG inhibits tumor regeneration of the $PSA^{-/lo}$ LAPC9 PCa cells in castrated host (25). In this experimental setting, NANOG promotes the expression of pro-survival, pro-proliferation, anti-apoptosis, pro-migration, and detoxification genes, including ABCG2, CD133, IGFBP-5, BCL-2, CXCR4, and ALDH1A1 (93). These functional studies (87) (93) suggest that the endogenously expressed 'master' transcription factor, NANOG, seems to be critical for tumorigenic properties of (prostate) cancer cells. More importantly, our studies with inducible NANOG in PCa cells (93) suggest that overexpression of NANOG alone appears to be sufficient to reprogram androgen-sensitive (androgen-dependent, AD) PCa cells to the androgen-independent (AI; castration-resistant) state that also acquires a variety of phenotypic and functional CSC properties (93). In this project, we made efforts to define the mechanisms underlying the NANOG-induced reprogramming of AD LNCaP cells to the AI LNCaP cells manifesting castration resistance and CSC characteristics.

4.2 Hypothesis

To a degree, the process of NANOG-induced reprogramming of LNCaP cells, within 2-3 weeks (93), into the castration-resistant state that has also acquired certain CSC properties resembles somatic cell reprogramming by OSKM factors (Oct4, Sox2, KLF4, and Myc) to generate the iPSCs. Therefore, our initial hypothesis is that NANOG may reprogram PCa cells via reactivating the pluripotency network to engage endogenous stemness factors such as Oct4 and SOX2 as in iPSC reprogramming. On the other hand, we are mindful that LNCaP cells, unlike normal 'differentiated' somatic cells such as fibroblasts, are metastatic cancer cells that already possess some stem cell properties such as longevity (at the population level). Therefore, we alternatively hypothesize that NANOG-induced LNCaP cell reprogramming may not necessarily involve reactivation of the pluripotency network and pluripotency factors but instead may engage PCa-specific lineage regulators. In this project, we test these hypotheses by performing NANOG ChIP-Seq experiments combined with time-resolved RNA-Seq analyses and some biological validations. I helped annotate and interpret the newly generated ChIP-Seq and RNA-Seq data sets, and the results tested our hypotheses and led to some new testable hypotheses.

4.3 Materials and Methods

Biological models

Inducible NANOG expression in LNCaP cells was established using a binary Tet-on system, in which the CMV promoter drives TetR expression and TRE drives the expression of NANOG1 or NP8 (Figure 4-1; (93)). In principle, the transgene

(NANOG1 or NP8) is turned on by addition of doxycycline (DOX) in the culture medium for various time intervals. LNCaP-NANOG1 or LNCaP-NP8 cells were cultured in normal serum-containing medium, i.e., AD condition, or in medium containing charcoal dextran-stripped serum (CDSS), i.e., the AI condition. As illustrated in Figure 4-1, samples were harvested at different time points for analysis. For RNA-Seq experiments, the LNCaP-NP8 cells cultured in the presence of DOX were collected at both early (i.e., 5 days in AD conditions and 7 days in AI conditions) and late (i.e., 12 days in AD conditions and 22 days in AI conditions) time points. For RNA-Seq in LNCaP-NANOG1 cells, we only used the DOX-induced cells for 5 days in the AD conditions. For ChIP-Seq experiments, samples with NANOG1 or NP8 induction for 5 days in AD condition were studied (Figure 4-1).

RNA-Seq

Total RNA was extracted using the RNeasy RNA purification kit (Oiagen, Valencia, CA, USA) and contaminating DNA was removed by on-column DNase digestion. One hundred nanogram of RNA was used to synthesize complementary DNA libraries using NuGEN's Ovation RNA-Seq System. The libraries were sequenced by 2x75 bases paired-end on HiSeq 2000 instrument (Illumina, San Diego, CA).

ChIP-Seq

Before DNA extraction, cells were fixed in formaldehyde and lysed. Genomic DNA was then extracted and sheared by sonication, captured by anti-NANOG antibody (H-155, cat# sc-33759; Santa Cruz Biotechnology, Santa Cruz, CA, USA)

and pulled down by protein-A beads. The libraries were sequenced by 36 bases single-end on HiSeq 2000 instrument (Illumina, San Diego, CA).

Bioinformatics analysis

For RNA-Seq, 26-36 million pairs of read from paired end sequencing were produced per sample, of which 76-91% of reads were mapped to the reference human genome sequence (hg18) downloaded from UCSC. The program TopHat (version 2.0.7) (209), which implements Bowtie (version 2.0.6), was employed for mapping (210). RNA sequence fragments (or read counts) were enumerated by htseq-count from HTSeq package (version 0.5.3p9) (155), and genes with fewer than 10 fragments in all the samples were removed for DEG analysis. A total of > 14,000 genes that remained and DEGs were called using edgeR (version 3.0.8) (156) and a threshold cutoff of p-value < 0.05 and fold change (FC) > 1.5.

We performed comprehensive unsupervised hierarchical clustering and generated heatmaps using the DEGs in all experimental groups (Figure 4-1). The log2 ratios of each comparison were calculated by rescaling sum of the squares of each gene as 1.0. The unsupervised hierarchical clustering was computed by hclust function in R with Euclidean distance and ward clustering method, and the heatmap was plotted by heatmap.2 function in R.

Detailed procedures for GSEA, Oncomine analysis, IPA, and survival analysis based on gene signatures were discussed in Chapter 2. In short, transcriptome-associated biological phenotype was determined by GSEA by interrogating multiple PCa datasets, including those in cell lines (e.g., LNCaP), xenografts (e.g., LAPC9), normal human prostate basal vs. luminal cells, and different patient samples. The

Oncomine ‘Concept analysis’ was performed to discover the gene expression patterns in 16 PCa datasets with tumor-to-normal comparison and 13 datasets with metastasis-to-primary tumor comparison. For IPA, the ‘Upstream Regulator’ analysis was conducted with input of the DEGs filtered by cutoff at p-value < 0.05 and FC > 1.5. For gene signature-based survival analysis, we used the genes that passed p-value < 0.05 and median rank < 2000 in meta-analysis of comparison between primary versus metastasis in Concept Analysis. The genes that passed the statistical threshold and were shared in the datasets with survival information were selected as the signatures. The signatures were derived from each unsupervised cluster from previous analysis and contained 33 genes from clusters 1 and 3 and 58 genes from cluster 5 (see Results below). The predictor to classify subjects was computed by pairwise, ridge, LASSO linear regression using LMC package in R. The risk coefficients were estimated by the pairwise linear coefficient [107]. The optimized risk coefficient with the best discrimination power measured by Area under Curve of the combined predictor was applied in the training set to define a threshold for high and low risk group separation. The largest cohort (Setlur) (211) was the training set and the Taylor dataset (212) was the testing set.

For ChIP-Seq, we obtained 22-26 million reads per sample, of which 87-90% were mapped to human reference genome (hg18) and 66-70% were uniquely mapped. To quantify the peak density, only uniquely mapped reads were counted. Also, only one copy of the reads was kept for all the reads mapped to the same genomic location. The peaks were called using model-based analysis of ChIP-Seq (MACS) 1.3.7.1 (213). The scanning window size was set at 300 bp and the cutoff p-value was set at $1e^{-5}$.

Differential peaks were identified by comparison between with and without DOX and subjected to NANOG and NP8 immunoprecipitation after the background peak subtraction. The background was the peak captured by IgG and NANOG in pLVX vector only with DOX treatment. In total, we obtained 14,331 NANOG1 and 14,449 NP8 peaks, of which 1,342 NANOG1 and 1,313 NP8 peaks were located in the promoter region defined as −5 kb to +0.5 kb from transcription start site (TSS).

The distribution heatmap was generated by centering the peaks on NP8 with a range of upstream and downstream 10 kb. Whole regions were divided into 250 bins, and the average reads per kilobase transcript per million reads (RPKM) were calculated in each bin and plotted by heatmap.2 function in R. The presentation order was sorted by intensity of NP8 peaks and the classification of co-occupancy of nuclear factors, including three factors (NP8/FOXA1/AR), two factors (NP8/FOXA1 or NP8/AR), and one factor (NP8 only). The enriched motifs of 100 bp flanking the summit of top 800 peaks in NANOG and NP8 binding region were identified by MEME-ChIP (151) and compared to known motifs by TOMTOM in MEME Suite (version 4.9.0).

Of note, both RNA-Seq and ChIP-Seq data were analyzed by collaboration with Dr. Yue Lu and her bioinformatics team of Next Generation Sequencing (NGS) core at Science Park.

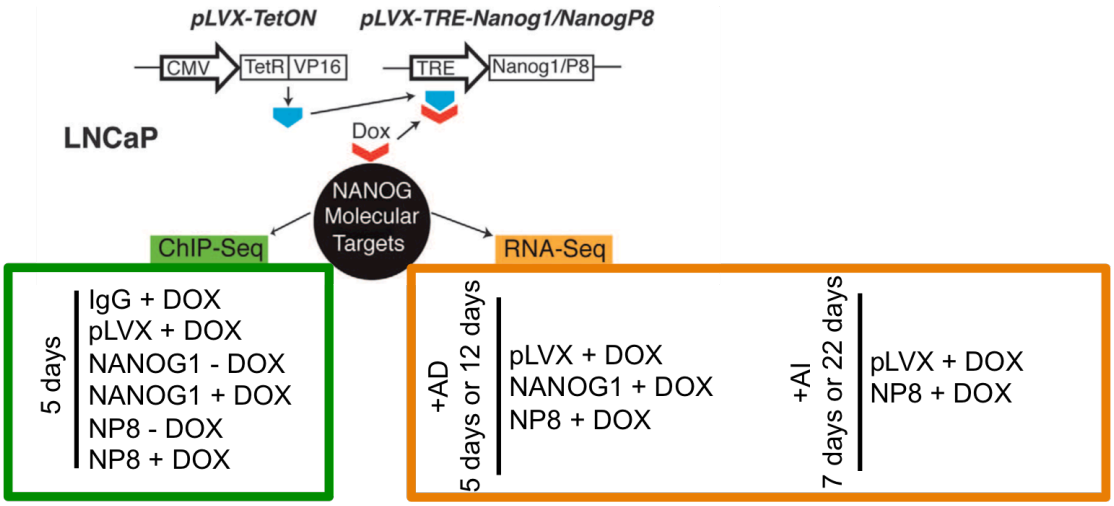


Figure 4-1. Experimental schema (adapted from (151) Jeter, C. R., B. Liu, Y. Lu, H. P. Chao, D. Zhang, X. Liu, X. Chen, Q. Li, K. Rycaj, T. Calhoun-Davis, L. Yan, Q. Hu, J. Wang, J. Shen, S. Liu, and D. G. Tang. 2016. NANOG reprograms prostate cancer cells to castration resistance via dynamically repressing and engaging the AR/FOXA1 signaling axis. *Cell discovery* 2: 16041).

The inducible NANOG1 or NP8 expression was controlled by the binary Tet-ON system, which overexpresses target genes via the tetracycline response element (TRE) after administration of doxycycline (DOX). The pLVX vector only was the control group. For ChIP-Seq, samples were collected at day 5 after DOX induction. For RNA-Seq, samples were harvested at two different time points for both AD and AI conditions.

4.4 Results

The main goal of this project is to understand how NANOG reprograms the bulk AD LNCaP cells to the AI, stem cell-like state (93). Practically, through transcriptome and ChIP-Seq analyses, we hope to test the two alternative, although not necessarily mutually exclusive, hypotheses stated earlier, i.e., through engaging the pluripotency network and/or mainly through engaging the lineage-specific TFs. We first focused on transcriptome analysis in LNCaP cells upon NANOG1 or NP8 induction for various intervals of time under either AD (serum-containing) and AI (in CDSS medium) conditions (Figure 4-1). Unsupervised hierarchal clustering showed that the overall DEG profile of NANOG1 induced cells was very similar to that in NP8-induced cells in AD condition. Dendrogram of hierarchical clustering of DEGs revealed that the overall transcriptional responses to NANOG induction were more similar in AD or AI conditions than between AD and AI conditions. Of interest, total DEGs were clustered into 11 groups (data not shown) with 6 clusters identified that could potentially be involved in mediating distinct transcriptional responses involved in NANOG-induced LNCaP cell reprogramming (Figure 4-2). Genes in clusters 1, 3, and 4 appeared to be only regulated by NANOG induction independently from androgen context and time course of induction, but genes in clusters 2, 5, and 6 showed time and androgen dependence coincident with the effect by NANOG induction (Figure 4-2).

The DEGs in clusters 1 (95 genes) and 3 (163 genes) exhibited an overall similar expression pattern, which was persistently suppressed by NP8 (Figure 4-2). The Oncomine concept analysis revealed that 42 of the total 258 genes in clusters 1 and 3 were downregulated in primary prostate tumors compared to normal tissues and

58 genes downregulated in metastases compared to primary tumors (Figure 4-3A). Also, 20 genes, including *AHNAK* and *AZGP1*, were commonly downregulated in both comparisons (Figure 4-3B). Both *AHNAK* and *AZGP1* have been reported to be tumor suppressors (214, 215), and downregulation of these genes in primary tumors and metastases imply potential tumor-suppressive functions for, at least, some of these genes. Consistent with this idea, GSEA revealed that NANOG-repressed genes in cluster 1 were associated with the gene expression profiles present in normal differentiated (mature) prostate luminal epitheilal, differentiated PSA^+ PCa, low-grade tumor, and AD PCa cells, as well as in patient tumors before ADT (Figure 4-3C). Thus, genes in cluster 1 are generally associated with more differentiated, low-grade, and androgen-sensitive tumor phenotypes. A similar association was also observed with the NANOG-repressed genes in cluster 3 (Figure 4-3D) and in the analysis of DEGs in both cluster 1 and cluster 3 (Figure 4-3E). Significantly, the cluster 1 DEGs repressed by NANOG contained classical AR downstream targets involved in cell differentiation, such as KLK3 (PSA), KLK2, NKX3.1, TMPRSS2, LRIG1, and ELL2. To determine the potential clinical significance of the NANOG-repressed genes in clusters 1 and 3, we derived a 33-gene signature using the clinical information in Oncomine applying comparison of differentially expressed genes in primary and metastatic tumor, which successfully classified patients into high- and low-risk groups in a training cohort (Figure 4-3). Cox proportional hazards model computing survivor function and Kaplan-Meier survival analysis successfully predicted clinical outcome in an independent testing cohort, with low expression of these genes associated with poor outcome (Figure 4-3F). These results suggest that the genes in cluster 1 and 3 are not only

associated with differentiated and low-grade tumor phenotypes but also are candidates of gene signatures to predict the clinical outcome.

DEGs in clusters 2 and 6 (136 and 73 genes, respectively) shared somewhat similar dynamic expression patterns, first upregulated by NP8 in AD conditions at day 5 and then moderately downregulated at day 12. Cluster 2 genes were significantly downregulated in AI conditions at both time points (Figure 4-2). Of interest, genes in these two clusters included AKR1C1, AKR1C3, UGT2B4, UGT2B15, UGT2B17, CREB5, BMP6, ID1 and ID3, many of which are well known to be involved in androgen metabolism and catabolism (see Discussion). In sharp contrast to cluster 1 and 3 DEGs comprising predominantly downregulated genes, the genes in cluster 4 (139 genes) were persistently activated by NP8 (Figure 4-2). Many of the genes in this cluster, exemplified by ABCG2, LMO2, TERT, NEDD9, and RDX, are known to be expressed in, and functionally regulate, SCs and CSCs. Indeed, GSEA correlated this DEG cluster with CSC and castration-resistant phenotypes (Figure 4-4). We also examined the mRNA levels of iPSC reprogramming factors including OCT4, SOX2, KLF4, and MYC, in our RNA-Seq data. As shown in Table 4-1, none of these factors except MYC were significantly upregulated. We previously showed that NANOG induction in LNCaP cells led to increased MYC protein levels in vitro and in vivo (93). The current RNA-Seq results indicated that NANOG induced MYC mRNA levels in AI d7 (2.14X p-value < 0.001) and AI d22 (1.63X p-value < 0.001), suggesting that NANOG upregulates MYC in androgen deprivation conditions.

Genes in cluster 5 were upregulated by NP8 in the absence of androgen but more prominently induced in long-term AI conditions (Figure 4-2). Of the total 240

DEGs in cluster 5, the Oncomine concept analysis showed that 59 genes were significantly upregulated in patient tumors compared to normal prostate tissues and 122 in metastases compared to primary tumors (Figure 4-5A). Also, 40 genes were commonly upregulated in both comparisons (Figure 4-5B). Annotation of the DEGs in cluster 5 unveiled that, strikingly, more than half of the genes were involved in DNA replication, cell cycle, and mitosis, highlighted by *CDK1, CKS2, UBE2C, BUB1, AURKB*, and *MCM5/7*. In support, GSEA demonstrated that the cluster 5 DEGs were associated with more aggressive, metastatic, and castration-resistant tumor phenotypes (Figure 4-5C). A 58-gene signature was derived from cluster 5, including all the genes significantly upregulated in metastasis versus primary comparison by Oncomine Concept analysis, which successfully predicted patients' survival outcome in a training set and independent testing dataset by Cox proportional hazards model and Kaplan-Meier analysis (Figure 4-5D).

The above time-related RNA-Seq analysis in both AD and AI conditions has revealed changes in AR transcriptome and in genes associated with SC/CSC regulation, DNA replication, cell-cycle control, and cell motility but not in genes associated with OCT4/SOX2 signaling. These results seem to suggest that NANOG-induced LNCaP cell reprogramming may primarily engage prostate/PCa-specific lineage factors such as AR rather than the endogenous reprogramming and pluripotency factors such as OCT4 and SOX2. To test this possibility, we examined NANOG genomic occupancy in LNCaP cells by performing NANOG1 and NP8 ChIP-Seq experiments (Figure 4-1). A total 14,449 NP8 and 14,331 NANOG1 peaks were identified, respectively. MEME motif analysis followed by TOMTOM analysis revealed

that, to our surprise, the FOXA1 motifs were the top and most highly enriched motifs, identified in 583 peaks of the top 800 NP8-binding sites (Figure 4-6A). FOXA1 is a well-established 'pioneer' factor for AR. A meta analysis of center of distribution (+/- 10 Kb) heatmaps using published data confirmed that, remarkably, AR and FOXA1 co-occupied more than 50% of NP8-binding sites, and AR or FOXA1 separately co-occupied another 40% of NP8-binding sites with each occupying 20% of the sites (Figure 4-6B) among all the 14,449 NP8 peaks. Examination of NANOG binding to the genomic loci of OSKM revealed only slightly increased Nanog binding to Myc at -10 kb (151). These analyses support the scenario that NANOG reprograms LNCaP PCa cells by physically binding to the AR and FOXA1 genomic loci and primarily 'interfacing' with the AR/FOXA1 signaling axis.

Integrated analysis of the ChIP-Seq and RNA-Seq data provides further support to this interpretation. For instance, the DEGs in clusters 1 and 3 were mostly (61%) co-occupied by NP8, AR, and FOXA1 simultaneously (Figure 4-6C). In cluster 4, two genomic binding patterns were enriched, NP8/AR/FOXA1 co-occupation or NP8-only (Figure 4-6C). On the other hand, the DEGs in clusters 2 and 6 were co-occupied by NP8/AR/FOXA1, NP8/AR, or NP8/FOXA1 (Figure 4-6C). These results, collectively, indicate that NANOG reprograms LNCaP cells to the castration-resistant 'stemness' state (93) via, primarily and dynamically, engaging the AR/FOX1 transcription factors in a dynamic manner. Surprisingly, however, none of these three transcription factors bound to the genomic regions of DEGs in cluster 5 (Figure 4-6C). When we performed IPA Upstream Regulator analysis in attempt to decipher the potential transcription factors involved in regulating the cluster 5 genes, we found that

MYC was the common upstream transcription factor of a cohort of mitosis genes upregulated by NP8 induction (Figure 4-6D). Reciprocally, these DEGs were also enriched in a MYC oncogenic signature (Figure 4-6E). Therefore, during NANOG-induced reprogramming of AD LNCaP cells, MYC might be the pivotal downstream transcriptional regulator of the mitosis and cell cycle-related genes.

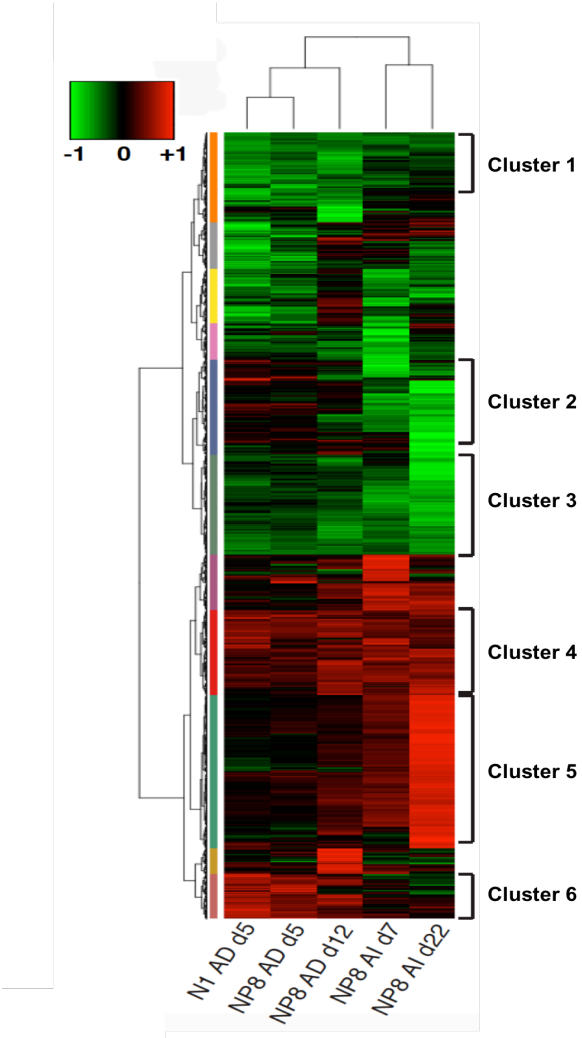


Figure 4-2. Unsupervised hierarchical clustering and heatmap for differentially expressed genes (adapted from (151) Jeter, C. R., B. Liu, Y. Lu, H. P. Chao, D. Zhang, X. Liu, X. Chen, Q. Li, K. Rycaj, T. Calhoun-Davis, L. Yan, Q. Hu, J. Wang, J. Shen, S. Liu, and D. G. Tang. 2016. NANOG reprograms prostate cancer cells to castration resistance via dynamically repressing and engaging the AR/FOXA1 signaling axis. *Cell discovery* 2: 16041). DEGs were defined as the gene expression that passed the cutoffs, FC ≥ 1.5 or ≤ 0.67 and p-value < 0.05. In total, there were 1,154 DEGs, including 215, 262, 420, and 605 DEGs, for NP8 induction in AD condition collected at day 5 and 12, and in AI condition collected at day 7 and 22, respectively.

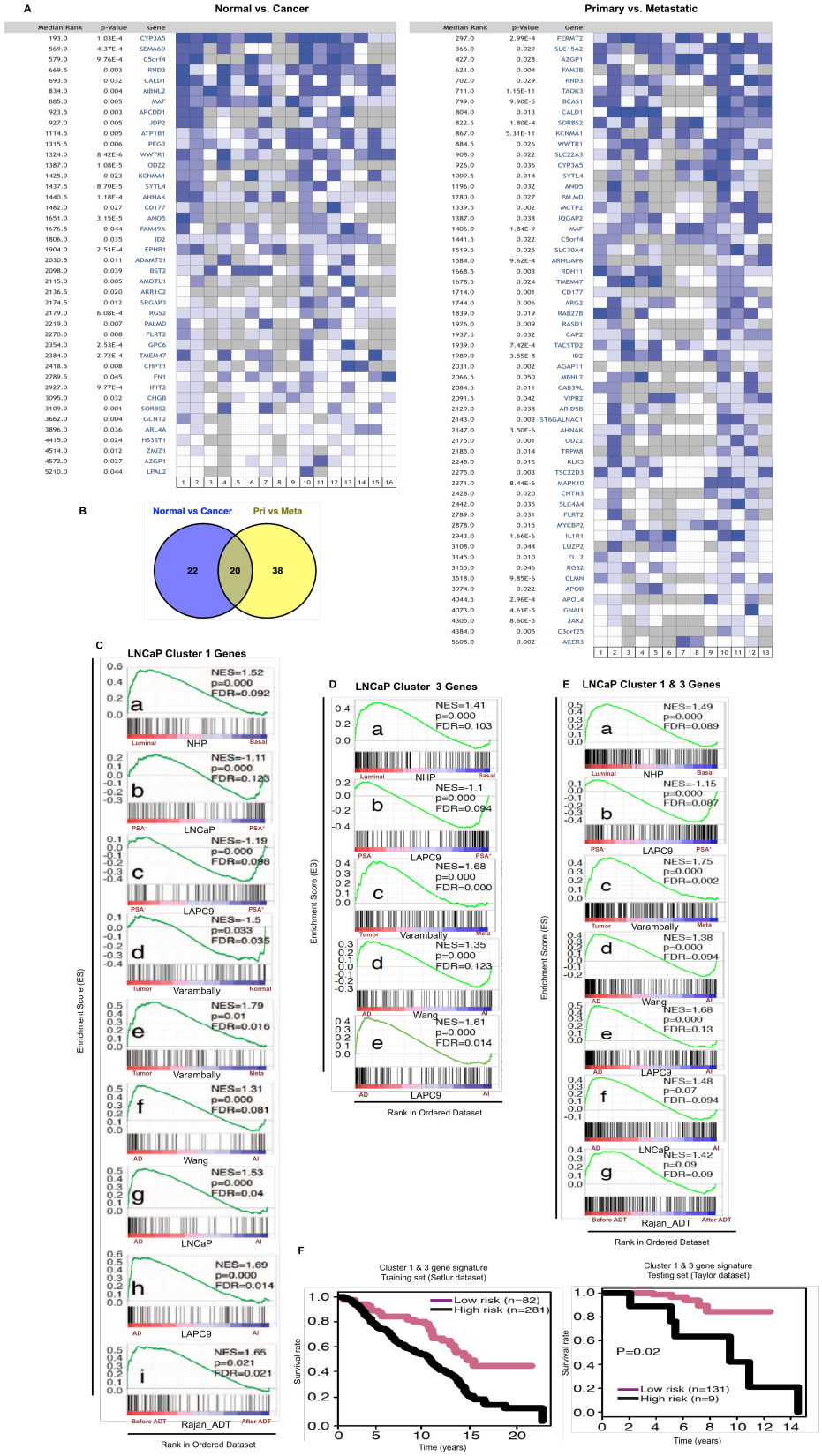

A
Normal vs. Cancer
Primary vs. Metastatic
B
Normal vs Cancer
Pri vs Meta
22
20
38
C
LNCaP Cluster 1 Genes
D
LNCaP Cluster 3 Genes
E
LNCaP Cluster 1 & 3 Genes
Enrichment Score (ES)
Rank in Ordered Dataset
F
Cluster 1 & 3 gene signature
Training set (Setlur dataset)
Low risk (n=82)
High risk (n=281)
Survival rate
Time (years)
Cluster 1 & 3 gene signature
Testing set (Taylor dataset)
P=0.02
Low risk (n=131)
High risk (n=9)

Figure 4-3. Deciphering transcriptome profiles of cluster 1 and 3 DEGs (adapted from (151) Jeter, C. R., B. Liu, Y. Lu, H. P. Chao, D. Zhang, X. Liu, X. Chen, Q. Li, K. Rycaj, T. Calhoun-Davis, L. Yan, Q. Hu, J. Wang, J. Shen, S. Liu, and D. G. Tang. 2016. NANOG reprograms prostate cancer cells to castration resistance via dynamically repressing and engaging the AR/FOXA1 signaling axis. *Cell discovery* 2: 16041).

A. Heatmaps of genes (in clusters 1 and 3) downregulated in prostate tumors compared to normal/benign tissues and metastases compared to primary tumors by Concept Analysis in Oncomine database. Only the genes with significant downregulation (p-value < 0.05) were presented in the heatmaps.

B. Venn diagram of the commonly downregulated genes in both comparisons. In total, 42 genes were significantly downregulated in normal vs. cancer comparisons, 58 genes in primary tumor vs. metastases comparisons, and 20 genes in both comparisons.

C-E. GSEA showing gene signature enrichment in cluster 1 (C), cluster 3 (D), and cluster 1 and 3 (E) of DEGs. The DEGs classified into cluster 1 were associated with more differentiated tumor type or cell type, AD tumor, and before ADT.

F. A 33-gene expression signature derived from clusters 1 and 3 as a predictor of survival in PCa patients. Thirty-three genes were selected from comparison of primary vs. metastatic tumors based on Concept Analysis from Oncomine. The Setlur dataset served as training set and the Taylor dataset as the testing set.

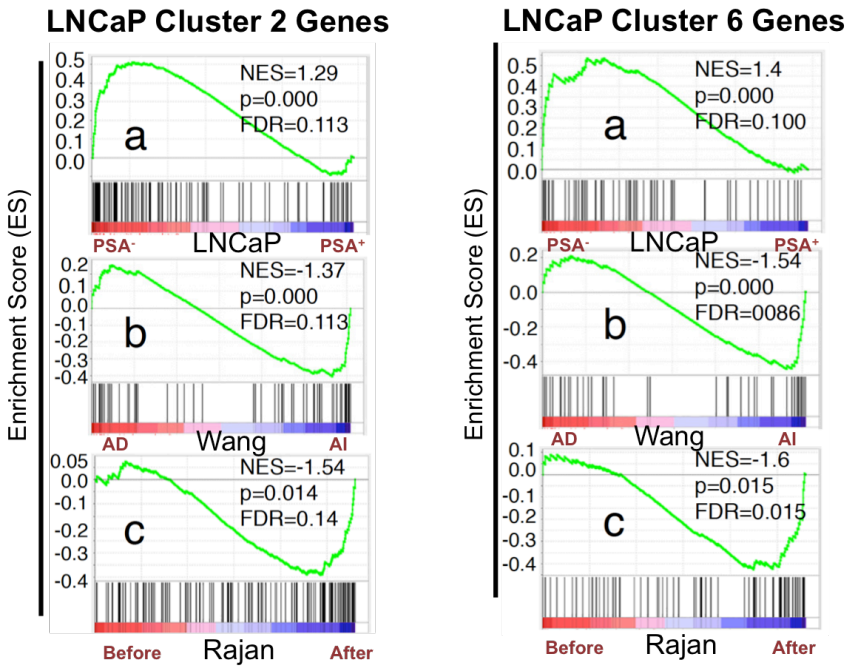


Figure 4-4. Gene signatures uncovered in the cluster 2 and cluster 6 DEGs (adapted from (151) Jeter, C. R., B. Liu, Y. Lu, H. P. Chao, D. Zhang, X. Liu, X. Chen, Q. Li, K. Rycaj, T. Calhoun-Davis, L. Yan, Q. Hu, J. Wang, J. Shen, S. Liu, and D. G. Tang. 2016. NANOG reprograms prostate cancer cells to castration resistance via dynamically repressing and engaging the AR/FOXA1 signaling axis. *Cell discovery* 2: 16041).

The DEGs classified into cluster 2 were associated with biological phenotypes, such as PSA^-, AI condition, and after ADT. On the other hand, the DEGs clustered into cluster 6 tend to associated with transcriptome profiles of PSA^+ cell, AD patient, and patient tumor before ADT.

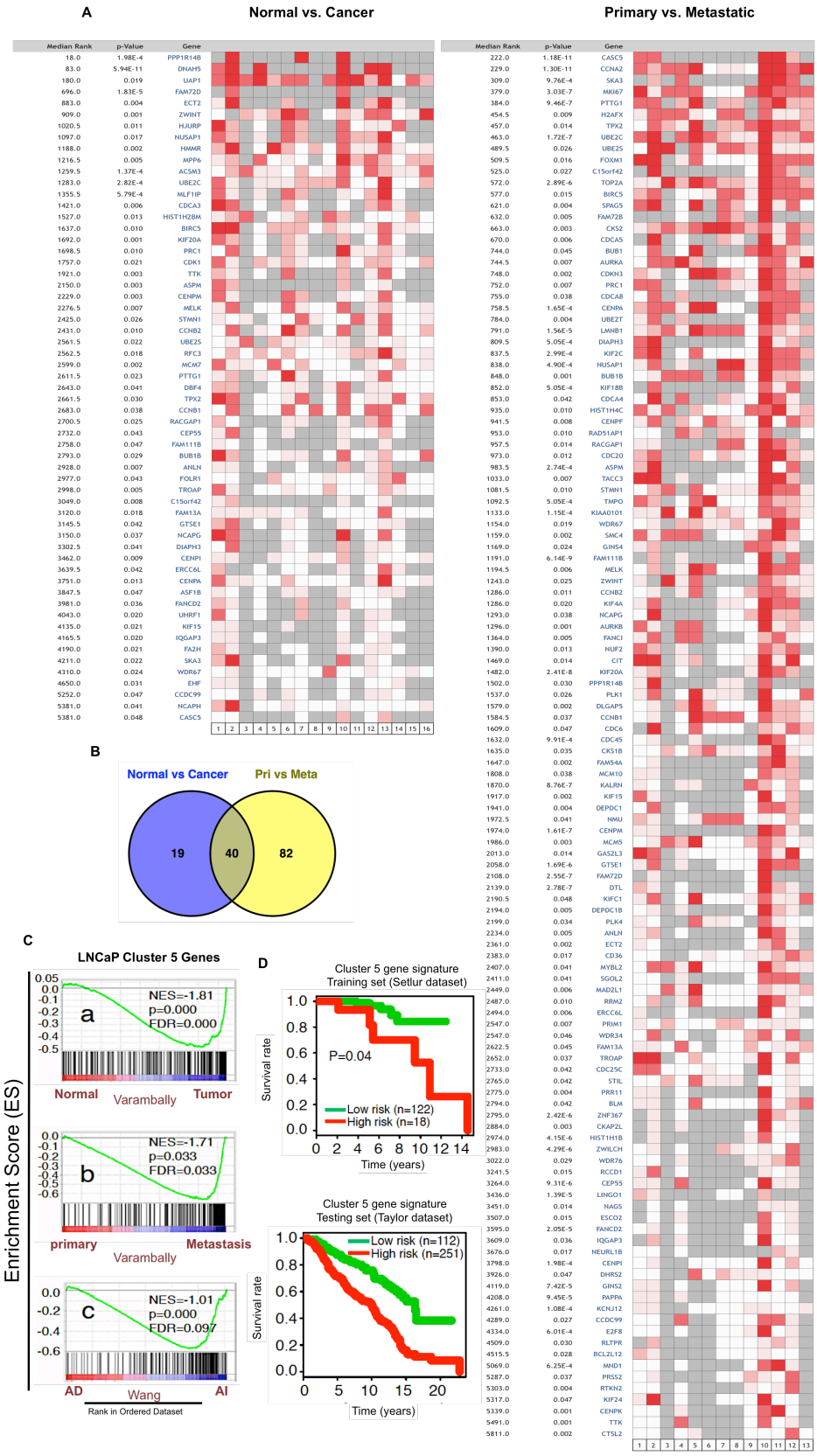

A
Normal vs. Cancer
Median Rank
p-Value
Gene
18.0 1.98E-4 PPP1R14B
83.0 5.94E-11 DNAH5
180.0 0.019 UAP1
696.0 1.83E-5 FAM72D
883.0 0.004 ECT2
909.0 0.001 ZWINT
1020.5 0.011 HJURP
1097.0 0.017 NUSAP1
1188.0 0.002 HMMR
1216.5 0.005 MPP6
1259.5 1.37E-4 ACSM3
1283.0 2.82E-4 UBE2C
1355.5 5.79E-4 MLF1IP
1421.0 0.006 CDCA3
1527.0 0.013 HIST1H2BM
1637.0 0.010 BIRC5
1692.0 0.001 KIF20A
1698.5 0.010 PRC1
1757.0 0.021 CDK1
1921.0 0.003 TTK
2150.0 0.003 ASPM
2229.0 0.003 CENPM
2276.5 0.007 MELK
2425.0 0.026 STMN1
2431.0 0.010 CCNB2
2561.5 0.022 UBE2S
2562.5 0.018 RFC3
2599.0 0.002 MCM7
2611.5 0.023 PTTG1
2643.0 0.041 DBF4
2661.5 0.030 TPX2
2683.0 0.038 CCNB1
2700.5 0.025 RACGAP1
2732.0 0.043 CEP55
2758.0 0.047 FAM111B
2793.0 0.029 BUB1B
2928.0 0.007 ANLN
2977.0 0.043 FOLR1
2998.0 0.005 TROAP
3049.0 0.008 C15orf42
3120.0 0.018 FAM13A
3145.5 0.042 GTSE1
3150.0 0.037 NCAPG
3302.5 0.041 DIAPH3
3462.0 0.009 CENPI
3639.5 0.042 ERCC6L
3751.0 0.013 CENPA
3847.5 0.047 ASF1B
3981.0 0.036 FANCD2
4043.0 0.020 UHRF1
4135.0 0.021 KIF15
4165.5 0.020 IQGAP3
4190.0 0.021 FA2H
4211.0 0.022 SKA3
4310.0 0.024 WDR67
4650.0 0.031 EHF
5252.0 0.047 CCDC99
5381.0 0.041 NCAPH
5381.0 0.048 CASC5
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Primary vs. Metastatic
Median Rank
p-Value
Gene
222.0 1.18E-11 CASC5
229.0 1.30E-11 CCNA2
309.0 9.76E-4 SKA3
379.0 3.03E-7 MKI67
384.0 9.46E-7 PTTG1
454.5 0.009 H2AFX
457.0 0.014 TPX2
463.0 1.72E-7 UBE2C
489.5 0.026 UBE2S
509.5 0.016 FOXM1
525.0 0.027 C15orf42
572.0 2.89E-6 TOP2A
577.0 0.015 BIRC5
621.0 0.004 SPAG5
632.0 0.005 FAM72B
663.0 0.003 CKS2
670.0 0.006 CDCA5
744.0 0.045 BUB1
744.5 0.007 AURKA
748.0 0.002 CDKN3
752.0 0.007 PRC1
755.0 0.038 CDCA8
758.5 1.65E-4 CENPA
784.0 0.004 UBE2T
791.0 1.56E-5 LMNB1
809.5 5.05E-4 DIAPH3
837.5 2.99E-4 KIF2C
838.0 4.90E-4 HJURP
848.0 0.001 BUB1B
852.0 5.05E-4 KIF18B
853.0 0.042 CDCA4
935.0 0.010 HIST1H4C
941.5 0.008 CENPF
953.0 0.010 RAD51AP1
957.5 0.014 RACGAP1
973.0 0.012 CDC20
983.5 2.74E-4 ASPM
1033.0 0.007 TACC3
1081.5 0.010 STMN1
1092.5 5.05E-4 TMPO
1133.0 1.15E-4 KIAA0101
1154.0 0.019 WDR67
1159.0 0.002 SMC4
1169.0 0.024 GINS4
1191.0 6.14E-9 FAM111B
1194.5 0.006 MELK
1243.0 0.025 ZWINT
1286.0 0.011 CCNB2
1286.0 0.020 KIF4A
1293.0 0.038 NCAPG
1296.0 0.001 AURKB
1364.0 0.005 FANCI
1390.0 0.013 NUF2
1469.0 0.014 CIT
1482.0 2.41E-8 KIF20A
1502.0 0.030 PPP1R14B
1537.0 0.026 PLK1
1579.0 0.002 DLGAP5
1584.5 0.037 CCNB1
1609.0 0.047 CDC6
1632.0 9.91E-4 CDC45
1635.0 0.035 CKS1B
1647.0 0.002 FAM54A
1808.0 0.038 MCM10
1870.0 8.76E-7 KALRN
1917.0 0.002 KIF15
1941.0 0.004 DEPDC1
1972.5 0.041 NMU
1974.0 1.61E-7 CENPM
1986.0 0.003 MCM5
2013.0 0.014 GAS2L3
2058.0 1.69E-6 GTSE1
2108.0 2.55E-7 FAM72D
2139.0 2.78E-7 DTL
2190.5 0.048 KIFC1
2194.0 0.005 DEPDC1B
2199.0 0.034 PLK4
2234.0 0.005 ANLN
2361.0 0.002 ECT2
2383.0 0.017 CD36
2407.0 0.041 MYBL2
2411.0 0.041 SGOL2
2449.0 0.006 MAD2L1
2487.0 0.010 RRM2
2494.0 0.006 ERCC6L
2547.0 0.007 PRIM1
2547.0 0.046 WDR34
2622.5 0.045 FAM13A
2652.0 0.037 TROAP
2733.0 0.042 CDC25C
2765.0 0.042 STIL
2775.0 0.004 PRR11
2794.0 0.042 BLM
2795.0 2.42E-6 ZNF367
2884.0 0.003 CKAP2L
2974.0 4.15E-6 HIST1H1B
2983.0 4.29E-6 ZWILCH
3022.0 0.029 WDR76
3241.5 0.015 RCCD1
3264.0 9.31E-6 CEP55
3436.0 1.39E-5 LINGO1
3451.0 0.014 NAGS
3507.0 0.015 ESCO2
3595.0 2.05E-5 FANCD2
3609.0 0.036 IQGAP3
3676.0 0.017 NEURL1B
3798.0 1.98E-4 CENPI
3926.0 0.047 DHRS2
4119.0 7.42E-5 GINS2
4208.0 9.45E-5 PAPPA
4261.0 1.08E-4 KCNJ12
4289.0 0.027 CCDC99
4334.0 6.01E-4 E2F8
4509.0 0.030 RLTPR
4515.5 0.028 BCL2L12
5069.0 6.25E-4 MND1
5287.0 0.037 PRSS2
5303.0 0.004 RTKN2
5317.0 0.047 KIF24
5339.0 0.001 CENPK
5491.0 0.001 TTK
5811.0 0.002 CTSL2
1 2 3 4 5 6 7 8 9 10 11 12 13
B
Normal vs Cancer
Pri vs Meta
19
40
82
C
LNCaP Cluster 5 Genes
Enrichment Score (ES)
a
NES=-1.81
p=0.000
FDR=0.000
Normal
Varambally
Tumor
b
NES=-1.71
p=0.033
FDR=0.033
primary
Varambally
Metastasis
c
NES=-1.01
p=0.000
FDR=0.097
AD
Wang
AI
Rank in Ordered Dataset
D
Cluster 5 gene signature
Training set (Setlur dataset)
Survival rate
P=0.04
Low risk (n=122)
High risk (n=18)
Time (years)
Cluster 5 gene signature
Testing set (Taylor dataset)
Survival rate
Low risk (n=112)
High risk (n=251)
Time (years)

Figure 4-5. Deciphering the transcriptome profile of cluster 5 DEGs (adapted from (151) Jeter, C. R., B. Liu, Y. Lu, H. P. Chao, D. Zhang, X. Liu, X. Chen, Q. Li, K. Rycaj, T. Calhoun-Davis, L. Yan, Q. Hu, J. Wang, J. Shen, S. Liu, and D. G. Tang. 2016. NANOG reprograms prostate cancer cells to castration resistance via dynamically repressing and engaging the AR/FOXA1 signaling axis. *Cell discovery* 2: 16041).

A. Heatmaps showing the genes in cluster 5 that were upregulated in prostate tumor vs. normal/benign tissue (left) and metastases vs. primary tumor (right) comparisons. Heatmaps presented the genes in cluster 5 with significant upregulation (median rank < 6000 and p-value < 0.05) in Oncomine by Concept Analysis.

B. Venn diagram of the common DEGs in both comparisons of prostate tumor vs. normal/benign tissue and metastases vs. primary tumor in (A). In total, 40 genes were significantly upregulated in both comparisons. Also, there were 19 genes uniquely upregulated in normal/benign vs. cancer and 82 genes in primary vs. metastatic tumor.

C. The gene signatures uncovered in the cluster 5. All the DEGs classified into cluster 5 were ranked and analyzed. The DEGs in cluster 5 were associated with tumor development, metastasis, and AI phenotype by GSEA.

F. A 58-gene expression signature derived from cluster 5 as a predictor of survival in PCa patients. The 58 genes were selected by comparisons of primary vs. metastatic tumors based on Oncomine Concept Analysis. The Setlur dataset was used as the training set and Taylor dataset was testing set.

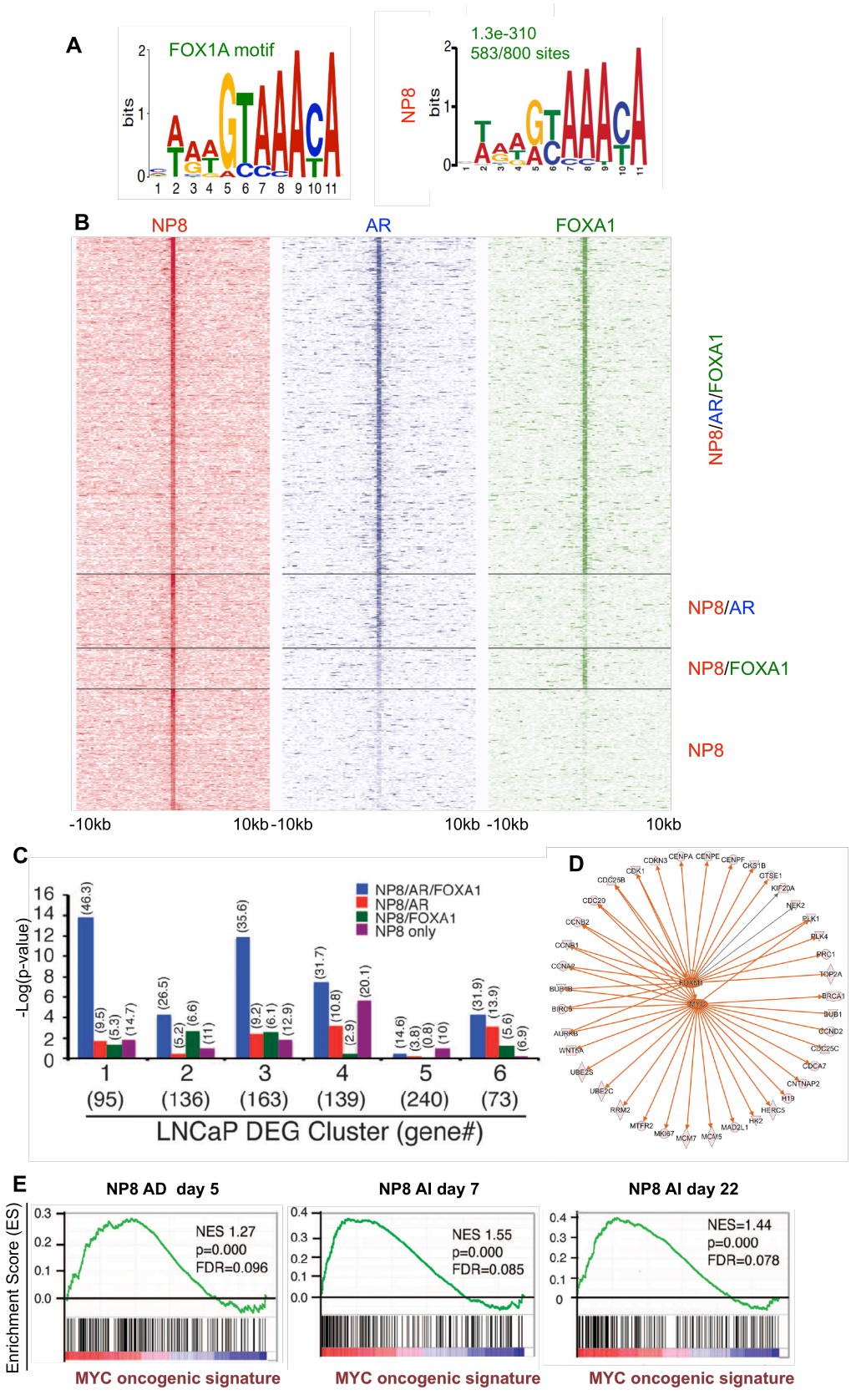

A
FOX1A motif
bits
NP8
1.3e-310
583/800 sites
B
NP8
AR
FOXA1
NP8/AR/FOXA1
NP8/AR
NP8/FOXA1
NP8
-10kb
10kb
C
-Log(p-value)
NP8/AR/FOXA1
NP8/AR
NP8/FOXA1
NP8 only
LNCaP DEG Cluster (gene#)
D
E
NP8 AD day 5
NP8 AI day 7
NP8 AI day 22
Enrichment Score (ES)
NES 1.27
p=0.000
FDR=0.096
NES 1.55
p=0.000
FDR=0.085
NES=1.44
p=0.000
FDR=0.078
MYC oncogenic signature

Figure 4-6. NANOG reprograms LNCaP cells via engaging FOXA1 and AR (adapted from (151) Jeter, C. R., B. Liu, Y. Lu, H. P. Chao, D. Zhang, X. Liu, X. Chen, Q. Li, K. Rycaj, T. Calhoun-Davis, L. Yan, Q. Hu, J. Wang, J. Shen, S. Liu, and D. G. Tang. 2016. NANOG reprograms prostate cancer cells to castration resistance via dynamically repressing and engaging the AR/FOXA1 signaling axis. *Cell discovery* 2: 16041).

A. The FOXA1 motif was identified as the most highly enriched in the top 800 peaks from NP8 ChIP-Seq, as analyzed by MEME followed by TOMTOM. The left is FOXA1 motif and the right is the frequency of FOXA1 motif bound by NP8 with the significant occurrence (E-value = 1.3e-310).

B. Intensity distribution heatmap of ChIP-Seq signal centered by NP8 peaks with a range of ±10 kb. The peaks were sorted by intensity of NP8 signal.

C. Inspection of NP8, AR, and FOXA1 co-occupancy on the DEGs in each cluster. The co-occupancy was examined by Fisher's exact test with range of ± 50kb from the center of the peak and y-axis was the p-values of NP8/AR/FOXA1, NP8/AR, NP8/FOXA1, or NP8 occupancy on the DEGs in each cluster.

D. The Upstream Regulator Analysis of gene expression reveals MYC and FOXM1 as the two major NP8 downstream transcription factors in regulating cluster 5 genes. The Upstream Regulator analysis was performed for a cohort of genes lack of binding by any of NP8, AR or FOXA1 to find the potential downstream mediators of Nanog signaling. MYC (Z-score = 4.262 and p-value = $6e^{-7}$) and FOXM1 (Z-score = 4.578 and p-value = $1e^{-23}$) were predicted as two major NP8 downstream transcription factor mediating its transcriptional network.

E. MYC oncogenic signature was enriched in NP8 induced transcriptome, under both AD and AI conditions.

4.5 Discussion

The main goal of this project was to understand how, at the molecular level, a single stemness factor, NANOG, could reprogram a largely differentiated somatic PCa cell line, LNCaP, to the castration-resistant and stem-like state. Bioinformatics analyses of the integrated RNA-Seq and ChIP-Seq results suggest that NANOG accomplishes this feat, primarily, by dynamically repressing and engaging the AR/FOXA1 signaling axis, although NANOG also engages MYC to regulate cell cycle/mitosis related genes, especially during late stages of reprogramming under AR-blocked conditions (Figure 4-7). Importantly, integrated analyses of RNA-Seq and ChIP-Seq data reveals an intricate portrait of NANOG genomic binding and its transcriptional output in the course of LNCaP cell reprogramming (Figure 4-7). Many DEGs in clusters 1 and 3 are conventional AR downstream genes, which are repressed by NP8 and co-occupied by NP8/AR/FOXA1. These genes include *KLK3*, which encodes PSA and represents the best know AR target and a clinically used diagnostic and prognostic biomarker for PCa, *NKX3.1*, an androgen-regulated homeobox TF, and *ELL2*, a transcriptional elongation factor. Both NKX3.1 and ELL are direct downstream target genes of AR contributing to prostate differentiation (216, 217). Reduced *KLK3* mRNA expression is consistent with earlier findings that overexpression of NANOG1 or NP8 reduces PSA protein levels (151). Altogether, these results suggest that early during, and throughout the process of, NANOG-induced LNCaP cell reprogramming, NANOG impedes AR/FOXA1 transcription machinery to suppresses 'conventional' AR target genes and genes associated with cell differentiation and tumor suppression (Figure 4-7).

Other interesting genes in clusters 1 and 3 that also possess tumor-suppressive functions include AHANK (214) and AZGP1 (215). Both genes function as tumor suppressors through TGFβ signaling. TGFβ is well known to function as a tumor suppressor via SMAD in early stages of tumorigenesis and as a tumor promoter in advanced and metastatic tumors via non-SMAD pathways (218). AHANK directly interacts with Smad2/3 and the complex translocates into the nucleus, resulting in downregulation of c-MYC, inhibition of cyclin D/CDK4 and subsequent cell-cycle arrest (214). The AHANK-MYC connection through Smad2/3 could indirectly contribute to NANOG-induced MYC expression and transcriptome. AZGP1, on the other hand, functions as gatekeeper for TGFβ1-mediated ERK2 phosphorylation, which blocks cross talk between TGFβ/ERK signaling and epithelial-mesenchymal transition (215). Although the biological functions of AHNAK and AZGP1 have been investigated in other cancers, our current findings also implicate both molecules, and, potentially, TGFβ, in NANOG-induced pro-differentiation and tumor suppression during the reprogramming process.

The DEGs in cluster 2 and, in particular, cluster 6, are upregulated at the early stage of NP8 induction in AD conditions but then downregulated by NP8 in AI conditions (Figure 4-2). Several members (e.g., AKR1C1, AKR1C3, UGT2B4, UGT2B15, and UGT2B17) of the two steroid hormone metabolic enzyme families, uridine diphosphate-glucuronosyltransferase (UGT) and aldo-keto reductase (AKR), are in the DEGs in these 2 clusters. AKR1Cs are known for catalyzing the reduction and inactivation of DHT (219). The glucuronidation activities of UGT2B15 and UGT2B17 locally inactivate androgens. Suppressing UGT2B15 and UGT2B17 leads

to increased bioactive DHT and stimulates the expression of AR dependent genes (220). Thus, NANOG-induced upregulation of AKR1C1, AKR1C3, UGT2B4, UGT2B15, and UGT2B17 genes in AD conditions might cooperate with repression of AR downstream pro-differentiation genes (in clusters 1 and 3) to dampen AR-mediated differentiation signaling at early time (day 5) (Figure 4-7). On the other hand, under AI conditions and especially at a later time point (day 22) when LNCaP cells are partially or even completely reprogramed (Figure 4-7), these gene are downregulated, which, theoretically, might lead to decreased androgen catabolism and increased androgen levels in the cells. This phenomenon, called intracrine androgen synthesis (221, 222), has been reported in PCa cell cultures in vitro and prostate tumors in vivo.

The DEGs in cluster 4 are very interesting and include many involved in SC and CSC genes, which are persistently co-occupied by NP8/AR/FOXA1 and upregulated by NANOG in both AD and AI conditions. For example, ABCG2, similar to CD133 and integrin α2β1 (223), is one of the PCa stem/progenitor cell markers and the ABCG2$^+$, CD133$^+$ and α2β1$^+$ cell populations overlap with one another (103). ABCG2 is an ABC transporter and functions as an energy-dependent efflux pump that has been implicated in multi-drug resistance (224). Upregulated ABCG2 might contribute to NANOG-induced castration resistance in LNCaP cells by pumping out steroid hormones, as reported by Huss et al (225). LMO2, a regulator of yolk sac erythropoiesis (226), has been implicated in PCa progression, and LMO2 mRNA expression correlates with Gleason score and metastasis in clinical samples and is high in androgen-independent cell and tumor lines (227). TERT, encoding the catalytic subunit of telomerase, is responsible for maintaining telomere length and promotes

cell immortality. TERT is highly expressed in ESC (228) and CSC (229, 230) and potentially contributes to indefinite self-renewal capacity of stem-like cells. In support, PCa cells with high TERT expression possess SC gene signatures, long-term tumor-propagating capacity, and ability of symmetric self-renewal of CSCs (231).

Contrasting with genes coordinately regulated by NP8/AR/FOXA1, roughly 20% cluster 4 genes are bound by NP8 alone. GO analysis implicates most of these genes in potentially regulating cell motility, invasion, and metastasis. For example, NEDD9 is a crucial mediator of TGFβ regulated EMT and invasion via upregulating EMT TFs and effectors including BCAR1, Snail, Slug, and MMP14, as well as activating ERK signaling (232). Another example is RDX, which is required for PCa cell migration by regulating epithelial polarity via Rac1 and forming adherens junctions via VavGEF (233). Of note, the androgen-context independent nature of these 20% NANOG-upregulated cluster 4 genes may relate to the absence of AR binding to *cis*-elements.

Genes in cluster 5 are strongly upregulated in the absence of androgen, especially at the later time point (AI d22), implying that the development and growth of AI tumors promoted by NANOG may rely, greatly, on these genes. Strikingly, more than half of the cluster 5 DEGs are involved in cell cycle progression and cell division, especially M-phase cell-cycle genes highlighted by UBE2C. Previous studies have shown that AR promotes G1/S transition in androgen-dependent PCa cells in presence of androgen (234), whereas AR selectively upregulates M-phase genes and accelerates M-phase transition in AI PCa cells (140). Furthermore, histone H3K4 methylation and FOXA1 are recruited to UBE2C and CDK1 enhancers in AI PCa cells. UBE2C is an E2 ubiquitin-conjugating enzyme that works with ubiquitin ligases,

APC/C, to degrade cyclin B and terminate the M phase (235). CDK1 is a kinase that partners with cyclin A and cyclin B during the G2/M stage of cell cycle (236). Both CDK1 and UBE2C are in cluster 5, significantly upregulated in the absence of androgen and co-occupied by NANOG/FOXA1/AR on their enhancers. Other M-phase and cell cycle genes such as Bub1 and AURKB are also upregulated in cluster 5. BUB1 function as a scaffold kinase to stabilize kinetochore during spindle assembly, and BUB1 coordinates spindle checkpoint signaling and response (237). AURKB participates in multiple functions during cytokinesis and in mitotic control during G2/M transition, via regulation of chromatin posttranslational modifications, and microtubule-kinetochore attachment and separation (238).

Surprisingly, the majority of the cluster 5 genes are not bound by any combination of NP8, AR, or FOXA1. MEME analysis infers that MYC might be a potential upstream regulator of these genes. Indeed, NANOG occupies a site -10 Kb from the TSS of c-MYC and induces c-MYC expression in PCa cells (93). Additionally, the DEGs in cluster 5 are highly associated with a MYC oncogenic signature (Figure 4-6E). MYC represents a common response hub downstream of many growth and proliferation promoting signaling pathways, including WNT/β-catenin, PI3K/AKT, ERK/MAPK, and SMAD (239). In PCa, MYC expression not only promotes PIN (Prostate Intraepithelial Neoplasia) precursor lesions, but MYC amplification and overexpression are also associated with CRPC development (240-242). Therefore, NP8 may take advantage of MYC and its downstream target genes to reprogram the PCa transcriptome in the absence of androgen, both directly and indirectly. In this regard, the NANOG-induced LNCaP cell reprogramming, compared with iPSC

reprogramming by OSKM factors from somatic cells such as fibroblasts, seems to involve a 'hybrid' mode of mechanism involving one of the OSKM quartet, i.e., MYC, and prostate/PCa-specific lineage factors, i.e., AR and FOXA1. Hence, both of our initial hypotheses seem to be operational in NANOG-mediated reprogramming of PCa cells.

It is striking and also informative that NANOG mediates PCa cell reprogramming via converging on the AR/FOXA1 signaling axis. FOXA1 is a pioneer TF crucial to endoderm development, prostatic cell differentiation, prostate glandular morphogenesis, and CRPC development (243). Previous studies have demonstrated FOXA1 to be an AR co-regulator co-occupying the majority of the AR binding sites (244) and physically interacting with AR (245) to coordinately activate gene expression required for prostatic differentiation. For example, FOXA1 occupies PSA enhancers and interacts with DNA-binding domain/hinge region of AR via its forkhead domain (245). In addition to direct interactions between FOXA1 and AR, FOXA1 also promotes chromatin accessibility via its winged helix domain, leading to increased AR accessibility to the nucleosomes (246). FOXA1 has been shown to work with AR in three different ways: independent from AR, as a pioneering factor for AR (i.e., FOXA1 required for AR recruitment), and competing with AR (FOXA1 depletion is required for AR recruitment) (247). Interestingly, the stoichiometric ratio of AR and FOXA1 seems to dictate the mode of AR functions. Thus, a relatively higher level of AR results in AR binding dominantly and enforces AR target gene transcription even without androgen; equilibrium of AR and FOXA1 results in cooperation between the two factors with FOXA1 opening up FKHD regions to interact with AR; and a relatively higher level of

FOXA1 leads to opening up of the genome broadly and decreased specific AR-chromatin binding (93). In our system, the expression levels of FOXA1 and AR are not significantly different between NP8 versus control or between AD versus AI conditions (data not shown). Hence, NANOG reprograms LNCaP cells not by simply changing the expression levels and ratios of AR and FOXA1.

FOXA1/AR signaling is also associated with specific histone modifications and DNA methylation patterns. FOXA1 preferentially occupy H3K4me1 and H3K4me2 (248, 249) and we found that NANOG is enriched at H3K4me1 and H3K4me2 enhancer sites (151). H3K4me2 distribution on enhancers is cell-type specific and defines lineage-specific FOXA1 recruitment on chromatin (250). Thus, FOXA1 and NANOG may converge on H3K4me2 marks in a cell-type specific manner ultimately regulating and altering the prostate lineage-specific AR cistrome. Furthermore, the enhancer regions bound by FOXA1 appears to have lower methylated DNA content than juxtaposing genomic regions (251). These mechanisms suggest that NANOG may reprogram the AR cistrome not only by functioning as a master TF but also by partnering with FOXA1 to broadly affect the chromatin and epigenetic environment.

In conclusion, our current study indicates that NANOG redirects PCa cell fate towards undifferentiated and castration-resistant CSCs via multiple mechanisms: hijacking the pre-existent prostate lineage TFs AR and FOXA1 to reprogram the AR transcirptome, engaging one of the iPSC reprogramming factors MYC to confer cell-cycle autonomy, and, likely, modulating chromatin structure to permit dynamic gene expression changes required for full reprogramming.

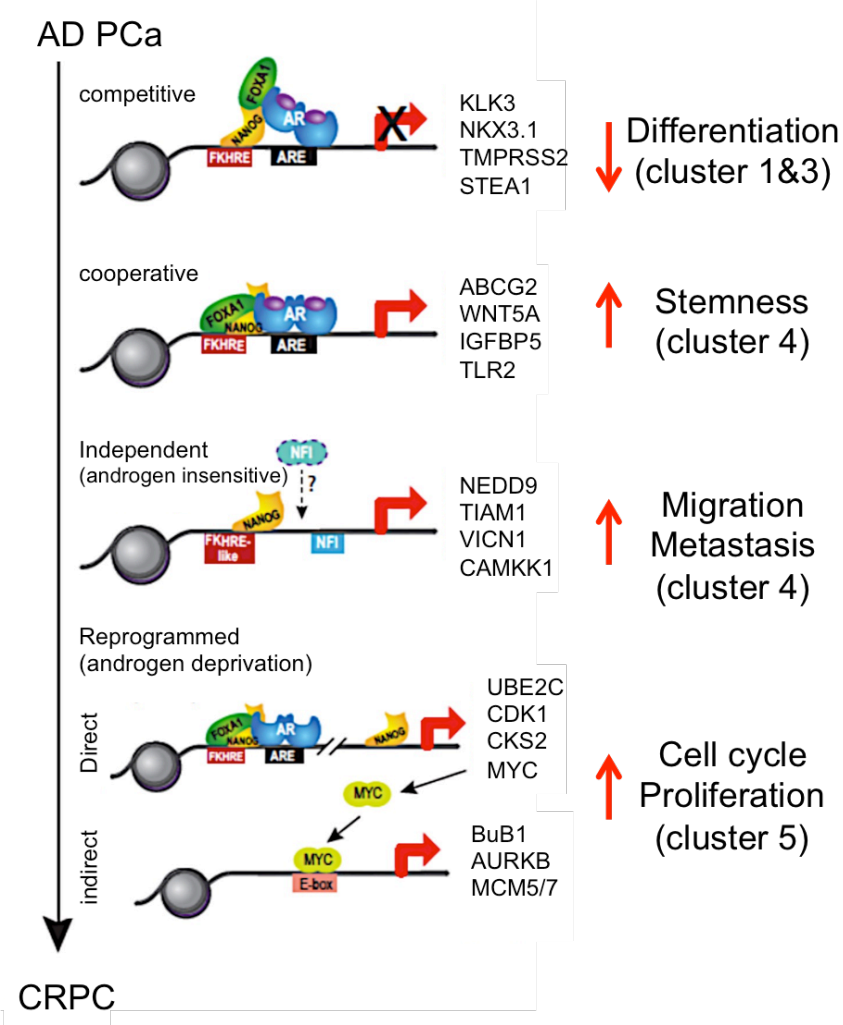


Figure 4-7. Overview of Nanog induced reprogramming model in AR^{+}PSA^{+} LNCaP cells (adapted from (151) Jeter, C. R., B. Liu, Y. Lu, H. P. Chao, D. Zhang, X. Liu, X. Chen, Q. Li, K. Rycaj, T. Calhoun-Davis, L. Yan, Q. Hu, J. Wang, J. Shen, S. Liu, and D. G. Tang. 2016. NANOG reprograms prostate cancer cells to castration resistance via dynamically repressing and engaging the AR/FOXA1 signaling axis. *Cell discovery* 2: 16041).

A hypothetical model showing that Nanog may interact with FOXA1 and/or AR and their responsive elements, forkhead responsive elements (FKHRE) or androgen responsive elements (ARE), respectively, to regulate downstream gene expression leading to reprogramming of PCa cells.

Table 4-1. Differential expression of OSKM genes in NANOGP8 overexpressing LNCaP cells in the short-term and long-term AD and AI conditions

Presented are fold changes and p-values derived from normalized read counts in LNCaP cells with or without NP8 induction under AD or AI conditions for the indicated time intervals. Note that the SOX2 level is undetectable in AD at day 5.

	AD day5		AD day12		AI day7		AI day22	
	Fold change	p-value	Fold change	p-value	Fold change	p-value	Fold change	p-value
Nanog	410.10	0	1024.16	5.80E-77	334.27	0	249.08	1.43E-130
OCT4	0.89	0.56	1.80	0.15	0.71	0.096	0.81	0.43
SOX2			0.83	0.75	0.71	0.36	1.02	0.94
KLF4	0.70	0.07	0.85	0.59	1.26	0.36	1.47	0.18
MYC	1.19	0.02	1.36	0.22	2.14	2.25E-12	1.63	8.95E-05

CHAPTER FIVE

Bioinformatics analyses uncover novel molecular mechanisms of castration resistance

5.1 Background and Preliminary Data

Due to the essential role of AR signaling in PCa development, progression, and therapy response, ADT (i.e., chemical castration) is the standard-of-care treatment for advanced and metastatic PCa as well as for recurrent tumors after radical prostatectomy and radiation therapy. The first-line ADT using gonadotropin-releasing hormone analogs (or antagonists) suppresses gonadal production of testosterone. The recurrent tumor after this initial ADT is referred to as primary CRPC. Further treatment of primary CRPC via enzalutamide (a high affinity AR antagonist, also called anti-androgen) and abiraterone aims to inhibit continued AR signaling and adrenal androgen biosynthesis, respectively. The inevitable emergence of recurrent tumor post enzalutamide and abiraterone treatment is termed secondary CRPC.

Current treatments for PCa mainly target the AR signaling pathway. However, AR heterogeneity has been observed in human PCa and is accentuated in advanced, metastatic, and relapsed PCa (25, 123, 252-254). How the heterogeneity in AR expression levels impacts PCa biology and PCa cell response to ADT and antiandrogens remains unclear. Indeed, when our laboratory analyzed AR expression by IHC (immunohistochemistry) in CRPC sections from 195 tissue microarray (TMA) cores (Figure 5-1A) and 8 whole-mount slides (Figure 5-1B) obtained from 89 patients, we observed three distinct AR expression patterns - overexpression, loss, and

subcellular redistribution. Also, when 4 AR^+ human PCa xenograft lines (i.e., LNCaP, VCaP, LAPC4, and LAPC9) are serially propagated in castrated NOD/SCID or NSG mice, we observed distinct 'evolutionary' changes in AR expression levels/patterns. Thus, castration-resistant LNCaP tumors are characterized by prominent nuclear AR ($AR^{+/hi}$ CRPC), VCaP and LAPC4-derived castration-resistant tumors display both nuclear and cytoplasmic distribution of AR (nuc/cyto-AR CRPC), and LAPC9 CRPC show marked reduction in AR levels ($AR^{-/lo}$ CRPC) (Figure 5-1C). Strikingly, the distinct AR expression in the 4 CRPC models has been liked to their different responses to second-line ADT therapeutics, enzalutamide: the $AR^{-/lo}$ LAPC9 CRPC is enzalutamide-resistant *de novo,* whereas the three AR^+ CRPCs are transiently sensitive to enzalutamide followed by emergence of enzalutamide-resistant secondary CRPC (Figure 5-1D).

As these xenograft models have different genetic backgrounds, my colleague, Dr. Qu Deng, employed a zinc finger nuclease mediated integration strategy to generate RFP-tagged homogeneously AR-positive (AR^+) LNCaP cell clones and utilized the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-cas9 system to generate AR knockout (KO) LNCaP clones (152). Using the genetically matched LNCaP cell clones, they performed side-by-side comparisons of (RFP^+)AR^+ and AR-KO cells with respect to their tumorigenic potential and responses to enzalutamide. The results showed that AR-KO LNCaP clones manifest high tumorigenicity and enzalutamide resistance upon transplantation in castrated mice, when compared to AR^+ LNCaP cells (Figure 5-1E). Interestingly, AR^+ LNCaP cells

possess competitive advantages in androgen-containing conditions *in vitro* in the presence of dihydrotestosterone (DHT) (Figure 5-1F).

5.2 Hypothesis

Undoubtedly, AR is a very critical therapeutic target for primary PCa in general. Nevertheless, AR expression is highly heterogeneous in primary tumors as well as in CRPC and there exists $AR^{-/lo}$ PCa cell population that is inherently insensitive to castration and antiandrogens(152). Of note, during progression of androgen-dependent into androgen-independent tumors, decreased PSA protein levels are observed in all 4 CRPC models regardless of changes in AR expression (Figure 5-1G), indicating decreased differentiation status and increased $PSA^{-/lo}$ PCa cells and 'stemness' in the experimental CRPC models.

These preliminary observations led to my overarching hypothesis that ADT and enzalutamide treatment represent a reprogramming process that results in enrichment of $PSA^{-/lo}$ PCSCs and increased stemness in CRPC regardless of evolution of the AR status. I further hypothesize that *ADT (castration) induced reprogramming of androgen-dependent AR^{+} PCa to the castration-resistant state, in both $AR^{+/hi}$ (e.g., the LNCaP type) and $AR^{-/lo}$ (e.g., the LAPC9 type) CRPCs, involves common enrichment of gene expression profiles known to be associated with, for example, castration, stemness, and neurogenesis. On the other hand, I also hypothesize that the $AR^{+/hi}$ or $AR^{-/lo}$ CRPCs induced by ADT may possess unique transcriptomes that facilitate their individual evolutionary trajectories.* In order to dissect the overlapping and distinct molecular circuitry of $AR^{+/hi}$ vs. $AR^{-/lo}$ CRPC development, I interrogated the RNA-Seq transcriptional profiles of xenografts in intact hosts (AD), and primary CRPC and

secondary CRPC in castrated hosts, which recapitulate PCa development and disease progression in patients. Analysis of these clinically relevant $AR^{+/hi}$ vs. $AR^{-/lo}$ CRPC models sheds light on the possible mechanisms by which these tumors progress to castration resistance, and provides potentially novel therapeutic targets.

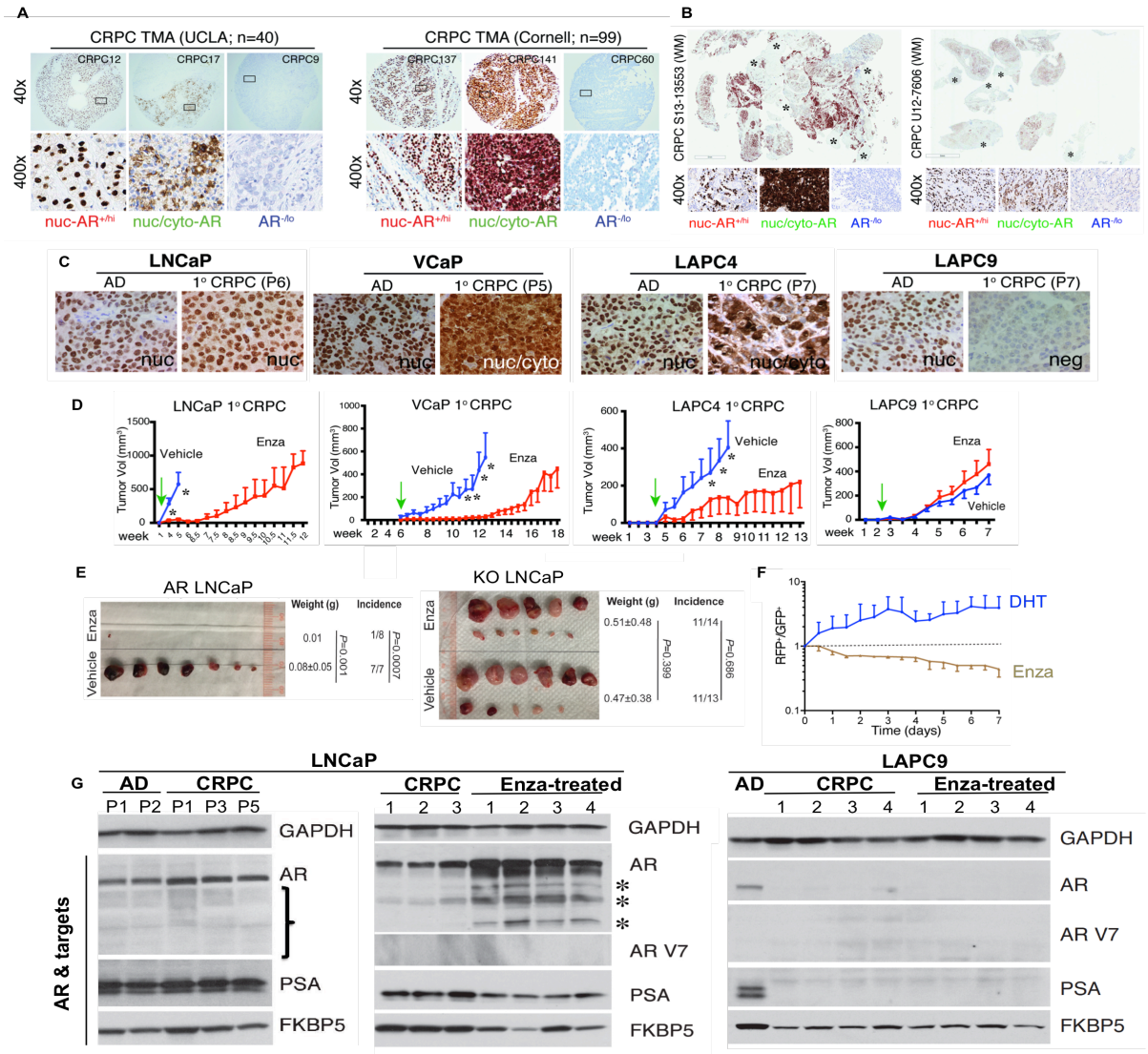


Figure 5-1. AR expression heterogeneity in CRPC and its impact on enzalutamide response (adapted from (152) Li Q, D. Q., Chao HP, Liu X, Lu Y, Lin K, Liu B, Tang GW, Zhang D, Tracz A, Jeter C, Rycaj K, Calhoun-Davis T, Huang J, Rubin MA, Beltran H, Shen J, Chatta G, Puzanov I, Mohler J, Wang J, Zhao R, Kirk J, Chen X, and Tang DG. 2018. Linking prostate cancer cell AR heterogeneity to distinct castration and Enzalutamide responses. *Nature communications*.) (performed by Dr. Qiuhui Li and Dr. Xin Chen)

A. IHC images of AR protein expression, representative of a total of 195 CRPC cores analyzed from three TMAs.

B. Representative IHC images of AR protein expression in whole-mount sections from two CRPC patients. The * indicates the $AR^{-/lo}$ areas.

C. Representative IHC images of AR protein expression in four xenograft models, LNCaP, VCaP, LAPC4, and LAPC9.

D. Therapeutic responses of CRPC to enzalutamide in four xenograft models. The green arrows indicate the enzalutamide treatment starting time and the * indicates the statistical significance with p-value < 0.01.

E. AR^{+} LNCaP cells are sensitive whereas AR-KO LNCaP cells are resistant to enzalutamide. Shown are endpoint tumor images with tumor weight, incidence and p-values (determined using unpaired Student's *t*-test and Chi-squared test for tumor weight and incidence, respectively) indicated.

F. Competitive advantages of AR^{+} and AR^{-}KO LNCaP cells in DHT vs. enzalutamide containing media. Shown is the ratio of RFP^{+} (AR^{+}) over GFP^{+} (KO, AR knockout) LNCaP cells.

G. Decreased PSA protein levels in CRPC models. Shown is the WB analysis of AR, AR-V7 and its targets in androgen-dependent (AD), castration-resistant prostate cancer (CRPC), and enzalutamide-resistant LNCaP and LAPC9 models. P1 - P5 refer to passage numbers. The numbers in other panels refer to individual tumors. The * indicates the potential AR variants. Note that in the LNCaP AD ➔ primary CRPC, PSA protein levels did not significantly change on WB (left). However, since PSA is a secreted protein, we examined PSA expression in tumor cells in immunofluorescence and confocal microscopy and results showed reduced intra-tumor cell PSA and increased $PSA^{-/lo}$ LNCaP cells in primary CRPC (123).

5.3 Materials and Methods

Biological models

Xenograft models of AD tumors were maintained by injecting purified single cells into immunodeficient NSG and NOD/SCID mice for LNCaP and LAPC9, respectively. Primary CRPCs were established by subcutaneous (s.c.) injection of purified single parental AD cells mixed with Matrigel and serially passaged in surgically castrated immunodeficient mice. Secondary CRPCs were generated by s.c. injection of primary CRPC cells into castrated immunodeficient mice concurrent with enzalutamide treatment (Figure 5-2). In total, four biological replicates of LNCaP (AD, primary, and secondary) tumors and five of LAPC9 (AD and CRPC) tumors were harvested and utilized in RNA-Seq analysis. Note that since the $AR^{-/lo}$ LAPC9 CRPC did not respond to enzalutamide treatment (Figure 5-1D), we did not harvest any enzalutamide-treated tumors for RNA-Seq analysis.

Tumor harvest and RNA-Seq

After tumors were harvested, total RNA was extracted by RNeasy mini kit (Qiagen) and genomic DNA was depleted by on-column DNA digestion. RNA quality was checked by bioRobot and 1 □g of total RNA was used to prepare cDNA libraries by Illumina TruSeq Stranded Total RNA LT Sample Prep kit following the manufacturer's guidelines. The libraries were ligated with adaptors and then sequenced by 2x75 bases paired-end on HiSeq 2000 instrument (Illumina, San Diego, CA). 75-98 and 47-70 million pairs of read were produced per sample in LNCaP and

LAPC9 samples, respectively. Each pair of reads represents a cDNA fragment from the library.

Bioinformatics analysis

All reads were quality checked by FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The reads were mapped to the reference human genome sequence (hg38) by TopHat (version 2.0.10) (209), which implements Bowtie (version 2.1.0) (210) with parameters "-r 50 --mate-std-dev 50 --library-type fr-unstranded –G", as mentioned in Chapter 2. More than 80% fragments were uniquely mapped to the human genome. The number of fragments located on all 60,119 known genes from GENCODE Release 21 (255) was enumerated by htseq-count from HTSeq package (version 0.6.1) (155) with parameters "-f bam -m union -s reverse". Genes with fewer than 10 fragments in all the samples were removed for differential expression analysis, resulting in a total of 23,933 genes for LNCaP and 20,842 genes for LAPC9 models. DEseq (version 1.14.0) (157) was applied for calling Differentially Expressed Genes (DEGs).

The DEGs were the genes passing statistical cutoff of fold change (FC) ≥ 2 and the FDR < 0.05. In LNCaP, there were 2,451 DEGs in primary CRPC over AD and 3,254 in secondary over AD. Also, there were 601 DEGs in secondary over primary LNCaP CRPCs by setting cutoff at FC ≥ 1.5 and FDR < 0.05. In LAPC9, there were 3,929 DEGs in (primary) CRPC over AD. The Venn diagram, heatmaps, and landscape profiles were generated as discussed in Chapter 2. For IPA, the DEGs with FC ≥ 2 and FDR < 0.05 were run for the Core Analysis. For GSEA, the normalized

read count of all genes was included to compare to the gene list enriched in multiple clinical and experimental datasets, including those in PCa patient tumors before versus after ADT (referred to as before ADT vs. after ADT, (175)), AD versus AI LNCaP cells (181), non-recurrent versus recurrent PCa (Rec vs. non-Rec, (256)), in-house PSA^{+} versus $PSA^{-/lo}$ (also see Chapter 3), normal prostate basal/stem cells versus luminal cells (112), and CRPC adenocarcinoma versus neuroendocrine-like CRPC (i.e., CRPC-Adeno vs. CRPC-NE, (257)). The curated gene set C2 of MSigDB was used for GO term enrichment.

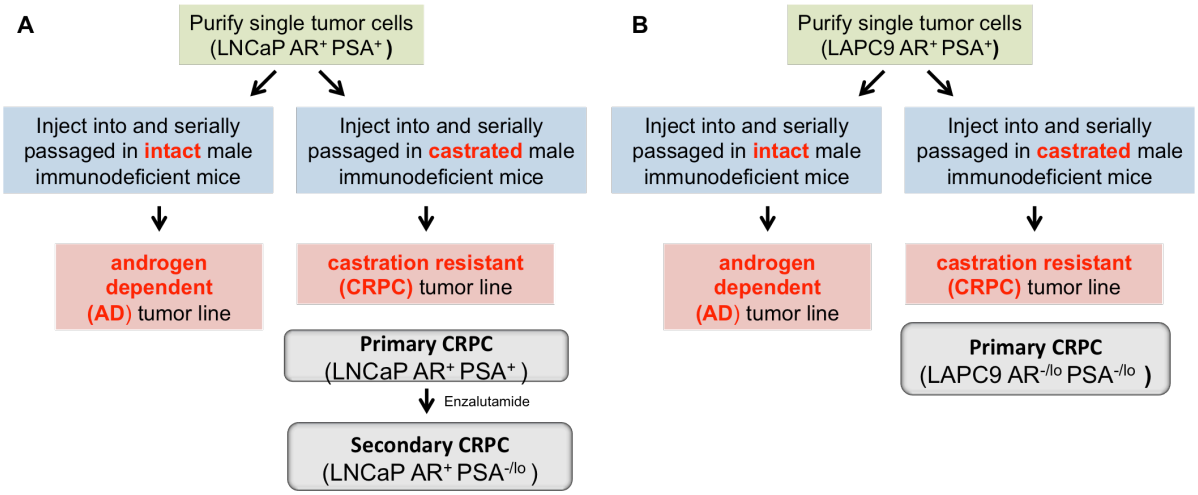


Figure 5-2. Experimental scheme of RNA-Seq in the xenograft models (adapted from (152) Li Q, D. Q., Chao HP, Liu X, Lu Y, Lin K, Liu B, Tang GW, Zhang D, Tracz A, Jeter C, Rycaj K, Calhoun-Davis T, Huang J, Rubin MA, Beltran H, Shen J, Chatta G, Puzanov I, Mohler J, Wang J, Zhao R, Kirk J, Chen X, and Tang DG. 2018. Linking prostate cancer cell AR heterogeneity to distinct castration and Enzalutamide responses. *Nature communications*.).

A. Scheme of generating $AR^{+/hi}$ LNCaP CRPC model. The xenograft tumor line generated in intact NSG mice with purified parent LNCaP cells is designated AD tumors and, in castrated mice, primary AI tumors. The primary AI tumors treated with, and subsequently became resistant to, enzalutamide are designated secondary AI tumors.

B. Scheme of generating $AR^{-/lo}$ LAPC9 CRPC. The xenograft tumor line generated in intact NOD/SCID mice with parent LAPC9 cells is designated AD tumors and, in castrated mice, AI tumors (i.e., CRPC). Note that primary $AR^{-/lo}$ AI LAPC9 tumors did not respond to enzalutamide treatment (see Figure 5-1D).

5.4 Results

To dissect the transcriptomes of different LNCaP tumors at various stages of CRPC reprogramming, I performed several types of analysis including: characterizing the global transcriptome features, annotating biological functions or pathways of gene expression profiles, and conducting more detailed analysis of target genes. First, I applied *the pair correlation analysis* to examine the correlation between samples and performed unsupervised clustering to investigate the similarity of the samples. The correlation coefficients between samples within the same condition were higher than 0.90, indicating consistency of the biological replicates (Figure 5-3A). Interestingly, the correlation coefficients of primary LNCaP CRPC over AD (0.89-0.91) were higher than those in secondary LNCaP CRPC over AD (0.84-0.89), suggesting that the global transcriptome profiles gradually diverged along the course of castration and enzalutamide treatment (Figure 5-3A). Strikingly, the correlation coefficients of secondary LNCaP CRPC over primary CRPC were very high (0.92-0.97), suggesting relatively minor changes in transcriptomes during transition from ADT (castration) resistance to enzalutamide resistance (Figure 5-3A). Both unsupervised clustering and the pair correlation analyses revealed a concordance among biological replicates. Furthermore, the global transcriptome features of AD tumors were distinct from those in CRPC, but transcriptome features of primary CRPC and secondary CRPC were very similar (Figure 5-3A and B).

To further analyze transcriptome variations in our biological samples, I performed PCA, which applies linear transformation to distinguish datasets according to variance. The first principal component successfully separated the AD and CRPC

groups and accounted for more than 30% variance of the total, whereas the second principal component separated the primary and secondary CRPC and underlie 18% of the variance (Figure 5-3C). PCA indicated that the consistency of biological replicates and transcriptome profiles gradually changed from AD to primary CRPC and secondary CRPC. Specifically, there were 2,451 DEGs in primary LNCaP CRPC over AD and 3,254 DEGs in secondary CRPC over AD, when using a cutoff at FC ≥ 2 and FDR < 0.05. When using the same cutoff parameters, we only observed slightly >100 DEGs between secondary vs. primary CRPC (data not shown), again suggesting that differences in transcriptomes are rather subtle when comparing enzalutamide-resistant (secondary) vs. ADT-resistant (primary) LNCaP CRPC. This is not particularly surprising as both ADT and enzalutamide target the AR signaling. On the other hand, we did observe 601 DEGs in secondary over primary LNCaP CRPCs when using the cutoff of FC ≥ 1.5 and FDR < 0.05. In total, 2,033 genes were shared between primary CRPC over AD and secondary CRPC over AD, including 1,437 upregulated and 596 downregulated genes (Figure 5-3D).

To help interpret the biological meaning of the transcriptomic changes during AD castration ➔ resistance □□□□□□□□□□□□ ➔ resistance transitions, functional annotations and comparisons were performed. GSEA, by comparing our xenograft CRPC transcriptomes with published datasets, suggests that the overall gene expression profiles of both primary and secondary LNCaP CRPCs are enriched in genes associated with ADT resistance and neuroendocrine PCa (i.e., CRPC-NE) phenotype in patients (Figure 5-4A). CRPC-NE is an aggressive form of CRPC with high epithelial plasticity, heterogeneous clinical features and poor prognosis. The

canonical pathway analysis of IPA revealed that the DEGs preferentially expressed in both primary and secondary LNCaP CRPC over AD were involved in Stem Cell Signaling, Lipid Signaling and Neurogenesis (Figure 5-4B left and middle panel). Stem cell signaling is enriched in not only secondary CRPC over AD but also in secondary over primary CRPC, highlighted by many genes in the STAT3, IGF-1, Wnt/β-catenin, TGFβ, and Notch signaling pathways (Figure 5-4B middle and right panel). Consistently, the DEGs of both primary CRPC over AD and secondary CRPC over AD are highly associated with the gene signatures in $PSA^{-/lo}$ PCSCs as well as normal human prostate basal/stem cells (Figure 5-4C).

In the LAPC9 model, the overall gene expression profiles in CRPC compared with the AD tumors were also enriched in genes that have been associated with ADT resistance and recurrence in clinical CRPC (Figure 5-5A). In common with the LNCaP CRPC gene expression profiles, the 3,929 DEGs in the LAPC9 CRPC over AD (FC ≥ 2; FDR < 0.05) were also enriched in lipid signaling and neurogenesis pathways (Figure 5-5B). Of note, the axonal guidance signaling, serotonin degradation pathway, phospholipase signaling, eicosanoid signaling, and FXR/RXR signaling were all enriched in both LNCaP and LAPC9 CRPC models.

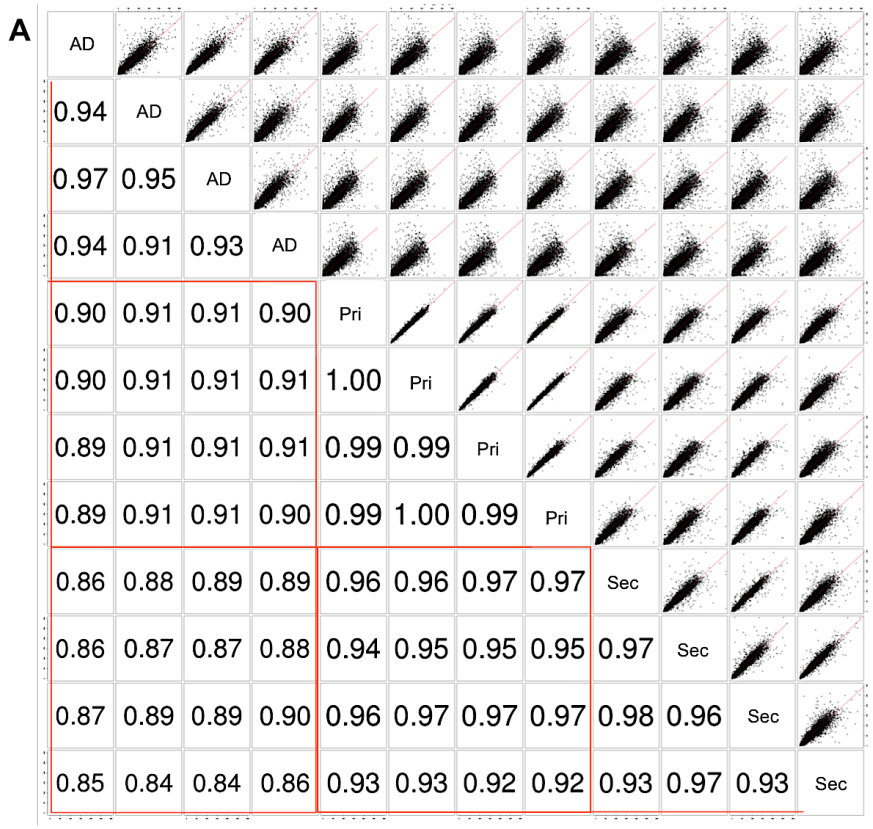
Paralleling IHC results, *AR* mRNA expression was progressively upregulated along the course of castration and enzalutamide treatment in the LNCaP progression model, but much reduced *AR* mRNA levels were observed in LAPC9 CRPC (Figure 5-6A and E). Strikingly, though, the two AR downstream target genes, *KLK3* and *FKBP5*, were decreased at the mRNA levels in both LNCaP and LAPC9 CRPC with *KLK3* barely detectable in LAPC9 CRPC (Figure 5-6B and E). This latter observation

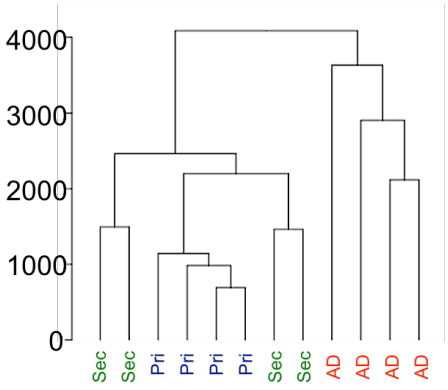
again indicates that the CRPC, regardless of the AR status, loses differentiation and becomes enriched in phenotypically undifferentiated $PSA^{-/lo}$ CSCs. In support, among the CSC genes examined including *ALDH7A, BCL-2, CDH2, ITGA2, MYC*, and *STAT3*, *BCL-2* was prominently and significantly upregulated in both primary and, particularly, secondary LNCaP CRPC, as well as in LAPC9 CRPC (Figure 5-6C and D). These results suggest that BCL-2 may be a pivotal factor that mediates castration and/or enzalutamide resistance in both $AR^{+/hi}$ and $AR^{-/lo}$ CRPC. In addition, BCL-2 was the only BCL family member found to be selectively upregulated in clinical CRPC (258). Thus, BCL-2 may be a common and critical therapeutic target for enzalutamide-resistant CRPC.

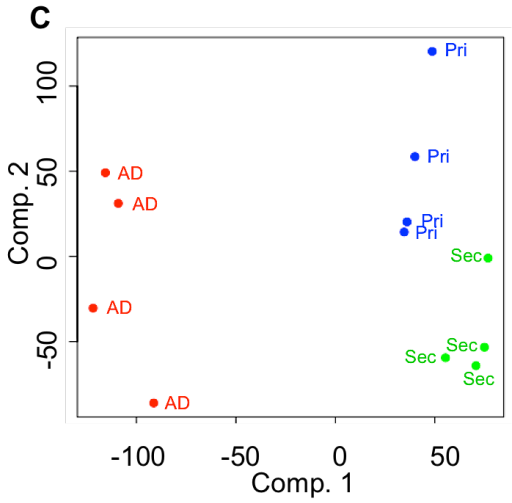
Among the CSC related target genes, BCL-2 is upregulated in both $AR^{+/hi}$ and $AR^{-/lo}$ CRPC, particularly associated with enzalutamide treatment (Figure 5-6C and D). BCL-2 is a pro-survival molecule regulating apoptosis. However, in both normal and transformed cells BCL-2 maintains viability by preventing cell death but does not affect proliferation. BCL-2-mediated premalignant cell survival benefit permits acquisition of additional oncogenic lesions such as MYC overexpression leading to tumorigeneisis (259). Furthermore, BCL-2 mRNA levels have been shown to increase in CRPC patients (175). These important findings lead my colleagues to perform combinatorial therapeutic experiments by treating mice harboring either $AR^{+/hi}$ or $AR^{-/lo}$ CRPC tumors with combination treatment of BCL-2 specific inhibitor, ABT-199, and other regimens. We found that combination of enzalutamide with ABT-199, but not GR (Glucocorticoid Receptor) antagonist, RU486, dramatically and significantly ($P< 0.05$) inhibited the $AR^{+/hi}$ CRPC tumor incidence (Figure 5-7A). ABT-199 alone inhibited AR^{-}

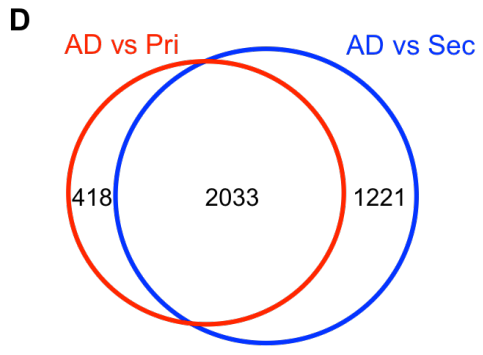
$^{/lo}$ CRPC growth and organoid expansion, and enzalutamide did not affect AR$^{-/lo}$ CRPC (Figure 5-7B and C). However, combination of ABT-199 with JQ1, BET (Bromodomain and Extra-Terminal motif) inhibitor, further suppressed the tumor growth (Figure 5-7B). JQ1 is known to transcriptionally downregulate Myc and its target genes and significantly attenuate MYC transcriptional program (260). Thus, AR$^{-/lo}$ CRPC might undergo MYC-assisted reprogramming (somewhat like NANOG-mediated LNCaP cell reprogramming; Chapter 4). In short, the results obtained with our xenograft models demonstrate that combination regimens with enzalutamide and inhibitors that target CSC signaling may be critical for the prevention and treatment of CRPC.

Of note, the AR$^{-/lo}$ LAPC9 CRPC is resistant to AR antagonist *de novo*, and, in fact, treating AR$^{-/lo}$ CRPC with ADT may do more harm than good (261). Interesting, however, the AR$^{-/lo}$ CRPC is sensitive to ABT-199 with JQ1 combination regime, whereas the GR inhibitors might be efficacious for AR^{+} CRPC because GR has been suggested to play a role in enzalutamide and docetaxel resistance (262). Thus, it is vital to tailor the combinatorial treatment strategies for CRPC patients to fit the precise tumor phenotype.


A
AD
0.94 AD
0.97 0.95 AD
0.94 0.91 0.93 AD
0.90 0.91 0.91 0.90 Pri
0.90 0.91 0.91 0.91 1.00 Pri
0.89 0.91 0.91 0.91 0.99 0.99 Pri
0.89 0.91 0.91 0.90 0.99 1.00 0.99 Pri
0.86 0.88 0.89 0.89 0.96 0.96 0.97 0.97 Sec
0.86 0.87 0.87 0.88 0.94 0.95 0.95 0.95 0.97 Sec
0.87 0.89 0.89 0.90 0.96 0.97 0.97 0.97 0.98 0.96 Sec
0.85 0.84 0.84 0.86 0.93 0.93 0.92 0.92 0.93 0.97 0.93 Sec


B
4000
3000
2000
1000
0
Sec Sec Pri Pri Pri Pri Sec Sec AD AD AD AD


C
Comp. 2
100
50
0
-50
Pri
Pri
AD
AD
Pri
Pri
Sec
AD
Sec
Sec
Sec
AD
-100
-50
0
50
Comp. 1


D
AD vs Pri
AD vs Sec
418
2033
1221

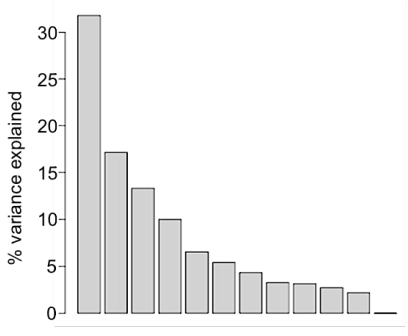

% variance explained
30
25
20
15
10
5
0

Figure 5-3. Global transcriptome changes in LNCaP CRPC model (adapted from (152) Li Q, D. Q., Chao HP, Liu X, Lu Y, Lin K, Liu B, Tang GW, Zhang D, Tracz A, Jeter C, Rycaj K, Calhoun-Davis T, Huang J, Rubin MA, Beltran H, Shen J, Chatta G, Puzanov I, Mohler J, Wang J, Zhao R, Kirk J, Chen X, and Tang DG. 2018. Linking prostate cancer cell AR heterogeneity to distinct castration and Enzalutamide responses. *Nature communications*.).

A. Pair correlation plots for all biological samples. The red line marks the biological replicates.
B. Unsupervised clustering of all experimental conditions. The results show good separation of AD tumors from AI tumors, and primary and secondary AI tumors are closer to each other transcriptomically. The cpm values with the top 5% most abundant genes across all the samples were removed, and unsupervised clustering was performed with Euclidean distance and complete clustering method.
C. PCA plots presentation of relatedness of all biological samples. The upper was the main PCA plots and the bottom bar chart was the eigenvalues of each component over sum that means the percentage of variance explained by each component.
D. Venn diagram showing the overlapping DEGs between primary and second CRPCs over AD (using the statistic cutoff of FC ≥ 2 and FDR < 0.05).

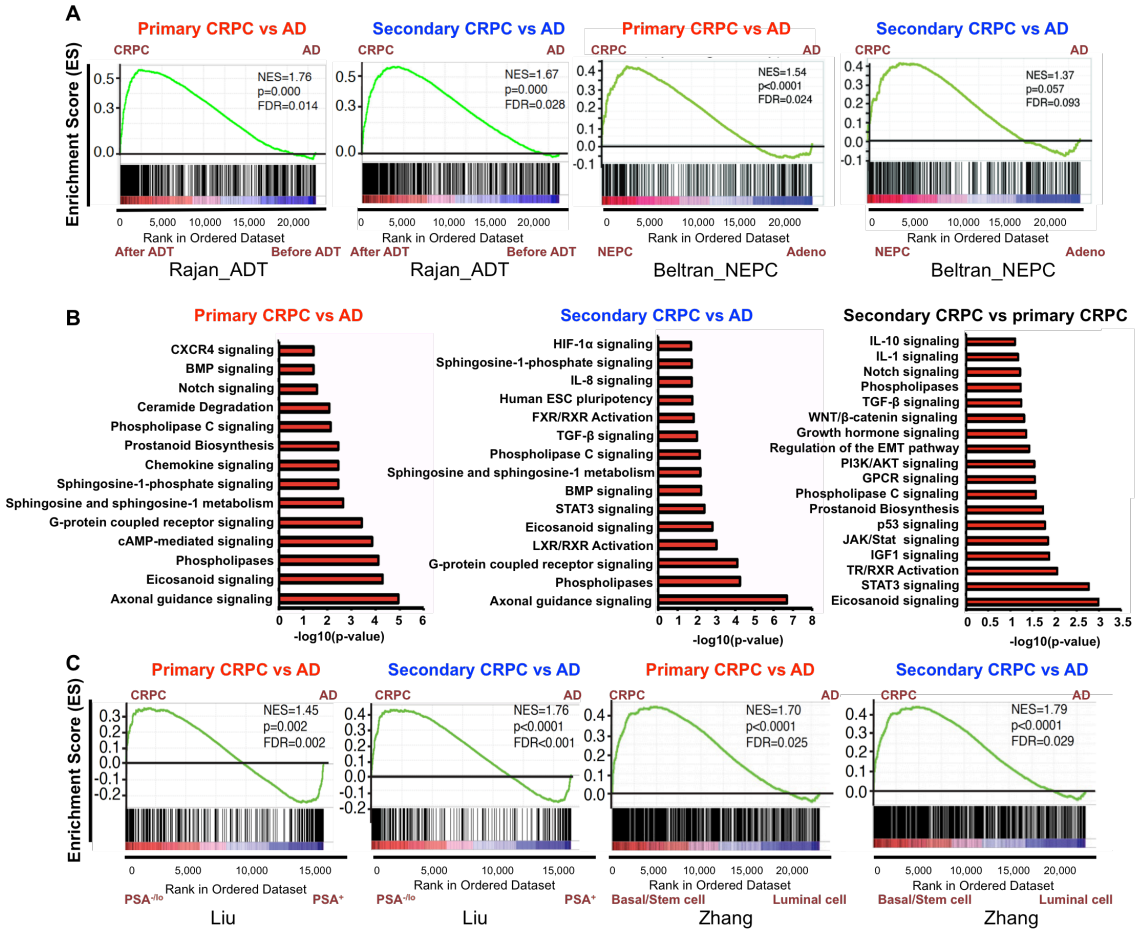


Figure 5-4. Functional annotation of the AR$^{+/hi}$ LNCaP CRPC transcriptome (adapted from (152) Li Q, D. Q., Chao HP, Liu X, Lu Y, Lin K, Liu B, Tang GW, Zhang D, Tracz A, Jeter C, Rycaj K, Calhoun-Davis T, Huang J, Rubin MA, Beltran H, Shen J, Chatta G, Puzanov I, Mohler J, Wang J, Zhao R, Kirk J, Chen X, and Tang DG. 2018. Linking prostate cancer cell AR heterogeneity to distinct castration and Enzalutamide responses. *Nature communications*.)).

A. The gene expression profile in LNCaP CRPC positively associates with gene sets in ADT-resistant patient CRPC and patient CRPC-NE. All the DEGs were used in GSEA against the gene set of Rajan dataset (175) (left) and Beltran dataset (263) (right).

B. Functional annotation of the transcriptome profiles of LNCaP AD tumors and CRPC. Presented are the top pathways enriched in primary CRPC over AD (left), and in secondary CRPC over AD (middle panel), in secondary over primary CRPC (right panel). The statistical cutoff of DEGs was FC ≥ 2 and FDR < 0.05. The right panel was the DEGs in secondary CRPC over primary CRPC with the cutoff at FC ≥ 1.5 and FDR < 0.05.

C. GSEA plots showing that the LNCaP CRPC is enriched in stem cell-associated gene signatures. The left two figures were comparison between the top 700 DEGs from primary CRPC over AD and secondary CRPC over AD to $PSA^{-/lo}$ stem cell gene signatures (*manuscript in preparation*). The right panel was comparison to the neurogenesis gene profile derived from normal prostate basal/stem cells (112).

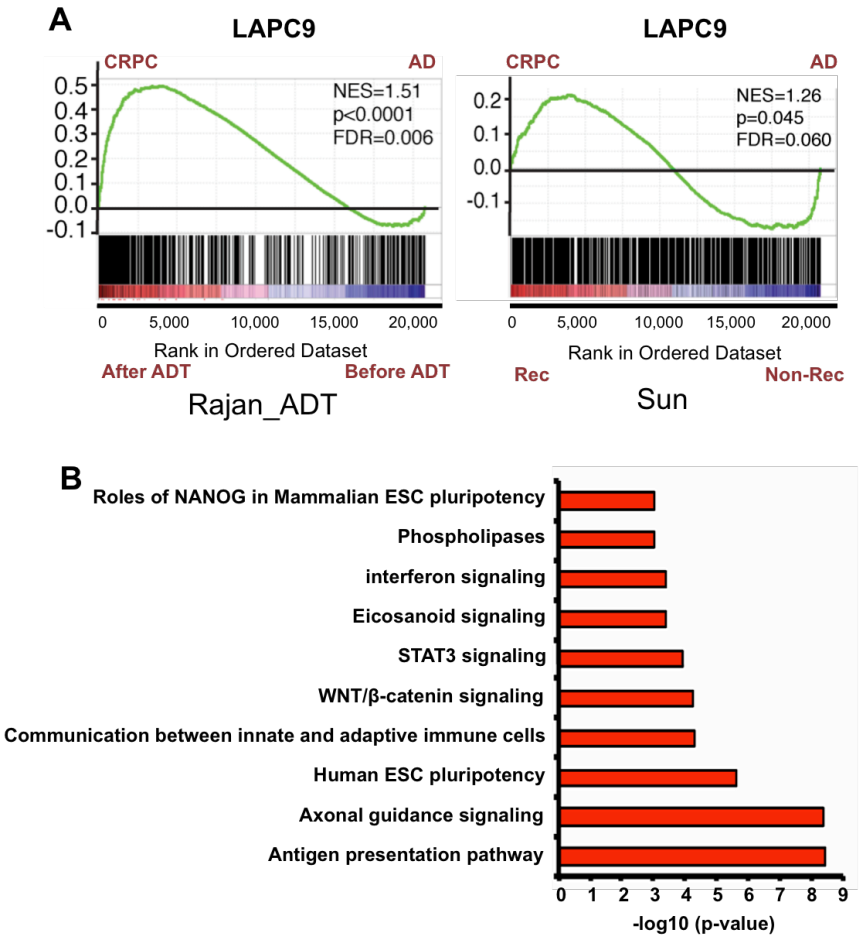


Figure 5-5. Functional annotation of the transcriptome of AR$^{-/lo}$ LAPC9 CRPC (adapted from (152) Li Q, D. Q., Chao HP, Liu X, Lu Y, Lin K, Liu B, Tang GW, Zhang D, Tracz A, Jeter C, Rycaj K, Calhoun-Davis T, Huang J, Rubin MA, Beltran H, Shen J, Chatta G, Puzanov I, Mohler J, Wang J, Zhao R, Kirk J, Chen X, and Tang DG. 2018. Linking prostate cancer cell AR heterogeneity to distinct castration and Enzalutamide responses. *Nature communications*.).

A. The LAPC9 gene expression profiles are enriched in castration-resistant patient tumors in the Rajan (175) (left) and Sun (256) (right) datasets.

B. IPA biological function profiles of the DEGs (using the statistic cutoff FC ≥ 2 and FDR < 0.05) in the AR$^{-/lo}$ LAPC9 CRPC.

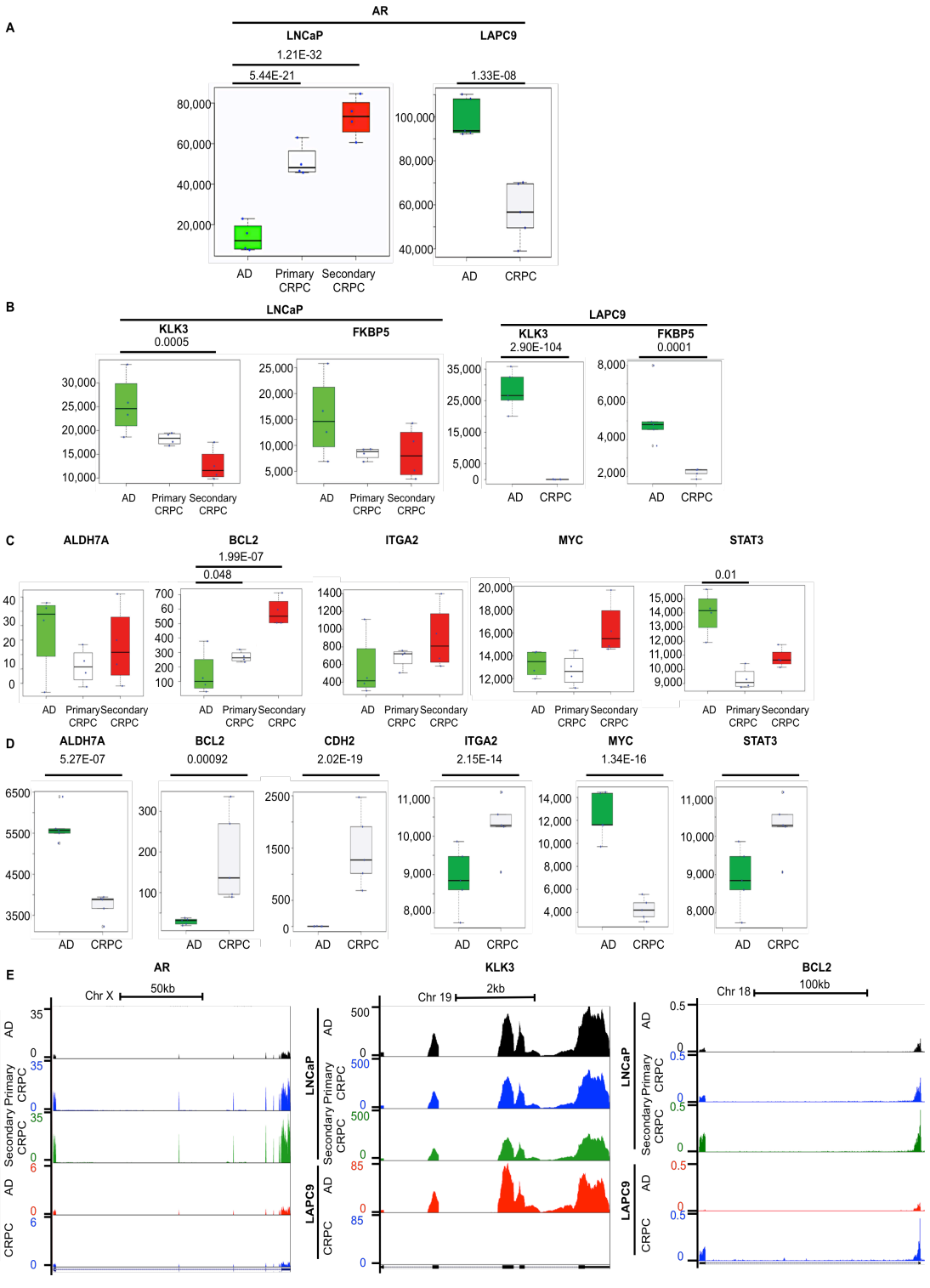

A
AR
LNCaP
LAPC9
1.21E-32
5.44E-21
1.33E-08
AD
Primary CRPC
Secondary CRPC
CRPC
B
LNCaP
LAPC9
KLK3
0.0005
FKBP5
KLK3
2.90E-104
FKBP5
0.0001
C
ALDH7A
BCL2
1.99E-07
0.048
ITGA2
MYC
STAT3
0.01
D
ALDH7A
5.27E-07
BCL2
0.00092
CDH2
2.02E-19
ITGA2
2.15E-14
MYC
1.34E-16
STAT3
E
AR
Chr X
50kb
KLK3
Chr 19
2kb
BCL2
Chr 18
100kb
LNCaP
LAPC9
AD
Primary CRPC
Secondary CRPC
CRPC

Figure 5-6. Examples of DEGs in $AR^{+/hi}$ LNCaP and $AR^{-/lo}$ LAPC9 CRPC models (adapted from (152) Li Q, D. Q., Chao HP, Liu X, Lu Y, Lin K, Liu B, Tang GW, Zhang D, Tracz A, Jeter C, Rycaj K, Calhoun-Davis T, Huang J, Rubin MA, Beltran H, Shen J, Chatta G, Puzanov I, Mohler J, Wang J, Zhao R, Kirk J, Chen X, and Tang DG. 2018. Linking prostate cancer cell AR heterogeneity to distinct castration and Enzalutamide responses. *Nature communications*.).

A-D. Box plots showing the mRNA expression levels of AR (A), several AR target genes (B), and representative CSC associated genes (C and D). Statistically significant FDR values were indicated in the figures. The y-axis of the box plots was normalized counts estimated by DESeq.

E. RNA-Seq tracks of *AR, KLK3*, and *BCL-2* genes in $AR^{+/hi}$ and $AR^{-/lo}$ CRPC models. The tracks were displayed in UCSC genome browser and the maximum height was adjusted by the highest expression of each gene in each tumor type.

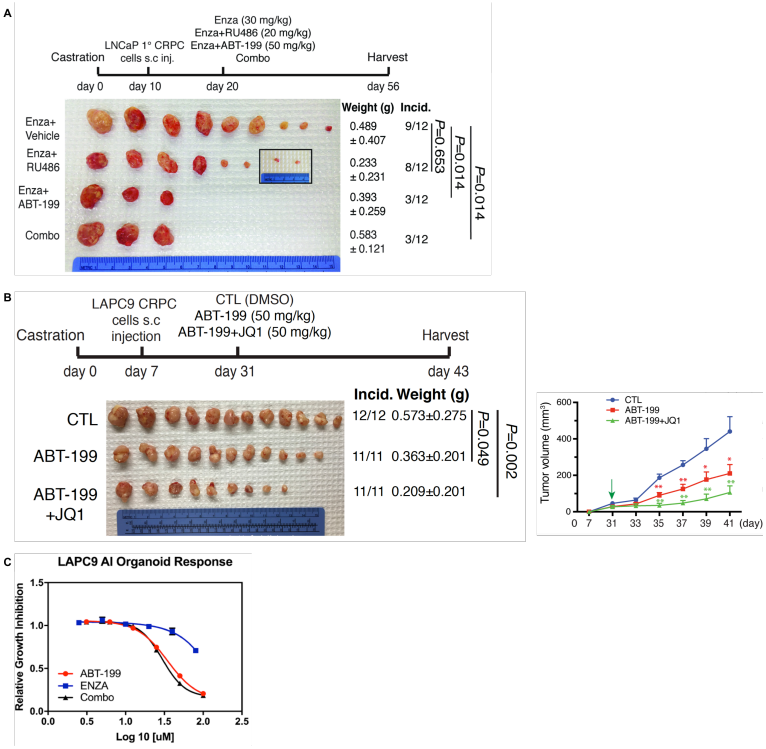


Figure 5-7. The combination treatment for AR^{+} and AR$^{-/lo}$ CRPC (adapted from (152) Li Q, D. Q., Chao HP, Liu X, Lu Y, Lin K, Liu B, Tang GW, Zhang D, Tracz A, Jeter C, Rycaj K, Calhoun-Davis T, Huang J, Rubin MA, Beltran H, Shen J, Chatta G, Puzanov I, Mohler J, Wang J, Zhao R, Kirk J, Chen X, and Tang DG. 2018. Linking prostate cancer cell AR heterogeneity to distinct castration and Enzalutamide responses. *Nature communications*.).

A. Treatment schema and results in AR^{+} CRPC. The top panel shows the experimental time line of LNCaP CRPC in NSG mice. The bottom panel presents the tumor incidence and weight and χ_2 test for tumor incidence at end point.

B. Treatment schema and results of AR$^{-/lo}$ CRPC. The top panel shows the experimental time line of LAPC9 CRPC in NOD/SCID mice. The bottom panel presents the tumor incidence and weight and Student's *t*-test for weight at end point. The right panel is the growth curve of tumor.

C. The growth curve of LAPC9 organoid in response to BCL-2 inhibitor, ABT-199, enzalutamide (Enza), or both.

Discussion

AR heterogeneity pre-exists in untreated prostate tumors such that there are both AR^{+} and $AR^{-/lo}$ PCa cells. These two subpopulations of PCa cells seem to co-evolve under the pressure from ADT/antiandrogens, resulting in the $AR^{+/hi}$ and $AR^{-/lo}$ PCa cell subpopulations in patient CRPC. Strikingly, our LNCaP and LAPC9 xenograft models, during AD ➔ primary CRPC (ADT resistance) ➔ secondary CRPC (enzalutamide resistance) progression, recapitulate this AR heterogeneity. Thus, castration gradually reprograms AR^{+} LNCaP AD tumors to the $AR^{+/hi}$ CRPC phenotype whereas castration in AR^{+} LAPC9 leads to the $AR^{-/lo}$ CRPC phenotype. Significantly, the distinct AR phenotypes in the two CRPC models are directly linked to their contrasting enzalutamide responses: while the $AR^{+/hi}$ LNCaP primary CRPC remains transiently sensitive to enzalutamide, the $AR^{-/lo}$ LAPC9 CRPC is enzalutamide resistant *de novo*. I took advantage of these two unique xenograft CRPC models and interrogated the transcriptomes during their progression to castration and/or enzalutamide resistance. My analysis overall corroborated my hypothesis and shows that castration-induced reprogramming of androgen-dependent AR^{+} PCa to the castration-resistant state, in both $AR^{+/hi}$ and $AR^{-/lo}$ CRPCs, involves common enrichment of gene expression profiles known to be associated with castration, stemness, and neurogenesis. On the other hand, castration-induced $AR^{+/hi}$ or $AR^{-/lo}$ CRPCs are also characterized with unique transcriptomes that facilitate their individual evolutionary trajectories.

GSEA reveals that gene signatures enriched in LNCaP primary $AR^{+/hi}$, LNCaP secondary $AR^{+/hi}$, and LAPC9 $AR^{-/lo}$ CRPCs are associated with those enriched in

patient CRPC that failed castration and antiandrogen treatment (175). Considering that both $AR^{+/hi}$ and $AR^{-/lo}$ xenograft CRPC gene expression profiles correlate similarly (NES are 1.76 and 1.67 and FDRs are 0.014 and 0.028, respectively) with the patient CRPC gene signatures, these data further demonstrates that both CRPC subtypes are present in patient CRPC, as validated by IHC analysis (Figure 5-1). Overall, bioinformatics analysis indicates that our experimental CRPC models, regardless of the AR expression status, preferentially and commonly express genes known to be involved in mediating castration resistance.

The second gene category commonly enriched in both $AR^{+/hi}$ LNCaP and $AR^{-/lo}$ LAPC9 CRPC is the "Stem Cell Signaling", including terms such as 'Human ESC Pluripotency' and 'STAT3 Signaling' pathways. Some of these stem cell signaling pathways are particularly enriched in enzalutamide-resistant $AR^{+/hi}$ secondary LNCaP and $AR^{-/lo}$ LAPC9 CRPC. Included in 'stemness genes' are, among many others, BCL-2, SMAD6, TDGF1, WNT8B, BMP5, FGFR3, BMP6, WNT5A, PDGFRB, and SALL4. For example, the secondary (enzalutamide-resistant) $AR^{+/hi}$ CRPC express high levels of *SALL4* mRNA. Sall4 is a 'pluripotency' transcription factor that directly interacts with NANOG, and, knocking down SALL4 or NANOG has been shown to alter ES cell morphology accompanied by reduced expression of SC genes and increased expression of differentiation-associated genes. (264). Although the functions of SALL4 in cancer are just emerging, its expression has been proposed as a biomarker for germ cell tumors such as embryonic carcinoma, sex cord stromal tumors, and yolk sac tumors (265). SALL4 is not abundantly expressed in PCa (266) but inhibition of SALL4 has been reported to reduce proliferation and promotes apoptosis in PCa cells via

modulating BCL-2 and BAX expression (267). SALL4 inhibition has also been shown to promote apoptosis and reduce proliferation, migration, clonogenicity, and tumorigenicity in esophageal squamous cell carcinoma cells (268). In theory, enzalutamide and other ADT drugs might reprogram PCa cells by inducing pluripotency factors such as SALL4 to mediate resistance.

Consistent with a common enrichment of stemness genes in both $AR^{+/hi}$ and $AR^{-/lo}$ CRPC, both types of CRPC have significantly reduced mRNA levels of *KLK3*, which encodes PSA, suggesting that both types of CRPC would be enriched in $PSA^{-/lo}$ PCa cells, which we had reported in IHC studies (123). We have previously demonstrated that the $PSA^{-/lo}$ cell population is enriched in PCSCs and intrinsically resistant to castration and antiandrogens (25). It remains obscure whether progression of CRPC reprograms the tumor cells to $PSA^{-/lo}$ and/or enriches the pre-existent refractory $PSA^{-/lo}$ cells. Regardless, the PSA expression level is markedly and progressively decreased and even undetectable in some CRPC patient samples. The heterogeneous expression and distribution of AR and discordant expression between AR and PSA in CRPC indicates that restored AR may reprogram castration resistant gene expression profiling. In AR^{+} CRPC, the androgen downstream signaling may be activated by other epigenetic or transcription factors hijacking the AR signaling pathway in the absence of androgen ligand. For example, FOXA1/H3K4 methyl marks (269), FOXA1/HOXB1/AR (270), and NANOG/FOXA1/AR (151) have all been shown to, independently or cooperatively, occupy regulatory element of AR target genes and reprogramming AR cistromes. In the $AR^{-/lo}$ CRPC, the androgen downstream signaling will be reactivated by alternative ways and bypassing AR. The possible mechanisms

include alternative androgen synthesis, modified AR or associated proteins, and unconventional initiation or regulation of downstream signaling cascades. The first resistance mechanism in enzalutamide-resistant CRPC is the “glucocorticoid receptor take-over”, in which glucocorticoid receptor binds to AR regulated genes owing to similarity of DNA binding domain of two nuclear receptors (271). The other resistance mechanism is unbalanced force on cell death, like deregulating autophagic cascade via activation of AMP-dependent protein kinase (AMPK) and the suppression of mammalian target of rapamycin (mTOR) (272). Also, there are enzalutamide resistance specific point mutations in AR such as the F876L mutation (273). Therefore, combination regimens may enhance efficacy by targeting key pathways in a synergistic or an additive manner (274) to prevent or tackle CRPC efficiently.

Neural/neuronal gene signatures are also highly enriched in both $PSA^{-/lo}$ CSC (see Chapter 3, *manuscript in preparation*) and PSA^{-} basal/stem cells (112). Perhaps that’s one of the reasons that the overall gene expression profiles of our xenograft CRPC (which have the $PSA^{-/lo}$ phenotype) correlate with the NE-cancer phenotype revealed by GSEA (Fig 5-4A). These proneural genes functionally regulate the stem/progenitor cell activities and the proneural differentiation *in vitro* in freshly purified human prostate basal cells (112). Particularly notable in the neurogenesis gene signature are some genes involved in regulating axonal guidance, which represents a neural development process that promotes axon growth and helps establish neuronal connections. In CRPC, several neurogenesis-associated morphogens are upregulated including WNT5A, WNT8B, BMP5, and BMP6. Interestingly, noncanonical Wnt ligand Wnt5a has been shown to be essential for

multiple processes of normal prostate gland development, including bud outgrowth, ductal elongation, branching, cell polarity, and lumenization in the mouse prostate (275). Expression of BMP6 is induced by WNT5A via PKC/NF-kB pathway, and then BMP6 activates BMP-RII and ALK2 leading to SMAD5 phosphorylation and interaction with β-catenin. This signaling pathway has been shown to facilitate CRPC development (276, 277). Semaphorin (SEMA) genes were also upregulated broadly in our CRPC models, including SEMA3A, SEMA3D, SEMA3E, and SEMA5A. SEMA3 proteins are in the extended family of VEGF molecules and function to compete for binding to neuropilin with VEGF family members. SEMA3A cleavage by MMP7 is involved in PCa metastasis by disrupting stability of SEMA3A and neuropilin-1-plexin complex and trigger downstream FAK signaling (278). SEMA3C activates multiple RTKs, EGFR, ERBB2, and MET via Plexin B1 and promotes CRPC growth and progression (279). Expression of SEMA3 family members is associated with angiogenesis in metastatic CRPC (280).

AR-mediated lipid biosynthesis has been shown to be reactivated in progression of clinical CRPC (281). Our RNA-Seq data shows that eicosanoid signaling pathway is commonly affected during the CRPC progression regardless of AR status and stage of progression. Eicosanoids are bioactive hormone-like lipid signaling molecules derived from arachidonic acid and released by phospholipase A2 (PLA2) and metabolized by cyclooxygenases (COXs) and lipoxygenases (LOXs). *AKR1C3, PLA2G7*, and *PTGS1* are the common eicosanoid pathway genes upregulated in both $AR^{+/hi}$ and $AR^{-/lo}$ CRPC. AKR1C3 is known as Aldo-keto reductase family 1 member C3 and plays pivotal roles in steroidogenesis, especially for

androgen biosynthesis in the human prostate (282). More importantly, AKR1C3 is overexpressed in response to ADT and contributes to CRPC (283, 284). PLA2G7 is known as phospholipase 2 group VII in PLA2 family and mediates arachidonic acid release from the plasma membrane (285). PLA2G7 suppresses apoptosis and induces tumorigenesis, cell adhesion, migration, and invasion, especially in ERG-positive PCa (285, 286). PTGS1 encodes COX-1, which catalyzes the conversion of arachidonate to prostaglandins. COX-1 expression and functions have been associated with several cancer types including cancers in the colon (287), ovary (288) and prostate (289). PTGS1 may cooperate with PTGS2, as knockdown of both molecules synergistically inhibited PCa cell growth *in vitro* (289).

In addition to the above-discussed signaling pathways commonly altered in the two subtypes of CRPC, $AR^{+/hi}$ CRPC seems to uniquely rely on steroid signaling whereas $AR^{-/lo}$ CRPC uniquely impacts immune/inflammation pathways (152). Although the detailed mechanisms are under investigation, these results suggest that $AR^{+/hi}$ CRPC remain responsive to steroid hormones that may impart a competitive advantage in surviving ADT (Figure 5-4). The $AR^{-/lo}$ as well as $AR^{-}NE^{-}$ (290) CRPCs, on the other hand, are enriched in immune/inflammation gene signatures such as antigen presentation pathway, adaptive immune cells, and interferon signaling (Figure 5-5). The exact meaning and significance of the enrichment of such immune and inflammation related signatures in $AR^{-/lo}$ CRPC remain to be elucidated. Interestingly, however, clinical bone metastatic CRPC appears to display a negative association between AR signaling and expression of some immune responses genes such as HLA-A (291). In our system, the $AR^{-/lo}$ LAPC9 CRPC expresses high levels of HLA-A

(7749 times higher than corresponding AD tumors) whereas AR^{+} LNCaP CRPC hardly expresses any HLA-A (FDR > 0.05).

In summary, by interrogating and annotating the RNA-Seq data related to two distinct CRPC models that recapitulate the evolution of AD tumors to either $AR^{+/hi}$ or $AR^{-/lo}$ CRPC phenotype, my bioinformatics analysis has helped to uncover transcriptomic changes both common to the two subtypes of CRPC and unique to each subtype. Thus, both $AR^{+/hi}$ LNCaP CRPC and $AR^{-/lo}$ LAPC9 CRPC are enriched in genes involved in Stemness, Neurogenesis, and Lipid Signaling. On the other hand, $AR^{+/hi}$ LNCaP CRPC preferentially expresses genes associated with Steroidogenesis (hormone biosynthesis and metabolism) whereas $AR^{-/lo}$ LAPC9 CRPC is enriched in genes that participate in Immune and Inflammation Signaling. These results have, overall, tested my starting hypothesis and suggest that, not surprisingly, ADT and enzalutamide treatment cause reprogramming of AD PCa cells through inducing widespread gene expression changes. There is little doubt that many of these altered genes and signaling pathways likely play causative roles in CRPC reprogramming and thus represent important regulators of CRPC development. Importantly, my bioinformatics analysis has also helped to pinpoint potential therapeutic target(s) and strategies for tackling the two subtypes of CRPC, as best illustrated by identification of BCL-2 upregulation in both enzalutamide-resistant $AR^{+/hi}$ LNCaP and $AR^{-/lo}$ LAPC9 CRPC and our therapeutic experiments using the BCL-2 inhibitor ABT-199. Consequently, ABT-199 alone inhibits the growth of LAPC9 CRPC while ABT-199 in combination with enzalutamide greatly suppresses the emergence of enzalutamide-

resistant secondary CRPC (152). These preclinical data, originated from bioinformatics-based discoveries, have set the stage for novel clinical trials.

CHAPTER SIX

Conclusions and perspectives

6.1 Conclusions

My Ph.D. thesis research has been centered on employing bioinformatics approaches and tools to help elucidate intrinsic PCa cell heterogeneity and induced PCa cell plasticity. For the former, we interrogated both the RNA-Seq transcriptomes and the histone modifications (H3K4me3 and H3K27me3) based ChIP-Seq data sets obtained in the CSC-enriched $PSA^{-/lo}$ LNCaP cells and the less tumorigenic PSA^{+} counterparts (Chapter 3). For the latter, I helped analyze and interpret the NGS data in PCa cell reprogramming caused by inducible NANOG overexpression *in vitro* (Chapter 4) or by chronic castration treatment *in vivo* (Chapter 5). The results from the 3 experimental systems have converged on several common findings and also revealed some novel insights on each unique model. For example, both unperturbed $PSA^{-/lo}$ cells and NANOG- and treatment-reprogrammed PCa cells are enriched in SC and, surprisingly, neurogenesis genes. The $PSA^{-/lo}$ and PSA^{+} subpopulations of LNCaP cells, though lineage-related, possess distinct histone modification patterns. NANOG-induced reprogramming of bulk LNCaP cells to the castration-resistant, stem-like state, surprisingly, appears to involve a 'hybrid' mode of action, i.e., engaging both prostate lineage master factors AR and FOXA1 at early phases and one of the OSKM factors, c-MYC, at a late stage. Finally, long-term castration *in vivo* led to two distinct, $AR^{+/hi}$ (LNCaP) or $AR^{-/lo}$ (LAPC9) CRPC models, both of which are enriched in SC/CSC as well as neurogenesis gene signatures.

Interestingly, although GO analysis reveals "Stem Cell Signaling" as a common

term to be enriched in PSA$^{-/lo}$ and the 'reprogrammed' PCa cells/tumor, these models seem to preferentially express different SC/development genes. For example, unperturbed (naïve) PSA$^{-/lo}$ LNCaP cells preferentially express genes such as TGFβ3, FZD2, and KLF5, and NANOG-reprogrammed LNCaP cells highly express ABCG2, WNT5A, and IGFBP5, whereas treatment-enriched CRPC cells express high levels of BCL-2, WNT5A, BMP5, and SALL4. Presumably, these SC genes, in a model-dependent manner, help sustain cell identity and stemness properties. As targeting CSCs has been proposed as an important strategy for cancer therapy (60-66), these stemness genes may represent potential therapeutic targets. This is best illustrated by the upregulation of BCL-2, a critical CSC gene, in both AR^{+} and AR$^{-/lo}$ CRPC cells. Thus, treatment of AR$^{+/hi}$ LNCaP CRPC with the BCL-2 inhibitor ABT-199 and AR-targeting enzalutamide significantly prevents the emergence of secondary CRPC, whereas treatment of AR$^{-/lo}$ LAPC9 CRPC with ABT-199 alone inhibits CRPC growth. In short, CSCs possess stemness properties that are supported by expression of SC genes, which can serve as effective therapeutic targets.

Our earlier studies have demonstrated that the PSA$^{-/lo}$ cell population is enriched in heterogeneous CSC subpopulations, which contribute to CRPC (20, 25). Our current studies indicate that naïve (i.e., untreated) PSA$^{-/lo}$ cells possess SC gene signatures (Chapter 3). NANOG-reprogramed CSCs and castration-selected PCa cells also displayed low cellular PSA expression. Also, cellular PSA expression is consistently low in CRPC regardless of the AR status (123, 152). In the Rajan dataset, log2 fold change of PSA is -2.68 with FDR 1.12e^{-09} (CRPCs over primary tumors), and log2 fold change of AR is 0.11 with FDR 0.96. These findings in patient ADT-failed

tumors support our xenograft studies and imply that cells with low PSA expression (i.e., losing differentiation) may play key roles in the evolution of PCa to CRPC. This emerging idea has been observed in other investigations that link $PSA^{-/lo}$ cells to aggressive PCa variants, such as neuroendocrine prostate cancer (NEPC), which is markedly increased in CRPC patients (292, 293). In short, $PSA^{-/lo}$ cells are enriched in CRPC, and are linked to advanced PCa phenotypes such as CSC properties and NEPC features.

Castration-induced NEPC (sometimes called CRPC-NE) is a hormone-refractory type of PCa with poor prognosis and represents about 25% of late-stage treatment-failed disease (294). CRPC-NE, like NEPC de novo (i.e., the primary tumor diagnosed as NEPC), frequently expresses mature NE markers such as CHA (chromogrannin A) and SYN (synaptophysin). The molecular underpinning of treatment-induced NEPC manifestation is not clear, but it may arise from transdifferentiation of PCa adenocarcinoma cells (292, 293). Molecularly, CRPC-NE generally lacks PSA expression and shows elevated BCL-2 expression, which echo our experimental systems. Nevertheless, $PSA^{-/lo}$ PCa cells, NANOG-reprogrammed cells, and castration-selected tumors all lack the expression of CHA and SYN. Strikingly, however, both naïve $PSA^{-/lo}$ and reprogrammed PCa cells, like the $AR^{-}PSA^{-}$ normal prostate basal/stem cells (112), are enriched in neurogenesis genes, which normally regulate the neural progenitor cell activities and neural cell development. It is tempting to speculate that these neurogenesis genes may promote $PSA^{-/lo}$ and treatment-reprogrammed PCa cells to undergo potential differentiation along neural lineages and differentiation into neuron-like or neural progenitor-like cells but without

pushing them to full NE differentiation. Of interest, NEPC cells are often found to have overexpression of cell cycle genes (292) such as UBE2C, cyclin D1, and AURKA, and these genes are also upregulated in NANOG-reprogrammed CSCs (151).

Enrichment of a neurogenesis gene signature appears to be a common phenomenon in PSA^{-} normal basal/stem epithelial cells, $PSA^{-/lo}$ PCa cells, and both AR^{+} and $AR^{-/lo}$ CRPC cells. Although the precise role of these genes in normal prostate development, prostate tumorigenesis, regulation of PCSCs, and CRPC emergence and evolution into NEPC remains to be elucidated, two pertinent points may be worth considering. First, at least some of the neurogensis genes seem to be functionally important in regulating the stem/progenitor activities of PCSCs and tumor aggressiveness. Thus, quite a few neurogenesis genes preferentially expressed in $PSA^{-/lo}$ PCa cells are inversely associated with poor patient survival. Moreover, knocking down expression of some of these genes, e.g., NRXN1 and CHRM3, attenuates CSC properties and tumorigenicity of $PSA^{-/lo}$ cells. Second, nerves are present throughout the prostate stroma, and various types of neural cells and neurotransmitters exist in the tumor microenvironment, which may regulate PCSC survival and functions through the expression of certain neurogenesis genes such as receptors. A recent study demonstrated that adrenergic fibers from the sympathetic nervous system had a major impact on tumor initiation via β2- and β3- adrenergic receptors, whereas cholinergic fibers from the parasympathetic nervous system were found to regulate PCa cell invasion, migration, and metastasis via Chrm1 signals (295). Therefore, the role of the neurogenesis program and neurogenesis genes in regulating PCSCs and development of CRPC and NEPC will be a fertile ground for

future investigation.

Lineage related $PSA^{-/lo}$ and PSA^{+} cells have quite different transcriptome profiles and epigenetic landscapes. For instance, many epigenetic factors (e.g., ARID2, JMJDIC, KDM5A, KDM6A, KDM6B, SETD2, SETD5, TET1, and TET3) are upregulated in $PSA^{-/lo}$ cells compared to PSA^{+} cells. Epigenetic modifications contribute to dynamic plasticity of cancer cells and heterogeneity of cell populations that advance tumor progression and treatment resistance (182). Our data suggest that various epigenetic regulators involved in multiple epigenetic mechanisms may play an important role in regulating the CSC properties of $PSA^{-/lo}$ cells. In PCa, KDMs are considered prognostic factors and therapeutic targets (296). For example, both KDM6B mRNA and protein expression levels are upregulated in human PCa samples, especially in metastatic tumors (297). KDM5A has been suggested as an oncogene and potential prognostic factor to predict disease recurrence (296). KDMs, which mediate histone lysine demethylation, together with other epigenetic machinery, have been implicated in maintaining tumor-initiating cell plasticity and metastatic potential (296). Thus, although it is unknown whether epigenetic changes are the cause or effect of transcriptomic differences, our data suggest that epigenetic and transcriptomic heterogeneity may likely support phenotypic and functional differences of lineage-related subpopulations of PCa cells and sustain the inherent CSC features of $PSA^{-/lo}$ cells.

Transcriptional networks are another critical regulatory mechanism contributing to CSC features. For example, NANOG, as a TF, re-directs PCa cell fate towards undifferentiated and castration-resistant CSCs via multiple mechanisms. First,

NANOG 'hijacks' the lineage TF, AR, and pioneer factor, FOXA1, to reprogram the AR transcriptome. Late in the reprogramming process, NANOG also engages one of the iPSC reprogramming factors MYC to re-establish cell-cycle autonomy. Moreover, NANOG may modulate chromatin structure to coordinate dynamic gene expression changes required for full reprogramming. It is interesting to note that reprogramming of differentiated somatic cells to pluripotent stage generally requires at least 3-4 TFs; intriguingly, however, NANOG alone can gradually (in ~2-3 weeks) push LNCaP PCa cells back to a more undifferentiated state through activation of transcriptomic reprogramming of multiple TF networks. These results indicate that somatic cancer cells most likely possess a more permissive epigenetic and transcriptomic environment compared with normal cells.

Therapeutic reprograming also reshapes a whole transcriptome and triggers cells to regain plasticity during the development of CRPC. Interestingly, though, long-term castration leads to two subtypes, i.e., $AR^{+/hi}$ and $AR^{-/lo}$, of CRPC, both of which are enriched in the expression of, commonly or uniquely, some SC/CSC genes. For example, SALL4 is only upregulated in AR^{+} CRPC. Knocking down SALL4 and/or its interacting protein, NANOG, has been shown to alter SC properties in ES cells by reducing expression of SC genes and increasing expression of differentiation-associated genes (264). Also, inhibition of SALL4 expression has been reported to reduce proliferation and promote apoptosis in PCa cells via modulation of BCL-2 and BAX expression (267). However, the $AR^{+/hi}$ LNCaP CRPC only express high levels of SALL4 (152) but not NANOG (151), again suggesting fundamental differences between normal ES cells and reprogrammed CSCs. On the other hand, some SC

genes (e.g., BCL-2) are commonly overexpressed in AR^{+} and $AR^{-/lo}$ CRPC subtypes, rendering both sensitive to the BCL-2 inhibitor, ABT-199.

6.2 Perspectives

My bioinformatics analysis of our cell and xenograft models has raised new questions pertinent to understanding PCa cell heterogeneity and plasticity and helped open new areas of research. For example, what is the real role of the neurogenesis program in regulating intrinsic and induced stemness and invasiveness in PCa cells? What is the master regulator(s) of the neurogenesis program? In NANOG-induced PCa cell reprogramming, how do epigenetic mechanisms contribute to the reprogramming process in addition to NANOG-induced transcriptomic changes? In castration-induced xenograft reprogramming, how can we dissect out the (murine) host contribution to the transcriptomic changes caused by ADT? Although RNA-Seq can identify global gene expression profiles that represent many molecular features of tumors (298), there are still some biological and computational confounding factors that complicate the translation of transcriptomic profiles into clinical diagnosis (299). For instance, untreated tumors may harbor only a small population of CSCs with various features and CSC subpopulations may dynamically change during tumor progression. Therefore, attempt to distinguish key features of such small subsets of CSCs from the random noise of bulk-processed tumors is one of the greatest challenges in the RNA-Seq field (300). The fast evolving single-cell RNA sequencing may provide high-resolution information of cancer heterogeneity and partially cancel the inherent noise from bulky tumor RNA-Seq. Integration of refined histological

examination of tumor morphology with various -omics tools and tailored bioinformatic and statistic analysis may also help mitigate the issue of tumor heterogeneity.

Cancer has been classically defined as an abnormal, uncontrolled, growth of transformed cells, so conventional chemotherapies (e.g., paclitaxel) mainly target rapidly proliferating cells. Recent appreciation and intensive studies of tumor cell heterogeneity have opened a new era in cancer research, leading to the development of precision oncology, which aims to personalize accurate and effective treatment strategies for each individual cancer patient based on integrative evidence from a tumor's molecular (-omics) profile (e.g., genomics, transcriptomics, proteomics, metabolomics, and epigenomics), disease theranostics (e.g., disease diagnosis and prognosis), and pharmacogenomics (301). Additionally, precision oncology may also aim to redirect evolving cancers from a devastating lethal disease to a manageable chronic disease (302). Due to the heterogeneous and dynamic nature of cancer cell subpopulations, emergence of treatment-resistant tumors is, theoretically, inevitable, and, for most tumor types, the response rate, progression-free survival, and overall survival of single-agent treatments has likely plateaued. Therefore, more and more investigations have been dedicated to identifying rationally designed combination and sequential therapy strategies. In metastatic breast cancer, patients treated with paclitaxel (targeting cell cycle) and trastuzumab (targeting HER2 overexpressing cells) have better overall survival compared to paclitaxel alone (303). In PCa, two recent clinical trials, the CHAARTED trial (304, 305) and the STAMPEDE trial (306), have demonstrated the benefit of ADT plus docetaxel in treating advanced or metastatic CRPC over ADT alone. ADT plus abiraterone/glucocorticoid combination therapy has

also demonstrated improved clinical outcomes compared to ADT alone (307, 308). In such new combinatorial treatment, transcriptomics and other -omics approaches, especially single cell RNA-Seq, should help map out the responsive and resistant cancer cell populations and delineate the evolutionary trajectory of the tumors during treatment.

In the context of tumor cell heterogeneity, preclinical and clinical observations have implicated CSCs as the potential driving force for tumor relapse and CSC-targeted therapeutic approaches can improve patient survival (309). Various CSC-targeting therapeutic strategies are being attempted, including targeting critical surface and adhesion molecules, altering the intrinsic epigenetic landscape, redirecting CSC signaling pathways, manipulating microenvironment signaling, suppressing drug-efflux pumps, and inducing CSC apoptosis and differentiation (310). Some efforts are also being made to target CSC in PCa, especially CRPC. Examples include developing inhibitors of Wnt and Shh signaling, which has been implicated in regulating PCSC stemness (63, 64).

Is targeting of CSCs alone sufficient to cure cancer? The answer is yes and no, based on our xenograft studies (124, 152). Theoretically, combinatorial targeting of multiple CSC regulators such as BCL-2, integrin α2, MYC, and N-cadherin could eliminate the $AR^{-/lo}$ LAPC9-subtype of CRPC. However, for the $AR^{+/hi}$ LNCaP-subtype of CRPC, anti-CSC strategies must be combined with AR-blocking agents such as enzalutamide (152). Because PCa patient primary (prostate) tumors harbor both AR^{+} and $AR^{-/lo}$ cells/clones and their metastases may be AR^{+} or $AR^{-/lo}$ (252), a combinatorial approach is entailed to co-target AR^{+} PCa cells/metastases with

enzalutamide and $AR^{-/lo}$ PCa cells with CSC inhibitors. By dissecting various CSC transcriptomic profiles, our results (152) lead to a proof-of-principle study combining a CSC-targeting drug (ABT-199, BCL-2 inhibitor) with ADT (enzalutamide), which showed beneficial effect over single-agent alone in mouse CRPC xenograft models. Importantly, these pre-clinical studies have supported the launch of a new Phase Ib/II clinical trial of treating metastatic CRPC with a combination of enzalutamide and ABT-199. RNA-Seq informing the design of combination therapeutic strategies and precision medicine is an effective strategy in cancer therapeutics and may lay the solid foundations for curing cancer.

Bibliography

1. 1987. Biological aspects of renal cancer and its treatments. *Seminars in nephrology* 7: 106-170.
2. Marusyk, A., V. Almendro, and K. Polyak. 2012. Intra-tumour heterogeneity: a looking glass for cancer? *Nature reviews. Cancer* 12: 323-334.
3. Negrini, S., V. G. Gorgoulis, and T. D. Halazonetis. 2010. Genomic instability--an evolving hallmark of cancer. *Nature reviews. Molecular cell biology* 11: 220-228.
4. Waddington, C. H. 1950. A discussion on the measurement of growth and form; the biological foundations of measurements of growth and form. *Proceedings of the Royal Society of London. Series B, Biological sciences* 137: 509-515.
5. Egeblad, M., E. S. Nakasone, and Z. Werb. 2010. Tumors as organs: complex tissues that interface with the entire organism. *Developmental cell* 18: 884-901.
6. Schwarz, R. F., C. K. Ng, S. L. Cooke, S. Newman, J. Temple, A. M. Piskorz, D. Gale, K. Sayal, M. Murtaza, P. J. Baldwin, N. Rosenfeld, H. M. Earl, E. Sala, M. Jimenez-Linan, C. A. Parkinson, F. Markowetz, and J. D. Brenton. 2015. Spatial and temporal heterogeneity in high-grade serous ovarian cancer: a phylogenetic analysis. *PLoS medicine* 12: e1001789.
7. Zhang, J., J. Fujimoto, J. Zhang, D. C. Wedge, X. Song, J. Zhang, S. Seth, C. W. Chow, Y. Cao, C. Gumbs, K. A. Gold, N. Kalhor, L. Little, H. Mahadeshwar, C. Moran, A. Protopopov, H. Sun, J. Tang, X. Wu, Y. Ye, W. N. William, J. J. Lee, J. V. Heymach, W. K. Hong, S. Swisher, Wistuba, II, and P. A. Futreal.

2014. Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science* 346: 256-259.

8. Bochtler, T., F. Stolzel, C. E. Heilig, C. Kunz, B. Mohr, A. Jauch, J. W. Janssen, M. Kramer, A. Benner, M. Bornhauser, A. D. Ho, G. Ehninger, M. Schaich, and A. Kramer. 2013. Clonal heterogeneity as detected by metaphase karyotyping is an indicator of poor prognosis in acute myeloid leukemia. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 31: 3898-3905.

9. Park, S. Y., M. Gonen, H. J. Kim, F. Michor, and K. Polyak. 2010. Cellular and genetic diversity in the progression of in situ human breast carcinomas to an invasive phenotype. *The Journal of clinical investigation* 120: 636-644.

10. Rasheed, Z. A., J. Kowalski, B. D. Smith, and W. Matsui. 2011. Concise review: Emerging concepts in clinical targeting of cancer stem cells. *Stem cells* 29: 883-887.

11. Zeppernick, F., R. Ahmadi, B. Campos, C. Dictus, B. M. Helmke, N. Becker, P. Lichter, A. Unterberg, B. Radlwimmer, and C. C. Herold-Mende. 2008. Stem cell marker CD133 affects clinical outcome in glioma patients. *Clinical cancer research : an official journal of the American Association for Cancer Research* 14: 123-129.

12. Ginestier, C., M. H. Hur, E. Charafe-Jauffret, F. Monville, J. Dutcher, M. Brown, J. Jacquemier, P. Viens, C. G. Kleer, S. Liu, A. Schott, D. Hayes, D. Birnbaum, M. S. Wicha, and G. Dontu. 2007. ALDH1 is a marker of normal and malignant

human mammary stem cells and a predictor of poor clinical outcome. *Cell stem cell* 1: 555-567.

13. Rasheed, Z. A., J. Yang, Q. Wang, J. Kowalski, I. Freed, C. Murter, S. M. Hong, J. B. Koorstra, N. V. Rajeshkumar, X. He, M. Goggins, C. Iacobuzio-Donahue, D. M. Berman, D. Laheru, A. Jimeno, M. Hidalgo, A. Maitra, and W. Matsui. 2010. Prognostic significance of tumorigenic cells with mesenchymal features in pancreatic adenocarcinoma. *Journal of the National Cancer Institute* 102: 340-351.

14. Sheridan, C., H. Kishimoto, R. K. Fuchs, S. Mehrotra, P. Bhat-Nakshatri, C. H. Turner, R. Goulet, Jr., S. Badve, and H. Nakshatri. 2006. CD44+/CD24- breast cancer cells exhibit enhanced invasive properties: an early step necessary for metastasis. *Breast cancer research : BCR* 8: R59.

15. Li, W., H. Ma, J. Zhang, L. Zhu, C. Wang, and Y. Yang. 2017. Unraveling the roles of CD44/CD24 and ALDH1 as cancer stem cell markers in tumorigenesis and metastasis. *Scientific reports* 7: 13856.

16. Liu, R., X. Wang, G. Y. Chen, P. Dalerba, A. Gurney, T. Hoey, G. Sherlock, J. Lewicki, K. Shedden, and M. F. Clarke. 2007. The prognostic role of a gene signature from tumorigenic breast-cancer cells. *The New England journal of medicine* 356: 217-226.

17. Kim, C., R. Gao, E. Sei, R. Brandt, J. Hartman, T. Hatschek, N. Crosetto, T. Foukakis, and N. E. Navin. 2018. Chemoresistance Evolution in Triple-Negative Breast Cancer Delineated by Single-Cell Sequencing. *Cell* 173: 879-893 e813.

18. Sun, C., L. Wang, S. Huang, G. J. Heynen, A. Prahallad, C. Robert, J. Haanen, C. Blank, J. Wesseling, S. M. Willems, D. Zecchin, S. Hobor, P. K. Bajpe, C. Lieftink, C. Mateus, S. Vagner, W. Grernrum, I. Hofland, A. Schlicker, L. F. Wessels, R. L. Beijersbergen, A. Bardelli, F. Di Nicolantonio, A. M. Eggermont, and R. Bernards. 2014. Reversible and adaptive resistance to BRAF(V600E) inhibition in melanoma. *Nature* 508: 118-122.

19. Greaves, M., and C. C. Maley. 2012. Clonal evolution in cancer. *Nature* 481: 306-313.

20. Tang, D. G. 2012. Understanding cancer stem cell heterogeneity and plasticity. *Cell research* 22: 457-472.

21. Medema, J. P. 2013. Cancer stem cells: the challenges ahead. *Nature cell biology* 15: 338-344.

22. Klonisch, T., E. Wiechec, S. Hombach-Klonisch, S. R. Ande, S. Wesselborg, K. Schulze-Osthoff, and M. Los. 2008. Cancer stem cell markers in common cancers - therapeutic implications. *Trends in molecular medicine* 14: 450-460.

23. O'Brien, C. A., A. Pollett, S. Gallinger, and J. E. Dick. 2007. A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. *Nature* 445: 106-110.

24. Magee, J. A., E. Piskounova, and S. J. Morrison. 2012. Cancer stem cells: impact, heterogeneity, and uncertainty. *Cancer cell* 21: 283-296.

25. Qin, J., X. Liu, B. Laffin, X. Chen, G. Choy, C. R. Jeter, T. Calhoun-Davis, H. Li, G. S. Palapattu, S. Pang, K. Lin, J. Huang, I. Ivanov, W. Li, M. V. Suraneni, and D. G. Tang. 2012. The PSA(-/lo) prostate cancer cell population harbors self-

renewing long-term tumor-propagating cells that resist castration. *Cell stem cell* 10: 556-569.

26. Singh, S. K., C. Hawkins, I. D. Clarke, J. A. Squire, J. Bayani, T. Hide, R. M. Henkelman, M. D. Cusimano, and P. B. Dirks. 2004. Identification of human brain tumour initiating cells. *Nature* 432: 396-401.

27. Al-Hajj, M., M. S. Wicha, A. Benito-Hernandez, S. J. Morrison, and M. F. Clarke. 2003. Prospective identification of tumorigenic breast cancer cells. *Proceedings of the National Academy of Sciences of the United States of America* 100: 3983-3988.

28. Ricci-Vitiani, L., D. G. Lombardi, E. Pilozzi, M. Biffoni, M. Todaro, C. Peschle, and R. De Maria. 2007. Identification and expansion of human colon-cancer-initiating cells. *Nature* 445: 111-115.

29. Dalerba, P., S. J. Dylla, I. K. Park, R. Liu, X. Wang, R. W. Cho, T. Hoey, A. Gurney, E. H. Huang, D. M. Simeone, A. A. Shelton, G. Parmiani, C. Castelli, and M. F. Clarke. 2007. Phenotypic characterization of human colorectal cancer stem cells. *Proceedings of the National Academy of Sciences of the United States of America* 104: 10158-10163.

30. Zhang, S., C. Balch, M. W. Chan, H. C. Lai, D. Matei, J. M. Schilder, P. S. Yan, T. H. Huang, and K. P. Nephew. 2008. Identification and characterization of ovarian cancer-initiating cells from primary human tumors. *Cancer research* 68: 4311-4320.

31. Li, C., C. J. Lee, and D. M. Simeone. 2009. Identification of human pancreatic cancer stem cells. *Methods in molecular biology* 568: 161-173.

32. Li, C., D. G. Heidt, P. Dalerba, C. F. Burant, L. Zhang, V. Adsay, M. Wicha, M. F. Clarke, and D. M. Simeone. 2007. Identification of pancreatic cancer stem cells. *Cancer research* 67: 1030-1037.

33. Visvader, J. E., and G. J. Lindeman. 2012. Cancer stem cells: current status and evolving complexities. *Cell stem cell* 10: 717-728.

34. Visvader, J. E. 2011. Cells of origin in cancer. *Nature* 469: 314-322.

35. Hanahan, D., and R. A. Weinberg. 2011. Hallmarks of cancer: the next generation. *Cell* 144: 646-674.

36. Biteau, B., C. E. Hochmuth, and H. Jasper. 2011. Maintaining tissue homeostasis: dynamic control of somatic stem cell activity. *Cell stem cell* 9: 402-411.

37. Held, M. A., D. P. Curley, D. Dankort, M. McMahon, V. Muthusamy, and M. W. Bosenberg. 2010. Characterization of melanoma cells capable of propagating tumors from a single cell. *Cancer research* 70: 388-397.

38. Hadjimichael, C., K. Chanoumidou, N. Papadopoulou, P. Arampatzi, J. Papamatheakis, and A. Kretsovali. 2015. Common stemness regulators of embryonic and cancer stem cells. *World journal of stem cells* 7: 1150-1184.

39. ten Berge, D., D. Kurek, T. Blauwkamp, W. Koole, A. Maas, E. Eroglu, R. K. Siu, and R. Nusse. 2011. Embryonic stem cells require Wnt proteins to prevent differentiation to epiblast stem cells. *Nature cell biology* 13: 1070-1075.

40. Wray, J., T. Kalkan, S. Gomez-Lopez, D. Eckardt, A. Cook, R. Kemler, and A. Smith. 2011. Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression

of the pluripotency network and increases embryonic stem cell resistance to differentiation. *Nature cell biology* 13: 838-845.

41. Niwa, H. 2011. Wnt: what's needed to maintain pluripotency? *Nature cell biology* 13: 1024-1026.

42. Kelly, K. F., D. Y. Ng, G. Jayakumaran, G. A. Wood, H. Koide, and B. W. Doble. 2011. beta-catenin enhances Oct-4 activity and reinforces pluripotency through a TCF-independent mechanism. *Cell stem cell* 8: 214-227.

43. Korkaya, H., A. Paulson, E. Charafe-Jauffret, C. Ginestier, M. Brown, J. Dutcher, S. G. Clouthier, and M. S. Wicha. 2009. Regulation of mammary stem/progenitor cells by PTEN/Akt/beta-catenin signaling. *PLoS biology* 7: e1000121.

44. Reya, T., and H. Clevers. 2005. Wnt signalling in stem cells and cancer. *Nature* 434: 843-850.

45. Zurawel, R. H., S. A. Chiappa, C. Allen, and C. Raffel. 1998. Sporadic medulloblastomas contain oncogenic beta-catenin mutations. *Cancer research* 58: 896-899.

46. Vermeulen, L., E. M. F. De Sousa, M. van der Heijden, K. Cameron, J. H. de Jong, T. Borovski, J. B. Tuynman, M. Todaro, C. Merz, H. Rodermond, M. R. Sprick, K. Kemper, D. J. Richel, G. Stassi, and J. P. Medema. 2010. Wnt activity defines colon cancer stem cells and is regulated by the microenvironment. *Nature cell biology* 12: 468-476.

47. Malanchi, I., A. Santamaria-Martinez, E. Susanto, H. Peng, H. A. Lehr, J. F. Delaloye, and J. Huelsken. 2011. Interactions between cancer stem cells and their niche govern metastatic colonization. *Nature* 481: 85-89.

48. Hochedlinger, K., Y. Yamada, C. Beard, and R. Jaenisch. 2005. Ectopic expression of Oct-4 blocks progenitor-cell differentiation and causes dysplasia in epithelial tissues. *Cell* 121: 465-477.

49. Takahashi, K., and S. Yamanaka. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126: 663-676.

50. Sancho-Martinez, I., and J. C. Izpisua Belmonte. 2013. Stem cells: Surf the waves of reprogramming. *Nature* 493: 310-311.

51. Polo, J. M., E. Anderssen, R. M. Walsh, B. A. Schwarz, C. M. Nefzger, S. M. Lim, M. Borkent, E. Apostolou, S. Alaei, J. Cloutier, O. Bar-Nur, S. Cheloufi, M. Stadtfeld, M. E. Figueroa, D. Robinton, S. Natesan, A. Melnick, J. Zhu, S. Ramaswamy, and K. Hochedlinger. 2012. A molecular roadmap of reprogramming somatic cells into iPS cells. *Cell* 151: 1617-1632.

52. Stadhouders, R., E. Vidal, F. Serra, B. Di Stefano, F. Le Dily, J. Quilez, A. Gomez, S. Collombet, C. Berenguer, Y. Cuartero, J. Hecht, G. J. Filion, M. Beato, M. A. Marti-Renom, and T. Graf. 2018. Transcription factors orchestrate dynamic interplay between genome topology and gene regulation during cell reprogramming. *Nature genetics* 50: 238-249.

53. Wang, R., K. Chadalavada, J. Wilshire, U. Kowalik, K. E. Hovinga, A. Geber, B. Fligelman, M. Leversha, C. Brennan, and V. Tabar. 2010. Glioblastoma stem-like cells give rise to tumour endothelium. *Nature* 468: 829-833.

54. Ricci-Vitiani, L., R. Pallini, M. Biffoni, M. Todaro, G. Invernici, T. Cenci, G. Maira, E. A. Parati, G. Stassi, L. M. Larocca, and R. De Maria. 2010. Tumour vascularization via endothelial differentiation of glioblastoma stem-like cells. *Nature* 468: 824-828.

55. Patrawala, L., T. Calhoun, R. Schneider-Broussard, J. Zhou, K. Claypool, and D. G. Tang. 2005. Side population is enriched in tumorigenic, stem-like cancer cells, whereas ABCG2+ and ABCG2- cancer cells are similarly tumorigenic. *Cancer research* 65: 6207-6219.

56. Patrawala, L., T. Calhoun, R. Schneider-Broussard, H. Li, B. Bhatia, S. Tang, J. G. Reilly, D. Chandra, J. Zhou, K. Claypool, L. Coghlan, and D. G. Tang. 2006. Highly purified CD44+ prostate cancer cells from xenograft human tumors are enriched in tumorigenic and metastatic progenitor cells. *Oncogene* 25: 1696-1708.

57. Blaylock, R. L. 2015. Cancer microenvironment, inflammation and cancer stem cells: A hypothesis for a paradigm change and new targets in cancer control. *Surgical neurology international* 6: 92.

58. Kalluri, R., and R. A. Weinberg. 2009. The basics of epithelial-mesenchymal transition. *The Journal of clinical investigation* 119: 1420-1428.

59. Krebs, A. M., J. Mitschke, M. Lasierra Losada, O. Schmalhofer, M. Boerries, H. Busch, M. Boettcher, D. Mougiakakos, W. Reichardt, P. Bronsert, V. G.

Brunton, C. Pilarsky, T. H. Winkler, S. Brabletz, M. P. Stemmler, and T. Brabletz. 2017. The EMT-activator Zeb1 is a key factor for cell plasticity and promotes metastasis in pancreatic cancer. *Nature cell biology* 19: 518-529.

60. Li, X., M. T. Lewis, J. Huang, C. Gutierrez, C. K. Osborne, M. F. Wu, S. G. Hilsenbeck, A. Pavlick, X. Zhang, G. C. Chamness, H. Wong, J. Rosen, and J. C. Chang. 2008. Intrinsic resistance of tumorigenic breast cancer cells to chemotherapy. *Journal of the National Cancer Institute* 100: 672-679.

61. Creighton, C. J., X. Li, M. Landis, J. M. Dixon, V. M. Neumeister, A. Sjolund, D. L. Rimm, H. Wong, A. Rodriguez, J. I. Herschkowitz, C. Fan, X. Zhang, X. He, A. Pavlick, M. C. Gutierrez, L. Renshaw, A. A. Larionov, D. Faratian, S. G. Hilsenbeck, C. M. Perou, M. T. Lewis, J. M. Rosen, and J. C. Chang. 2009. Residual breast cancers after conventional therapy display mesenchymal as well as tumor-initiating features. *Proceedings of the National Academy of Sciences of the United States of America* 106: 13820-13825.

62. Gupta, P. B., T. T. Onder, G. Jiang, K. Tao, C. Kuperwasser, R. A. Weinberg, and E. S. Lander. 2009. Identification of selective inhibitors of cancer stem cells by high-throughput screening. *Cell* 138: 645-659.

63. Yun, E. J., J. Zhou, C. J. Lin, E. Hernandez, L. Fazli, M. Gleave, and J. T. Hsieh. 2016. Targeting Cancer Stem Cells in Castration-Resistant Prostate Cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* 22: 670-679.

64. Merchant, A. A., and W. Matsui. 2010. Targeting Hedgehog--a cancer stem cell pathway. *Clinical cancer research : an official journal of the American Association for Cancer Research* 16: 3130-3140.

65. Clement, V., P. Sanchez, N. de Tribolet, I. Radovanovic, and A. Ruiz i Altaba. 2007. HEDGEHOG-GLI1 signaling regulates human glioma growth, cancer stem cell self-renewal, and tumorigenicity. *Current biology : CB* 17: 165-172.

66. Agliano, A., A. Calvo, and C. Box. 2017. The challenge of targeting cancer stem cells to halt metastasis. *Seminars in cancer biology* 44: 25-42.

67. Kreso, A., and J. E. Dick. 2014. Evolution of the cancer stem cell model. *Cell stem cell* 14: 275-291.

68. Chen, T., and S. Y. Dent. 2014. Chromatin modifiers and remodellers: regulators of cellular differentiation. *Nature reviews. Genetics* 15: 93-106.

69. Johnson, D. G., and S. Y. Dent. 2013. Chromatin: receiver and quarterback for cellular signals. *Cell* 152: 685-689.

70. Bernstein, B. E., T. S. Mikkelsen, X. Xie, M. Kamal, D. J. Huebert, J. Cuff, B. Fry, A. Meissner, M. Wernig, K. Plath, R. Jaenisch, A. Wagschal, R. Feil, S. L. Schreiber, and E. S. Lander. 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 125: 315-326.

71. Azuara, V., P. Perry, S. Sauer, M. Spivakov, H. F. Jorgensen, R. M. John, M. Gouti, M. Casanova, G. Warnes, M. Merkenschlager, and A. G. Fisher. 2006. Chromatin signatures of pluripotent cell lines. *Nature cell biology* 8: 532-538.

72. Ringrose, L., and R. Paro. 2007. Polycomb/Trithorax response elements and epigenetic memory of cell identity. *Development* 134: 223-232.

73. Mikkelsen, T. S., M. Ku, D. B. Jaffe, B. Issac, E. Lieberman, G. Giannoukos, P. Alvarez, W. Brockman, T. K. Kim, R. P. Koche, W. Lee, E. Mendenhall, A. O'Donovan, A. Presser, C. Russ, X. Xie, A. Meissner, M. Wernig, R. Jaenisch, C. Nusbaum, E. S. Lander, and B. E. Bernstein. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448: 553-560.

74. Fazzio, T. G., J. T. Huff, and B. Panning. 2008. An RNAi screen of chromatin proteins identifies Tip60-p400 as a regulator of embryonic stem cell identity. *Cell* 134: 162-174.

75. Boyer, L. A., T. I. Lee, M. F. Cole, S. E. Johnstone, S. S. Levine, J. P. Zucker, M. G. Guenther, R. M. Kumar, H. L. Murray, R. G. Jenner, D. K. Gifford, D. A. Melton, R. Jaenisch, and R. A. Young. 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* 122: 947-956.

76. Loh, Y. H., Q. Wu, J. L. Chew, V. B. Vega, W. Zhang, X. Chen, G. Bourque, J. George, B. Leong, J. Liu, K. Y. Wong, K. W. Sung, C. W. Lee, X. D. Zhao, K. P. Chiu, L. Lipovich, V. A. Kuznetsov, P. Robson, L. W. Stanton, C. L. Wei, Y. Ruan, B. Lim, and H. H. Ng. 2006. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nature genetics* 38: 431-440.

77. Mitsui, K., Y. Tokuzawa, H. Itoh, K. Segawa, M. Murakami, K. Takahashi, M. Maruyama, M. Maeda, and S. Yamanaka. 2003. The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* 113: 631-642.

78. Hnisz, D., B. J. Abraham, T. I. Lee, A. Lau, V. Saint-Andre, A. A. Sigova, H. A. Hoke, and R. A. Young. 2013. Super-enhancers in the control of cell identity and disease. *Cell* 155: 934-947.

79. Chen, J., Z. Zhang, L. Li, B. C. Chen, A. Revyakin, B. Hajj, W. Legant, M. Dahan, T. Lionnet, E. Betzig, R. Tjian, and Z. Liu. 2014. Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell* 156: 1274-1285.

80. Bernardo, G. M., and R. A. Keri. 2012. FOXA1: a transcription factor with parallel functions in development and cancer. *Bioscience reports* 32: 113-130.

81. Ye, D. Z., and K. H. Kaestner. 2009. Foxa1 and Foxa2 control the differentiation of goblet and enteroendocrine L- and D-cells in mice. *Gastroenterology* 137: 2052-2062.

82. Gualdi, R., P. Bossard, M. Zheng, Y. Hamada, J. R. Coleman, and K. S. Zaret. 1996. Hepatic specification of the gut endoderm in vitro: cell signaling and transcriptional control. *Genes & development* 10: 1670-1682.

83. Augello, M. A., T. E. Hickey, and K. E. Knudsen. 2011. FOXA1: master of steroid receptor function in cancer. *The EMBO journal* 30: 3885-3894.

84. Pei, D. 2009. Regulation of pluripotency and reprogramming by transcription factors. *The Journal of biological chemistry* 284: 3365-3369.

85. Liu, Y., B. Clem, E. K. Zuba-Surma, S. El-Naggar, S. Telang, A. B. Jenson, Y. Wang, H. Shao, M. Z. Ratajczak, J. Chesney, and D. C. Dean. 2009. Mouse fibroblasts lacking RB1 function form spheres and undergo reprogramming to a cancer stem cell phenotype. *Cell stem cell* 4: 336-347.

86. Lin, Y. L., Z. B. Han, F. Y. Xiong, L. Y. Tian, X. J. Wu, S. W. Xue, Y. R. Zhou, J. X. Deng, and H. X. Chen. 2011. Malignant transformation of 293 cells induced by ectopic expression of human Nanog. *Molecular and cellular biochemistry* 351: 109-116.

87. Jeter, C. R., M. Badeaux, G. Choy, D. Chandra, L. Patrawala, C. Liu, T. Calhoun-Davis, H. Zaehres, G. Q. Daley, and D. G. Tang. 2009. Functional evidence that the self-renewal gene NANOG regulates human tumor development. *Stem cells* 27: 993-1005.

88. Lee, T. K., A. Castilho, V. C. Cheung, K. H. Tang, S. Ma, and I. O. Ng. 2011. CD24(+) liver tumor-initiating cells drive self-renewal and tumor initiation through STAT3-mediated NANOG regulation. *Cell stem cell* 9: 50-63.

89. Bourguignon, L. Y., K. Peyrollier, W. Xia, and E. Gilad. 2008. Hyaluronan-CD44 interaction activates stem cell marker Nanog, Stat-3-mediated MDR1 gene expression, and ankyrin-regulated multidrug efflux in breast and ovarian tumor cells. *The Journal of biological chemistry* 283: 17635-17651.

90. Zbinden, M., A. Duquet, A. Lorente-Trigos, S. N. Ngwabyt, I. Borges, and A. Ruiz i Altaba. 2010. NANOG regulates glioma stem cells and is essential in vivo acting in a cross-functional network with GLI1 and p53. *The EMBO journal* 29: 2659-2674.

91. Niu, C. S., D. X. Li, Y. H. Liu, X. M. Fu, S. F. Tang, and J. Li. 2011. Expression of NANOG in human gliomas and its relationship with undifferentiated glioma cells. *Oncology reports* 26: 593-601.

92. Xu, C. X., M. Xu, L. Tan, H. Yang, J. Permuth-Wey, P. A. Kruk, R. M. Wenham, S. V. Nicosia, J. M. Lancaster, T. A. Sellers, and J. Q. Cheng. 2016. MicroRNA MiR-214 regulates ovarian cancer cell stemness by targeting p53/Nanog. *The Journal of biological chemistry* 291: 22851.

93. Jeter, C. R., B. Liu, X. Liu, X. Chen, C. Liu, T. Calhoun-Davis, J. Repass, H. Zaehres, J. J. Shen, and D. G. Tang. 2011. NANOG promotes cancer stem cell characteristics and prostate cancer resistance to androgen deprivation. *Oncogene* 30: 3833-3845.

94. Jeter, C. R., T. Yang, J. Wang, H. P. Chao, and D. G. Tang. 2015. Concise Review: NANOG in Cancer Stem Cells and Tumor Development: An Update and Outstanding Questions. *Stem cells* 33: 2381-2390.

95. Suva, M. L., N. Riggi, and B. E. Bernstein. 2013. Epigenetic reprogramming in cancer. *Science* 339: 1567-1570.

96. Sharma, S., T. K. Kelly, and P. A. Jones. 2010. Epigenetics in cancer. *Carcinogenesis* 31: 27-36.

97. Wainwright, E. N., and P. Scaffidi. 2017. Epigenetics and Cancer Stem Cells: Unleashing, Hijacking, and Restricting Cellular Plasticity. *Trends in cancer* 3: 372-386.

98. Marker, P. C., A. A. Donjacour, R. Dahiya, and G. R. Cunha. 2003. Hormonal, cellular, and molecular control of prostatic development. *Developmental biology* 253: 165-174.

99. Tan, M. H., J. Li, H. E. Xu, K. Melcher, and E. L. Yong. 2015. Androgen receptor: structure, role in prostate cancer and drug discovery. *Acta pharmacologica Sinica* 36: 3-23.

100. Prins, G. S., and O. Putz. 2008. Molecular signaling pathways that regulate prostate gland development. *Differentiation; research in biological diversity* 76: 641-659.

101. Gao, N., K. Ishii, J. Mirosevich, S. Kuwajima, S. R. Oppenheimer, R. L. Roberts, M. Jiang, X. Yu, S. B. Shappell, R. M. Caprioli, M. Stoffel, S. W. Hayward, and R. J. Matusik. 2005. Forkhead box A1 regulates prostate ductal morphogenesis and promotes epithelial cell maturation. *Development* 132: 3431-3443.

102. Kyprianou, N., and J. T. Isaacs. 1987. Quantal relationship between prostatic dihydrotestosterone and prostatic cell content: critical threshold concept. *The Prostate* 11: 41-50.

103. Collins, A. T., F. K. Habib, N. J. Maitland, and D. E. Neal. 2001. Identification and isolation of human prostate epithelial stem cells based on alpha(2)beta(1)-integrin expression. *Journal of cell science* 114: 3865-3872.

104. Ousset, M., A. Van Keymeulen, G. Bouvencourt, N. Sharma, Y. Achouri, B. D. Simons, and C. Blanpain. 2012. Multipotent and unipotent progenitors contribute to prostate postnatal development. *Nature cell biology* 14: 1131-1138.

105. Choi, N., B. Zhang, L. Zhang, M. Ittmann, and L. Xin. 2012. Adult murine prostate basal and luminal cells are self-sustained lineages that can both serve as targets for prostate cancer initiation. *Cancer cell* 21: 253-265.

106. Xin, L., D. A. Lawson, and O. N. Witte. 2005. The Sca-1 cell surface marker enriches for a prostate-regenerating cell subpopulation that can initiate prostate tumorigenesis. *Proceedings of the National Academy of Sciences of the United States of America* 102: 6942-6947.

107. Goldstein, A. S., J. Huang, C. Guo, I. P. Garraway, and O. N. Witte. 2010. Identification of a cell of origin for human prostate cancer. *Science* 329: 568-571.

108. Ratnacaram, C. K., M. Teletin, M. Jiang, X. Meng, P. Chambon, and D. Metzger. 2008. Temporally controlled ablation of PTEN in adult mouse prostate epithelium generates a model of invasive prostatic adenocarcinoma. *Proceedings of the National Academy of Sciences of the United States of America* 105: 2521-2526.

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

110. Wang, X., M. Kruithof-de Julio, K. D. Economides, D. Walker, H. Yu, M. V. Halili, Y. P. Hu, S. M. Price, C. Abate-Shen, and M. M. Shen. 2009. A luminal epithelial stem cell that is a cell of origin for prostate cancer. *Nature* 461: 495-500.

111. Wang, Z. A., A. Mitrofanova, S. K. Bergren, C. Abate-Shen, R. D. Cardiff, A. Califano, and M. M. Shen. 2013. Lineage analysis of basal epithelial cells reveals their unexpected plasticity and supports a cell-of-origin model for prostate cancer heterogeneity. *Nature cell biology* 15: 274-283.

112. Zhang, D., D. Park, Y. Zhong, Y. Lu, K. Rycaj, S. Gong, X. Chen, X. Liu, H. P. Chao, P. Whitney, T. Calhoun-Davis, Y. Takata, J. Shen, V. R. Iyer, and D. G. Tang. 2016. Stem cell and neurogenic gene-expression profiles link prostate basal cells to aggressive prostate cancer. *Nature communications* 7: 10798.

113. Park, J. W., J. K. Lee, J. W. Phillips, P. Huang, D. Cheng, J. Huang, and O. N. Witte. 2016. Prostate epithelial cell of origin determines cancer differentiation state in an organoid transformation assay. *Proceedings of the National Academy of Sciences of the United States of America* 113: 4482-4487.

114. Zhang, D., K. Lin, Y. Lu, K. Rycaj, Y. Zhong, H. P. Chao, T. Calhoun-Davis, J. Shen, and D. G. Tang. 2017. Developing a Novel Two-Dimensional Culture System to Enrich Human Prostate Luminal Progenitors that Can Function as a Cell of Origin for Prostate Cancer. *Stem cells translational medicine* 6: 748-760.

115. Karthaus, W. R., P. J. Iaquinta, J. Drost, A. Gracanin, R. van Boxtel, J. Wongvipat, C. M. Dowling, D. Gao, H. Begthel, N. Sachs, R. G. J. Vries, E. Cuppen, Y. Chen, C. L. Sawyers, and H. C. Clevers. 2014. Identification of multipotent luminal progenitor cells in human prostate organoid cultures. *Cell* 159: 163-175.

116. Li, H., X. Chen, T. Calhoun-Davis, K. Claypool, and D. G. Tang. 2008. PC3 human prostate carcinoma cell holoclones contain self-renewing tumor-initiating cells. *Cancer research* 68: 1820-1825.

117. Collins, A. T., P. A. Berry, C. Hyde, M. J. Stower, and N. J. Maitland. 2005. Prospective identification of tumorigenic prostate cancer stem cells. *Cancer research* 65: 10946-10951.

118. Dubrovska, A., S. Kim, R. J. Salamone, J. R. Walker, S. M. Maira, C. Garcia-Echeverria, P. G. Schultz, and V. A. Reddy. 2009. The role of PTEN/Akt/PI3K signaling in the maintenance and viability of prostate cancer stem-like cell populations. *Proceedings of the National Academy of Sciences of the United States of America* 106: 268-273.

119. Wei, C., W. Guomin, L. Yujun, and Q. Ruizhe. 2007. Cancer stem-like cells in human prostate carcinoma cells DU145: the seeds of the cell line? *Cancer biology & therapy* 6: 763-768.

120. Hurt, E. M., B. T. Kawasaki, G. J. Klarmann, S. B. Thomas, and W. L. Farrar. 2008. CD44+ CD24(-) prostate cells are early cancer progenitor/stem cells that provide a model for patients with poor prognosis. *British journal of cancer* 98: 756-765.

121. Domingo-Domenech, J., S. J. Vidal, V. Rodriguez-Bravo, M. Castillo-Martin, S. A. Quinn, R. Rodriguez-Barrueco, D. M. Bonal, E. Charytonowicz, N. Gladoun, J. de la Iglesia-Vicente, D. P. Petrylak, M. C. Benson, J. M. Silva, and C. Cordon-Cardo. 2012. Suppression of acquired docetaxel resistance in prostate

cancer through depletion of notch- and hedgehog-dependent tumor-initiating cells. *Cancer cell* 22: 373-388.

122. Rycaj, K., E. J. Cho, X. Liu, H. P. Chao, B. Liu, Q. Li, A. K. Devkota, D. Zhang, X. Chen, J. Moore, K. N. Dalby, and D. G. Tang. 2016. Longitudinal tracking of subpopulation dynamics and molecular changes during LNCaP cell castration and identification of inhibitors that could target the PSA-/lo castration-resistant cells. *Oncotarget* 7: 14220-14240.

123. Liu, X., X. Chen, K. Rycaj, H. P. Chao, Q. Deng, C. Jeter, C. Liu, S. Honorio, H. Li, T. Davis, M. Suraneni, B. Laffin, J. Qin, Q. Li, T. Yang, P. Whitney, J. Shen, J. Huang, and D. G. Tang. 2015. Systematic dissection of phenotypic, functional, and tumorigenic heterogeneity of human prostate cancer cells. *Oncotarget* 6: 23959-23986.

124. Chen, X., Q. Li, X. Liu, C. Liu, R. Liu, K. Rycaj, D. Zhang, B. Liu, C. Jeter, T. Calhoun-Davis, K. Lin, Y. Lu, H. P. Chao, J. Shen, and D. G. Tang. 2016. Defining a Population of Stem-like Human Prostate Cancer Cells That Can Generate and Propagate Castration-Resistant Prostate Cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* 22: 4505-4516.

125. Saad, F., and S. J. Hotte. 2010. Guidelines for the management of castrate-resistant prostate cancer. *Canadian Urological Association journal = Journal de l'Association des urologues du Canada* 4: 380-384.

126. Lowrance, W. T., B. J. Roth, E. Kirkby, M. H. Murad, and M. S. Cookson. 2016. Castration-Resistant Prostate Cancer: AUA Guideline Amendment 2015. *The Journal of urology* 195: 1444-1452.

127. Scher, H. I., C. Liebertz, W. K. Kelly, M. Mazumdar, C. Brett, L. Schwartz, G. Kolvenbag, L. Shapiro, and M. Schwartz. 1997. Bicalutamide for advanced prostate cancer: the natural versus treated history of disease. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 15: 2928-2938.

128. Scher, H. I., T. M. Beer, C. S. Higano, A. Anand, M. E. Taplin, E. Efstathiou, D. Rathkopf, J. Shelkey, E. Y. Yu, J. Alumkal, D. Hung, M. Hirmand, L. Seely, M. J. Morris, D. C. Danila, J. Humm, S. Larson, M. Fleisher, C. L. Sawyers, and C. Prostate Cancer Foundation/Department of Defense Prostate Cancer Clinical Trials. 2010. Antitumour activity of MDV3100 in castration-resistant prostate cancer: a phase 1-2 study. *Lancet* 375: 1437-1446.

129. Scher, H. I., K. Fizazi, F. Saad, M. E. Taplin, C. N. Sternberg, K. Miller, R. de Wit, P. Mulders, K. N. Chi, N. D. Shore, A. J. Armstrong, T. W. Flaig, A. Flechon, P. Mainwaring, M. Fleming, J. D. Hainsworth, M. Hirmand, B. Selby, L. Seely, J. S. de Bono, and A. Investigators. 2012. Increased survival with enzalutamide in prostate cancer after chemotherapy. *The New England journal of medicine* 367: 1187-1197.

130. Tran, C., 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, and C. L.

Sawyers. 2009. Development of a second-generation antiandrogen for treatment of advanced prostate cancer. *Science* 324: 787-790.

131. Beer, T. M., A. J. Armstrong, D. E. Rathkopf, Y. Loriot, C. N. Sternberg, C. S. Higano, P. Iversen, S. Bhattacharya, J. Carles, S. Chowdhury, I. D. Davis, J. S. de Bono, C. P. Evans, K. Fizazi, A. M. Joshua, C. S. Kim, G. Kimura, P. Mainwaring, H. Mansbach, K. Miller, S. B. Noonberg, F. Perabo, D. Phung, F. Saad, H. I. Scher, M. E. Taplin, P. M. Venner, B. Tombal, and P. Investigators. 2014. Enzalutamide in metastatic prostate cancer before chemotherapy. *The New England journal of medicine* 371: 424-433.

132. van der Kwast, T. H., J. Schalken, J. A. Ruizeveld de Winter, C. C. van Vroonhoven, E. Mulder, W. Boersma, and J. Trapman. 1991. Androgen receptors in endocrine-therapy-resistant human prostate cancer. *International journal of cancer* 48: 189-193.

133. Ruizeveld de Winter, J. A., P. J. Janssen, H. M. Sleddens, M. C. Verleun-Mooijman, J. Trapman, A. O. Brinkmann, A. B. Santerse, F. H. Schroder, and T. H. van der Kwast. 1994. Androgen receptor status in localized and locally progressive hormone refractory human prostate cancer. *The American journal of pathology* 144: 735-746.

134. Sadi, M. V., and E. R. Barrack. 1993. Image analysis of androgen receptor immunostaining in metastatic prostate cancer. Heterogeneity as a predictor of response to hormonal therapy. *Cancer* 71: 2574-2580.

135. de Winter, J. A., J. Trapman, A. O. Brinkmann, W. J. Boersma, E. Mulder, F. H. Schroeder, E. Claassen, and T. H. van der Kwast. 1990. Androgen receptor

heterogeneity in human prostatic carcinomas visualized by immunohistochemistry. *The Journal of pathology* 160: 329-332.

136. Masai, M., H. Sumiya, S. Akimoto, R. Yatani, C. S. Chang, S. S. Liao, and J. Shimazaki. 1990. Immunohistochemical study of androgen receptor in benign hyperplastic and cancerous human prostates. *The Prostate* 17: 293-300.

137. Chodak, G. W., D. M. Kranc, L. A. Puy, H. Takeda, K. Johnson, and C. Chang. 1992. Nuclear localization of androgen receptor in heterogeneous samples of normal, hyperplastic and neoplastic human prostate. *The Journal of urology* 147: 798-803.

138. Shah, R. B., R. Mehra, A. M. Chinnaiyan, R. Shen, D. Ghosh, M. Zhou, G. R. Macvicar, S. Varambally, J. Harwood, T. A. Bismar, R. Kim, M. A. Rubin, and K. J. Pienta. 2004. Androgen-independent prostate cancer is a heterogeneous group of diseases: lessons from a rapid autopsy program. *Cancer research* 64: 9209-9216.

139. Brocks, D., Y. Assenov, S. Minner, O. Bogatyrova, R. Simon, C. Koop, C. Oakes, M. Zucknick, D. B. Lipka, J. Weischenfeldt, L. Feuerbach, R. Cowper-Sal Lari, M. Lupien, B. Brors, J. Korbel, T. Schlomm, A. Tanay, G. Sauter, C. Gerhauser, C. Plass, and I. E. O. P. C. Project. 2014. Intratumor DNA methylation heterogeneity reflects clonal evolution in aggressive prostate cancer. *Cell reports* 8: 798-806.

140. Wang, Q., W. Li, Y. Zhang, X. Yuan, K. Xu, J. Yu, Z. Chen, R. Beroukhim, H. Wang, M. Lupien, T. Wu, M. M. Regan, C. A. Meyer, J. S. Carroll, A. K. Manrai, O. A. Janne, S. P. Balk, R. Mehra, B. Han, A. M. Chinnaiyan, M. A. Rubin, L.

True, M. Fiorentino, C. Fiore, M. Loda, P. W. Kantoff, X. S. Liu, and M. Brown. 2009. Androgen receptor regulates a distinct transcription program in androgen-independent prostate cancer. *Cell* 138: 245-256.

141. Cieslik, M., and A. M. Chinnaiyan. 2018. Cancer transcriptome profiling at the juncture of clinical translation. *Nature reviews. Genetics* 19: 93-109.

142. Ben-Porath, I., M. W. Thomson, V. J. Carey, R. Ge, G. W. Bell, A. Regev, and R. A. Weinberg. 2008. An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. *Nature genetics* 40: 499-507.

143. McLean, C. Y., D. Bristor, M. Hiller, S. L. Clarke, B. T. Schaar, C. B. Lowe, A. M. Wenger, and G. Bejerano. 2010. GREAT improves functional interpretation of cis-regulatory regions. *Nature biotechnology* 28: 495-501.

144. Subramanian, A., P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, and J. P. Mesirov. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* 102: 15545-15550.

145. Eppert, K., K. Takenaka, E. R. Lechman, L. Waldron, B. Nilsson, P. van Galen, K. H. Metzeler, A. Poeppl, V. Ling, J. Beyene, A. J. Canty, J. S. Danska, S. K. Bohlander, C. Buske, M. D. Minden, T. R. Golub, I. Jurisica, B. L. Ebert, and J. E. Dick. 2011. Stem cell gene expression programs influence clinical outcome in human leukemia. *Nature medicine* 17: 1086-1093.

146. Merlos-Suarez, A., F. M. Barriga, P. Jung, M. Iglesias, M. V. Cespedes, D. Rossell, M. Sevillano, X. Hernando-Momblona, V. da Silva-Diz, P. Munoz, H.

Clevers, E. Sancho, R. Mangues, and E. Batlle. 2011. The intestinal stem cell signature identifies colorectal cancer stem cells and predicts disease relapse. *Cell stem cell* 8: 511-524.

147. Guryanova, O. A., Q. Wu, L. Cheng, J. D. Lathia, Z. Huang, J. Yang, J. MacSwords, C. E. Eyler, R. E. McLendon, J. M. Heddleston, W. Shou, D. Hambardzumyan, J. Lee, A. B. Hjelmeland, A. E. Sloan, M. Bredel, G. R. Stark, J. N. Rich, and S. Bao. 2011. Nonreceptor tyrosine kinase BMX maintains self-renewal and tumorigenic potential of glioblastoma stem cells by activating STAT3. *Cancer cell* 19: 498-511.

148. Patel, A. P., I. Tirosh, J. J. Trombetta, A. K. Shalek, S. M. Gillespie, H. Wakimoto, D. P. Cahill, B. V. Nahed, W. T. Curry, R. L. Martuza, D. N. Louis, O. Rozenblatt-Rosen, M. L. Suva, A. Regev, and B. E. Bernstein. 2014. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* 344: 1396-1401.

149. Robinson, D., 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, and A. M. Chinnaiyan. 2015. Integrative clinical genomics of advanced prostate cancer. *Cell* 161: 1215-1228.

150. Rycaj, K., and D. G. Tang. 2015. Cell-of-Origin of Cancer versus Cancer Stem Cells: Assays and Interpretations. *Cancer research* 75: 4003-4011.

151. Jeter, C. R., B. Liu, Y. Lu, H. P. Chao, D. Zhang, X. Liu, X. Chen, Q. Li, K. Rycaj, T. Calhoun-Davis, L. Yan, Q. Hu, J. Wang, J. Shen, S. Liu, and D. G. Tang. 2016. NANOG reprograms prostate cancer cells to castration resistance via dynamically repressing and engaging the AR/FOXA1 signaling axis. *Cell discovery* 2: 16041.

152. Li Q, D. Q., Chao HP, Liu X, Lu Y, Lin K, Liu B, Tang GW, Zhang D, Tracz A, Jeter C, Rycaj K, Calhoun-Davis T, Huang J, Rubin MA, Beltran H, Shen J, Chatta G, Puzanov I, Mohler J, Wang J, Zhao R, Kirk J, Chen X, and Tang DG. 2018. Linking prostate cancer cell AR heterogeneity to distinct castration and Enzalutamide responses. *Nature communications*.

153. Edgar, R., M. Domrachev, and A. E. Lash. 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic acids research* 30: 207-210.

154. Brazma, A., P. Hingamp, J. Quackenbush, G. Sherlock, P. Spellman, C. Stoeckert, J. Aach, W. Ansorge, C. A. Ball, H. C. Causton, T. Gaasterland, P. Glenisson, F. C. Holstege, I. F. Kim, V. Markowitz, J. C. Matese, H. Parkinson,

A. Robinson, U. Sarkans, S. Schulze-Kremer, J. Stewart, R. Taylor, J. Vilo, and M. Vingron. 2001. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nature genetics* 29: 365-371.

155. Anders, S., P. T. Pyl, and W. Huber. 2015. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31: 166-169.

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

157. Anders, S., and W. Huber. 2010. Differential expression analysis for sequence count data. *Genome biology* 11: R106.

158. Yan, L., L. Tian, and S. Liu. 2015. Combining large number of weak biomarkers based on AUC. *Statistics in medicine* 34: 3811-3830.

159. Gene Ontology, C. 2015. Gene Ontology Consortium: going forward. *Nucleic acids research* 43: D1049-1056.

160. Jantzen, S. G., B. J. Sutherland, D. R. Minkley, and B. F. Koop. 2011. GO Trimming: Systematically reducing redundancy in large Gene Ontology datasets. *BMC research notes* 4: 267.

161. Feiner, H. D., and R. Gonzalez. 1986. Carcinoma of the prostate with atypical immunohistological features. Clinical and histologic correlates. *The American journal of surgical pathology* 10: 765-770.

162. Roudier, M. P., L. D. True, C. S. Higano, H. Vesselle, W. Ellis, P. Lange, and R. L. Vessella. 2003. Phenotypic heterogeneity of end-stage prostate carcinoma metastatic to bone. *Human pathology* 34: 646-653.

163. Easwaran, H., S. E. Johnstone, L. Van Neste, J. Ohm, T. Mosbruger, Q. Wang, M. J. Aryee, P. Joyce, N. Ahuja, D. Weisenberger, E. Collisson, J. Zhu, S. Yegnasubramanian, W. Matsui, and S. B. Baylin. 2012. A DNA hypermethylation module for the stem/progenitor cell signature of cancer. *Genome research* 22: 837-849.

164. Cui, K., C. Zang, T. Y. Roh, D. E. Schones, R. W. Childs, W. Peng, and K. Zhao. 2009. Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation. *Cell stem cell* 4: 80-93.

165. Meissner, A., T. S. Mikkelsen, H. Gu, M. Wernig, J. Hanna, A. Sivachenko, X. Zhang, B. E. Bernstein, C. Nusbaum, D. B. Jaffe, A. Gnirke, R. Jaenisch, and E. S. Lander. 2008. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* 454: 766-770.

166. Hawkins, R. D., G. C. Hon, L. K. Lee, Q. Ngo, R. Lister, M. Pelizzola, L. E. Edsall, S. Kuan, Y. Luu, S. Klugman, J. Antosiewicz-Bourget, Z. Ye, C. Espinoza, S. Agarwahl, L. Shen, V. Ruotti, W. Wang, R. Stewart, J. A. Thomson, J. R. Ecker, and B. Ren. 2010. Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell stem cell* 6: 479-491.

167. Fang, F., S. Turcan, A. Rimner, A. Kaufman, D. Giri, L. G. Morris, R. Shen, V. Seshan, Q. Mo, A. Heguy, S. B. Baylin, N. Ahuja, A. Viale, J. Massague, L. Norton, L. T. Vahdat, M. E. Moynahan, and T. A. Chan. 2011. Breast cancer methylomes establish an epigenomic foundation for metastasis. *Science translational medicine* 3: 75ra25.

168. Ogino, S., K. Nosho, G. J. Kirkner, T. Kawasaki, J. A. Meyerhardt, M. Loda, E. L. Giovannucci, and C. S. Fuchs. 2009. CpG island methylator phenotype, microsatellite instability, BRAF mutation and clinical outcome in colon cancer. *Gut* 58: 90-96.

169. Weisenberger, D. J., K. D. Siegmund, M. Campan, J. Young, T. I. Long, M. A. Faasse, G. H. Kang, M. Widschwendter, D. Weener, D. Buchanan, H. Koh, L. Simms, M. Barker, B. Leggett, J. Levine, M. Kim, A. J. French, S. N. Thibodeau, J. Jass, R. Haile, and P. W. Laird. 2006. CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nature genetics* 38: 787-793.

170. Lu, C., S. Venneti, A. Akalin, F. Fang, P. S. Ward, R. G. Dematteo, A. M. Intlekofer, C. Chen, J. Ye, M. Hameed, K. Nafa, N. P. Agaram, J. R. Cross, R. Khanin, C. E. Mason, J. H. Healey, S. W. Lowe, G. K. Schwartz, A. Melnick, and C. B. Thompson. 2013. Induction of sarcomas by mutant IDH2. *Genes & development* 27: 1986-1998.

171. Lu, C., P. S. Ward, G. S. Kapoor, D. Rohle, S. Turcan, O. Abdel-Wahab, C. R. Edwards, R. Khanin, M. E. Figueroa, A. Melnick, K. E. Wellen, D. M. O'Rourke, S. L. Berger, T. A. Chan, R. L. Levine, I. K. Mellinghoff, and C. B. Thompson. 2012. IDH mutation impairs histone demethylation and results in a block to cell differentiation. *Nature* 483: 474-478.

172. Turcan, S., D. Rohle, A. Goenka, L. A. Walsh, F. Fang, E. Yilmaz, C. Campos, A. W. Fabius, C. Lu, P. S. Ward, C. B. Thompson, A. Kaufman, O. Guryanova, R. Levine, A. Heguy, A. Viale, L. G. Morris, J. T. Huse, I. K. Mellinghoff, and T.

A. Chan. 2012. IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. *Nature* 483: 479-483.

173. Sharma, S. V., D. Y. Lee, B. Li, M. P. Quinlan, F. Takahashi, S. Maheswaran, U. McDermott, N. Azizian, L. Zou, M. A. Fischbach, K. K. Wong, K. Brandstetter, B. Wittner, S. Ramaswamy, M. Classon, and J. Settleman. 2010. A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell* 141: 69-80.

174. Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics* 25: 25-29.

175. Rajan, P., I. M. Sudbery, M. E. Villasevil, E. Mui, J. Fleming, M. Davis, I. Ahmad, J. Edwards, O. J. Sansom, D. Sims, C. P. Ponting, A. Heger, R. M. McMenemin, I. D. Pedley, and H. Y. Leung. 2014. Next-generation sequencing of advanced prostate cancer treated with androgen-deprivation therapy. *European urology* 66: 32-39.

176. Voigt, P., W. W. Tee, and D. Reinberg. 2013. A double take on bivalent promoters. *Genes & development* 27: 1318-1338.

177. Craig, A. M., and Y. Kang. 2007. Neurexin-neuroligin signaling in synapse development. *Current opinion in neurobiology* 17: 43-52.

178. Park, S., J. Lee, Y. H. Kim, J. Park, J. W. Shin, and S. Nam. 2016. Clinical Relevance and Molecular Phenotypes in Gastric Cancer, of TP53 Mutations

and Gene Expressions, in Combination With Other Gene Mutations. *Scientific reports* 6: 34822.

179. Ockenga, W., S. Kuhne, S. Bocksberger, A. Banning, and R. Tikkanen. 2013. Non-neuronal functions of the m2 muscarinic acetylcholine receptor. *Genes* 4: 171-197.

180. Wang, N., M. Yao, J. Xu, Y. Quan, K. Zhang, R. Yang, and W. Q. Gao. 2015. Autocrine Activation of CHRM3 Promotes Prostate Cancer Growth and Castration Resistance via CaM/CaMKK-Mediated Phosphorylation of Akt. *Clinical cancer research : an official journal of the American Association for Cancer Research* 21: 4676-4685.

181. Wang, Y., Y. Wang, Q. Liu, G. Xu, F. Mao, T. Qin, H. Teng, W. Cai, P. Yu, T. Cai, M. Zhao, Z. S. Sun, and C. Xie. 2014. Comparative RNA-seq analysis reveals potential mechanisms mediating the conversion to androgen independence in an LNCaP progression cell model. *Cancer letters* 342: 130-138.

182. Easwaran, H., H. C. Tsai, and S. B. Baylin. 2014. Cancer epigenetics: tumor heterogeneity, plasticity of stem-like states, and drug resistance. *Molecular cell* 54: 716-727.

183. Smith, B. A., A. Sokolov, V. Uzunangelov, R. Baertsch, Y. Newton, K. Graim, C. Mathis, D. Cheng, J. M. Stuart, and O. N. Witte. 2015. A basal stem cell signature identifies aggressive prostate cancer phenotypes. *Proceedings of the National Academy of Sciences of the United States of America* 112: E6544-6552.

184. Ring, K. L., L. M. Tong, M. E. Balestra, R. Javier, Y. Andrews-Zwilling, G. Li, D. Walker, W. R. Zhang, A. C. Kreitzer, and Y. Huang. 2012. Direct reprogramming of mouse and human fibroblasts into multipotent neural stem cells with a single factor. *Cell stem cell* 11: 100-109.

185. Chambers, I., D. Colby, M. Robertson, J. Nichols, S. Lee, S. Tweedie, and A. Smith. 2003. Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* 113: 643-655.

186. Shi, Y., H. Inoue, J. C. Wu, and S. Yamanaka. 2017. Induced pluripotent stem cell technology: a decade of progress. *Nature reviews. Drug discovery* 16: 115-130.

187. Saunders, A., F. Faiola, and J. Wang. 2013. Concise review: pursuing self-renewal and pluripotency with the stem cell factor Nanog. *Stem cells* 31: 1227-1236.

188. Costa, Y., J. Ding, T. W. Theunissen, F. Faiola, T. A. Hore, P. V. Shliaha, M. Fidalgo, A. Saunders, M. Lawrence, S. Dietmann, S. Das, D. N. Levasseur, Z. Li, M. Xu, W. Reik, J. C. Silva, and J. Wang. 2013. NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. *Nature* 495: 370-374.

189. Booth, H. A., and P. W. Holland. 2004. Eleven daughters of NANOG. *Genomics* 84: 229-238.

190. Chang, D. F., S. C. Tsai, X. C. Wang, P. Xia, D. Senadheera, and C. Lutzko. 2009. Molecular characterization of the human NANOG protein. *Stem cells* 27: 812-821.

191. Scerbo, P., G. V. Markov, C. Vivien, L. Kodjabachian, B. Demeneix, L. Coen, and F. Girardot. 2014. On the origin and evolutionary history of NANOG. *PloS one* 9: e85104.

192. Ling, G. Q., D. B. Chen, B. Q. Wang, and L. S. Zhang. 2012. Expression of the pluripotency markers Oct3/4, Nanog and Sox2 in human breast cancer cell lines. *Oncology letters* 4: 1264-1268.

193. Zhang, Y., Z. Wang, J. Yu, J. Shi, C. Wang, W. Fu, Z. Chen, and J. Yang. 2012. Cancer stem-like cells contribute to cisplatin resistance and progression in bladder cancer. *Cancer letters* 322: 70-77.

194. Hoei-Hansen, C. E., S. M. Kraggerud, V. M. Abeler, J. Kaern, E. Rajpert-De Meyts, and R. A. Lothe. 2007. Ovarian dysgerminomas are characterised by frequent KIT mutations and abundant expression of pluripotency markers. *Molecular cancer* 6: 12.

195. Miyazawa, K., T. Tanaka, D. Nakai, N. Morita, and K. Suzuki. 2014. Immunohistochemical expression of four different stem cell markers in prostate cancer: High expression of NANOG in conjunction with hypoxia-inducible factor-1alpha expression is involved in prostate epithelial malignancy. *Oncology letters* 8: 985-992.

196. Ambady, S., C. Malcuit, O. Kashpur, D. Kole, W. F. Holmes, E. Hedblom, R. L. Page, and T. Dominko. 2010. Expression of NANOG and NANOGP8 in a variety of undifferentiated and differentiated human cells. *The International journal of developmental biology* 54: 1743-1754.

197. Chiou, S. H., M. L. Wang, Y. T. Chou, C. J. Chen, C. F. Hong, W. J. Hsieh, H. T. Chang, Y. S. Chen, T. W. Lin, H. S. Hsu, and C. W. Wu. 2010. Coexpression of Oct4 and Nanog enhances malignancy in lung adenocarcinoma by inducing cancer stem cell-like properties and epithelial-mesenchymal transdifferentiation. *Cancer research* 70: 10433-10444.

198. Luo, W., S. Li, B. Peng, Y. Ye, X. Deng, and K. Yao. 2013. Embryonic stem cells markers SOX2, OCT4 and Nanog expression and their correlations with epithelial-mesenchymal transition in nasopharyngeal carcinoma. *PloS one* 8: e56324.

199. Watanabe, M., Y. Ohnishi, H. Inoue, M. Wato, A. Tanaka, K. Kakudo, and M. Nozaki. 2014. NANOG expression correlates with differentiation, metastasis and resistance to preoperative adjuvant therapy in oral squamous cell carcinoma. *Oncology letters* 7: 35-40.

200. Imai, T., K. Tamai, S. Oizumi, K. Oyama, K. Yamaguchi, I. Sato, K. Satoh, K. Matsuura, S. Saijo, K. Sugamura, and N. Tanaka. 2013. CD271 defines a stem cell-like population in hypopharyngeal cancer. *PloS one* 8: e62002.

201. Chae, H. D., M. R. Lee, and H. E. Broxmeyer. 2012. 5-Aminoimidazole-4-carboxyamide ribonucleoside induces G(1)/S arrest and Nanog downregulation via p53 and enhances erythroid differentiation. *Stem cells* 30: 140-149.

202. Chen, T., J. Du, and G. Lu. 2012. Cell growth arrest and apoptosis induced by Oct4 or Nanog knockdown in mouse embryonic stem cells: a possible role of Trp53. *Molecular biology reports* 39: 1855-1861.

203. You, J. S., J. K. Kang, D. W. Seo, J. H. Park, J. W. Park, J. C. Lee, Y. J. Jeon, E. J. Cho, and J. W. Han. 2009. Depletion of embryonic stem cell signature by histone deacetylase inhibitor in NCCIT cells: involvement of Nanog suppression. *Cancer research* 69: 5716-5725.

204. Han, M. K., E. K. Song, Y. Guo, X. Ou, C. Mantel, and H. E. Broxmeyer. 2008. SIRT1 regulates apoptosis and Nanog expression in mouse embryonic stem cells by controlling p53 subcellular localization. *Cell stem cell* 2: 241-251.

205. Hasmim, M., M. Z. Noman, Y. Messai, D. Bordereaux, G. Gros, V. Baud, and S. Chouaib. 2013. Cutting edge: Hypoxia-induced Nanog favors the intratumoral infiltration of regulatory T cells and macrophages via direct regulation of TGF-beta1. *Journal of immunology* 191: 5802-5806.

206. Hasmim, M., M. Z. Noman, J. Lauriol, H. Benlalam, A. Mallavialle, F. Rosselli, F. Mami-Chouaib, C. Alcaide-Loridan, and S. Chouaib. 2011. Hypoxia-dependent inhibition of tumor cell susceptibility to CTL-mediated lysis involves NANOG induction in target cells. *Journal of immunology* 187: 4031-4039.

207. Wood, N. J. 2014. Pancreatic cancer: pancreatic tumour formation and recurrence after radiotherapy are blocked by targeting CD44. *Nature reviews. Gastroenterology & hepatology* 11: 73.

208. Shan, J., J. Shen, L. Liu, F. Xia, C. Xu, G. Duan, Y. Xu, Q. Ma, Z. Yang, Q. Zhang, L. Ma, J. Liu, S. Xu, X. Yan, P. Bie, Y. Cui, X. W. Bian, and C. Qian. 2012. Nanog regulates self-renewal of cancer stem cells through the insulin-like growth factor pathway in human hepatocellular carcinoma. *Hepatology* 56: 1004-1014.

209. Kim, D., G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, and S. L. Salzberg. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology* 14: R36.

210. Langmead, B., and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. *Nature methods* 9: 357-359.

211. Setlur, S. R., K. D. Mertz, Y. Hoshida, F. Demichelis, M. Lupien, S. Perner, A. Sboner, Y. Pawitan, O. Andren, L. A. Johnson, J. Tang, H. O. Adami, S. Calza, A. M. Chinnaiyan, D. Rhodes, S. Tomlins, K. Fall, L. A. Mucci, P. W. Kantoff, M. J. Stampfer, S. O. Andersson, E. Varenhorst, J. E. Johansson, M. Brown, T. R. Golub, and M. A. Rubin. 2008. Estrogen-dependent signaling in a molecularly distinct subclass of aggressive prostate cancer. *Journal of the National Cancer Institute* 100: 815-825.

212. Taylor, B. S., N. Schultz, H. Hieronymus, A. Gopalan, Y. Xiao, B. S. Carver, V. K. Arora, P. Kaushik, E. Cerami, B. Reva, Y. Antipin, N. Mitsiades, T. Landers, I. Dolgalev, J. E. Major, M. Wilson, N. D. Socci, A. E. Lash, A. Heguy, J. A. Eastham, H. I. Scher, V. E. Reuter, P. T. Scardino, C. Sander, C. L. Sawyers, and W. L. Gerald. 2010. Integrative genomic profiling of human prostate cancer. *Cancer cell* 18: 11-22.

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

214. Lee, I. H., M. Sohn, H. J. Lim, S. Yoon, H. Oh, S. Shin, J. H. Shin, S. H. Oh, J. Kim, D. K. Lee, D. Y. Noh, D. S. Bae, J. K. Seong, and Y. S. Bae. 2014. Ahnak

functions as a tumor suppressor via modulation of TGFbeta/Smad signaling pathway. *Oncogene* 33: 4675-4684.

215. Kong, B., C. W. Michalski, X. Hong, N. Valkovskaya, S. Rieder, I. Abiatari, S. Streit, M. Erkan, I. Esposito, H. Friess, and J. Kleeff. 2010. AZGP1 is a tumor suppressor in pancreatic cancer inducing mesenchymal-to-epithelial transdifferentiation by inhibiting TGF-beta-mediated ERK signaling. *Oncogene* 29: 5146-5158.

216. Bhatia-Gaur, R., A. A. Donjacour, P. J. Sciavolino, M. Kim, N. Desai, P. Young, C. R. Norton, T. Gridley, R. D. Cardiff, G. R. Cunha, C. Abate-Shen, and M. M. Shen. 1999. Roles for Nkx3.1 in prostate development and cancer. *Genes & development* 13: 966-977.

217. Pascal, L. E., K. Z. Masoodi, J. Liu, X. Qiu, Q. Song, Y. Wang, Y. Zang, T. Yang, Y. Wang, L. H. Rigatti, U. Chandran, L. M. Colli, R. Z. N. Vencio, Y. Lu, J. Zhang, and Z. Wang. 2017. Conditional deletion of ELL2 induces murine prostate intraepithelial neoplasia. *The Journal of endocrinology* 235: 123-136.

218. Jones, E., H. Pu, and N. Kyprianou. 2009. Targeting TGF-beta in prostate cancer: therapeutic possibilities during tumor progression. *Expert opinion on therapeutic targets* 13: 227-234.

219. Steckelbroeck, S., Y. Jin, S. Gopishetty, B. Oyesanmi, and T. M. Penning. 2004. Human cytosolic 3alpha-hydroxysteroid dehydrogenases of the aldo-keto reductase superfamily display significant 3beta-hydroxysteroid dehydrogenase activity: implications for steroid hormone metabolism and action. *The Journal of biological chemistry* 279: 10784-10795.

220. Chouinard, S., O. Barbier, and A. Belanger. 2007. UDP-glucuronosyltransferase 2B15 (UGT2B15) and UGT2B17 enzymes are major determinants of the androgen response in prostate cancer LNCaP cells. *The Journal of biological chemistry* 282: 33466-33474.

221. Mostaghel, E. A. 2013. Steroid hormone synthetic pathways in prostate cancer. *Translational andrology and urology* 2: 212-227.

222. Schiffer, L., W. Arlt, and K. H. Storbeck. 2018. Intracrine androgen biosynthesis, metabolism and action revisited. *Molecular and cellular endocrinology* 465: 4-26.

223. Pascal, L. E., A. J. Oudes, T. W. Petersen, Y. A. Goo, L. S. Walashek, L. D. True, and A. Y. Liu. 2007. Molecular and cellular characterization of ABCG2 in the prostate. *BMC urology* 7: 6.

224. Mo, W., and J. T. Zhang. 2012. Human ABCG2: structure, function, and its role in multidrug resistance. *International journal of biochemistry and molecular biology* 3: 1-27.

225. Huss, W. J., D. R. Gray, N. M. Greenberg, J. L. Mohler, and G. J. Smith. 2005. Breast cancer resistance protein-mediated efflux of androgen in putative benign and malignant prostate stem cells. *Cancer research* 65: 6640-6650.

226. Warren, A. J., W. H. Colledge, M. B. Carlton, M. J. Evans, A. J. Smith, and T. H. Rabbitts. 1994. The oncogenic cysteine-rich LIM domain protein rbtn2 is essential for erythroid development. *Cell* 78: 45-57.

227. Ma, S., X. Y. Guan, P. S. Beh, K. Y. Wong, Y. P. Chan, H. F. Yuen, J. Vielkind, and K. W. Chan. 2007. The significance of LMO2 expression in the progression of prostate cancer. *The Journal of pathology* 211: 278-285.

228. Armstrong, L., G. Saretzki, H. Peters, I. Wappler, J. Evans, N. Hole, T. von Zglinicki, and M. Lako. 2005. Overexpression of telomerase confers growth advantage, stress resistance, and enhanced differentiation of ESCs toward the hematopoietic lineage. *Stem cells* 23: 516-529.

229. Ohyashiki, J. H., G. Sashida, T. Tauchi, and K. Ohyashiki. 2002. Telomeres and telomerase in hematologic neoplasia. *Oncogene* 21: 680-687.

230. Yamada, O. 1996. Telomeres and telomerase in human hematologic neoplasia. *International journal of hematology* 64: 87-99.

231. Zhang, K., Y. Guo, X. Wang, H. Zhao, Z. Ji, C. Cheng, L. Li, Y. Fang, D. Xu, H. H. Zhu, and W. Q. Gao. 2017. WNT/beta-Catenin Directs Self-Renewal Symmetric Cell Division of hTERT(high) Prostate Cancer Stem Cells. *Cancer research* 77: 2534-2547.

232. Morimoto, K., T. Tanaka, Y. Nitta, K. Ohnishi, H. Kawashima, and T. Nakatani. 2014. NEDD9 crucially regulates TGF-beta-triggered epithelial-mesenchymal transition and cell invasion in prostate cancer cells: involvement in cancer progressiveness. *The Prostate* 74: 901-910.

233. Valderrama, F., S. Thevapala, and A. J. Ridley. 2012. Radixin regulates cell migration and cell-cell adhesion through Rac1. *Journal of cell science* 125: 3310-3319.

234. Comstock, C. E., and K. E. Knudsen. 2007. The complex role of AR signaling after cytotoxic insult: implications for cell-cycle-based chemotherapeutics. *Cell cycle* 6: 1307-1313.

235. Hao, Z., H. Zhang, and J. Cowell. 2012. Ubiquitin-conjugating enzyme UBE2C: molecular biology, role in tumorigenesis, and potential as a biomarker. *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine* 33: 723-730.

236. Evans, T., E. T. Rosenthal, J. Youngblom, D. Distel, and T. Hunt. 1983. Cyclin: a protein specified by maternal mRNA in sea urchin eggs that is destroyed at each cleavage division. *Cell* 33: 389-396.

237. Elowe, S. 2011. Bub1 and BubR1: at the interface between chromosome attachment and the spindle checkpoint. *Molecular and cellular biology* 31: 3085-3093.

238. Fu, J., M. Bian, Q. Jiang, and C. Zhang. 2007. Roles of Aurora kinases in mitosis and tumorigenesis. *Molecular cancer research : MCR* 5: 1-10.

239. Dang, C. V. 2012. MYC on the path to cancer. *Cell* 149: 22-35.

240. Nupponen, N. N., L. Kakkola, P. Koivisto, and T. Visakorpi. 1998. Genetic alterations in hormone-refractory recurrent prostate carcinomas. *The American journal of pathology* 153: 141-148.

241. Visakorpi, T., A. H. Kallioniemi, A. C. Syvanen, E. R. Hyytinen, R. Karhu, T. Tammela, J. J. Isola, and O. P. Kallioniemi. 1995. Genetic changes in primary and recurrent prostate cancer by comparative genomic hybridization. *Cancer research* 55: 342-347.

242. Bernard, D., A. Pourtier-Manzanedo, J. Gil, and D. H. Beach. 2003. Myc confers androgen-independent prostate cancer cell growth. *The Journal of clinical investigation* 112: 1724-1731.

243. Yang, Y. A., and J. Yu. 2015. Current perspectives on FOXA1 regulation of androgen receptor signaling and prostate cancer. *Genes & diseases* 2: 144-151.

244. Lupien, M., and M. Brown. 2009. Cistromics of hormone-dependent cancer. *Endocrine-related cancer* 16: 381-389.

245. Gao, N., J. Zhang, M. A. Rao, T. C. Case, J. Mirosevich, Y. Wang, R. Jin, A. Gupta, P. S. Rennie, and R. J. Matusik. 2003. The role of hepatocyte nuclear factor-3 alpha (Forkhead Box A1) and androgen receptor in transcriptional regulation of prostatic genes. *Molecular endocrinology* 17: 1484-1507.

246. Clark, K. L., E. D. Halay, E. Lai, and S. K. Burley. 1993. Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature* 364: 412-420.

247. Sahu, B., M. Laakso, K. Ovaska, T. Mirtti, J. Lundin, A. Rannikko, A. Sankila, J. P. Turunen, M. Lundin, J. Konsti, T. Vesterinen, S. Nordling, O. Kallioniemi, S. Hautaniemi, and O. A. Janne. 2011. Dual role of FoxA1 in androgen receptor binding to chromatin, androgen signalling and prostate cancer. *The EMBO journal* 30: 3962-3976.

248. Robinson, J. L., and J. S. Carroll. 2012. FoxA1 is a key mediator of hormonal response in breast and prostate cancer. *Frontiers in endocrinology* 3: 68.

249. Wang, D., I. Garcia-Bassets, C. Benner, W. Li, X. Su, Y. Zhou, J. Qiu, W. Liu, M. U. Kaikkonen, K. A. Ohgi, C. K. Glass, M. G. Rosenfeld, and X. D. Fu. 2011. Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* 474: 390-394.

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

251. Serandour, A. A., S. Avner, F. Percevault, F. Demay, M. Bizot, C. Lucchetti-Miganeh, F. Barloy-Hubler, M. Brown, M. Lupien, R. Metivier, G. Salbert, and J. Eeckhoute. 2011. Epigenetic switch involved in activation of pioneer factor FOXA1-dependent enhancers. *Genome research* 21: 555-565.

252. Deng, Q., and D. G. Tang. 2015. Androgen receptor and prostate cancer stem cells: biological mechanisms and clinical implications. *Endocrine-related cancer* 22: T209-220.

253. Hobisch, A., Z. Culig, C. Radmayr, G. Bartsch, H. Klocker, and A. Hittmair. 1995. Distant metastases from prostatic carcinoma express androgen receptor protein. *Cancer research* 55: 3068-3072.

254. Mostaghel, E. A., S. T. Page, D. W. Lin, L. Fazli, I. M. Coleman, L. D. True, B. Knudsen, D. L. Hess, C. C. Nelson, A. M. Matsumoto, W. J. Bremner, M. E. Gleave, and P. S. Nelson. 2007. Intraprostatic androgens and androgen-regulated gene expression persist after testosterone suppression: therapeutic implications for castration-resistant prostate cancer. *Cancer research* 67: 5033-5041.

255. Harrow, J., A. Frankish, J. M. Gonzalez, E. Tapanari, M. Diekhans, F. Kokocinski, B. L. Aken, D. Barrell, A. Zadissa, S. Searle, I. Barnes, A. Bignell, V. Boychenko, T. Hunt, M. Kay, G. Mukherjee, J. Rajan, G. Despacio-Reyes, G. Saunders, C. Steward, R. Harte, M. Lin, C. Howald, A. Tanzer, T. Derrien, J. Chrast, N. Walters, S. Balasubramanian, B. Pei, M. Tress, J. M. Rodriguez, I. Ezkurdia, J. van Baren, M. Brent, D. Haussler, M. Kellis, A. Valencia, A. Reymond, M. Gerstein, R. Guigo, and T. J. Hubbard. 2012. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome research* 22: 1760-1774.

256. Sun, Y., and S. Goodison. 2009. Optimizing molecular signatures for predicting prostate cancer recurrence. *The Prostate* 69: 1119-1127.

257. Beltran, H., D. Prandi, J. M. Mosquera, M. Benelli, L. Puca, J. Cyrta, C. Marotz, E. Giannopoulou, B. V. Chakravarthi, S. Varambally, S. A. Tomlins, D. M. Nanus, S. T. Tagawa, E. M. Van Allen, O. Elemento, A. Sboner, L. A. Garraway, M. A. Rubin, and F. Demichelis. 2016. Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. *Nature medicine* 22: 298-305.

258. Kirollos, M. M. 2009. Re: Urs E. Studer, Laurence Collette, Peter Whelan, et al. Using PSA to guide timing of androgen deprivation in patients with T0-4 N0-2 M0 prostate cancer not suitable for local curative treatment (EORTC 30891). Eur Urol 2008;53:941-9 and re: John Anderson. Androgen-deprivation therapy--It's all a matter of timing. Eur Urol 2008;53:869-71. *European urology* 55: e40-42; author reply e45-46.

259. Strasser, A., A. W. Harris, M. L. Bath, and S. Cory. 1990. Novel primitive lymphoid tumours induced in transgenic mice by cooperation between myc and bcl-2. *Nature* 348: 331-333.

260. Delmore, J. E., G. C. Issa, M. E. Lemieux, P. B. Rahl, J. Shi, H. M. Jacobs, E. Kastritis, T. Gilpatrick, R. M. Paranal, J. Qi, M. Chesi, A. C. Schinzel, M. R. McKeown, T. P. Heffernan, C. R. Vakoc, P. L. Bergsagel, I. M. Ghobrial, P. G. Richardson, R. A. Young, W. C. Hahn, K. C. Anderson, A. L. Kung, J. E. Bradner, and C. S. Mitsiades. 2011. BET bromodomain inhibition as a therapeutic strategy to target c-Myc. *Cell* 146: 904-917.

261. Liu, J., L. Shi, O. Sartor, and R. Culbertson. 2013. Androgen-deprivation therapy versus radical prostatectomy as monotherapy among clinically localized prostate cancer patients. *OncoTargets and therapy* 6: 725-732.

262. Puhr, M., J. Hoefer, A. Eigentler, C. Ploner, F. Handle, G. Schaefer, J. Kroon, A. Leo, I. Heidegger, I. Eder, Z. Culig, G. Van der Pluijm, and H. Klocker. 2018. The Glucocorticoid Receptor Is a Key Player for Prostate Cancer Cell Survival and a Target for Improved Antiandrogen Therapy. *Clinical cancer research : an official journal of the American Association for Cancer Research* 24: 927-938.

263. Beltran, H., D. S. Rickman, K. Park, S. S. Chae, A. Sboner, T. Y. MacDonald, Y. Wang, K. L. Sheikh, S. Terry, S. T. Tagawa, R. Dhir, J. B. Nelson, A. de la Taille, Y. Allory, M. B. Gerstein, S. Perner, K. J. Pienta, A. M. Chinnaiyan, Y. Wang, C. C. Collins, M. E. Gleave, F. Demichelis, D. M. Nanus, and M. A. Rubin. 2011. Molecular characterization of neuroendocrine prostate cancer and identification of new drug targets. *Cancer discovery* 1: 487-495.

264. Wu, Q., X. Chen, J. Zhang, Y. H. Loh, T. Y. Low, W. Zhang, W. Zhang, S. K. Sze, B. Lim, and H. H. Ng. 2006. Sall4 interacts with Nanog and co-occupies Nanog genomic sites in embryonic stem cells. *The Journal of biological chemistry* 281: 24090-24094.

265. Lin, F., and H. Liu. 2014. Immunohistochemistry in undifferentiated neoplasm/tumor of uncertain origin. *Archives of pathology & laboratory medicine* 138: 1583-1610.

266. Miettinen, M., Z. Wang, P. A. McCue, M. Sarlomo-Rikala, J. Rys, W. Biernat, J. Lasota, and Y. S. Lee. 2014. SALL4 expression in germ cell and non-germ cell tumors: a systematic immunohistochemical study of 3215 cases. *The American journal of surgical pathology* 38: 410-420.

267. Liu, K. F., and Y. X. Shan. 2016. Effects of siRNA-mediated silencing of Sal-like 4 expression on proliferation and apoptosis of prostate cancer C4-2 cells. *Genetics and molecular research : GMR* 15.

268. He, J., M. Zhou, X. Chen, D. Yue, L. Yang, G. Qin, Z. Zhang, Q. Gao, D. Wang, C. Zhang, L. Huang, L. Wang, B. Zhang, J. Yu, and Y. Zhang. 2016. Inhibition of SALL4 reduces tumorigenicity involving epithelial-mesenchymal transition via Wnt/beta-catenin pathway in esophageal squamous cell carcinoma. *Journal of experimental & clinical cancer research : CR* 35: 98.

269. Jozwik, K. M., I. Chernukhin, A. A. Serandour, S. Nagarajan, and J. S. Carroll. 2016. FOXA1 Directs H3K4 Monomethylation at Enhancers via Recruitment of the Methyltransferase MLL3. *Cell reports* 17: 2715-2723.

270. Pomerantz, M. M., 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, and M. L. Freedman. 2015. The androgen receptor cistrome is extensively reprogrammed in human prostate tumorigenesis. *Nature genetics* 47: 1346-1351.

271. Sharifi, N. 2014. Steroid receptors aplenty in prostate cancer. *The New England journal of medicine* 370: 970-971.

272. Nguyen, H. G., J. C. Yang, H. J. Kung, X. B. Shi, D. Tilki, P. N. Lara, Jr., R. W. DeVere White, A. C. Gao, and C. P. Evans. 2014. Targeting autophagy overcomes Enzalutamide resistance in castration-resistant prostate cancer cells and improves therapeutic response in a xenograft model. *Oncogene* 33: 4521-4530.

273. Korpal, M., J. M. Korn, X. Gao, D. P. Rakiec, D. A. Ruddy, S. Doshi, J. Yuan, S. G. Kovats, S. Kim, V. G. Cooke, J. E. Monahan, F. Stegmeier, T. M. Roberts, W. R. Sellers, W. Zhou, and P. Zhu. 2013. An F876L mutation in androgen receptor confers genetic and phenotypic resistance to MDV3100 (enzalutamide). *Cancer discovery* 3: 1030-1043.

274. Bayat Mokhtari, R., T. S. Homayouni, N. Baluch, E. Morgatskaya, S. Kumar, B. Das, and H. Yeger. 2017. Combination therapy in combating cancer. *Oncotarget* 8: 38022-38043.

275. Huang, L., Y. Pu, W. Y. Hu, L. Birch, D. Luccio-Camelo, T. Yamaguchi, and G. S. Prins. 2009. The role of Wnt5a in prostate gland development. *Developmental biology* 328: 188-199.

276. Lee, G. T., D. I. Kang, Y. S. Ha, Y. S. Jung, J. Chung, K. Min, T. H. Kim, K. H. Moon, J. M. Chung, D. H. Lee, W. J. Kim, and I. Y. Kim. 2014. Prostate cancer bone metastases acquire resistance to androgen deprivation via WNT5A-mediated BMP-6 induction. *British journal of cancer* 110: 1634-1644.

277. Lee, G. T., S. J. Kwon, J. Kim, Y. S. Kwon, N. Lee, J. H. Hong, C. Jamieson, W. J. Kim, and I. Y. Kim. 2018. WNT5A induces castration-resistant prostate cancer via CCL2 and tumour-infiltrating macrophages. *British journal of cancer* 118: 670-678.

278. Grindel, B. J., J. R. Martinez, T. V. Tellman, D. A. Harrington, H. Zafar, L. Nakhleh, L. W. Chung, and M. C. Farach-Carson. 2018. Matrilysin/MMP-7 Cleavage of Perlecan/HSPG2 Complexed with Semaphorin 3A Supports FAK-Mediated Stromal Invasion by Prostate Cancer Cells. *Scientific reports* 8: 7262.

279. Peacock, J. W., A. Takeuchi, N. Hayashi, L. Liu, K. J. Tam, N. Al Nakouzi, N. Khazamipour, T. Tombe, T. Dejima, K. C. Lee, M. Shiota, D. Thaper, W. C. Lee, D. H. Hui, H. Kuruma, L. Ivanova, P. Yenki, I. Z. Jiao, S. Khosravi, A. L. Mui, L. Fazli, A. Zoubeidi, M. Daugaard, M. E. Gleave, and C. J. Ong. 2018. SEMA3C drives cancer growth by transactivating multiple receptor tyrosine kinases via Plexin B1. *EMBO molecular medicine* 10: 219-238.

280. Bender, R. J., and F. Mac Gabhann. 2015. Dysregulation of the vascular endothelial growth factor and semaphorin ligand-receptor families in prostate cancer metastasis. *BMC systems biology* 9: 55.

281. Han, W., S. Gao, D. Barrett, M. Ahmed, D. Han, J. A. Macoska, H. H. He, and C. Cai. 2018. Reactivation of androgen receptor-regulated lipid biosynthesis

drives the progression of castration-resistant prostate cancer. *Oncogene* 37: 710-721.

282. Penning, T. M. 2014. Androgen biosynthesis in castration-resistant prostate cancer. *Endocrine-related cancer* 21: T67-78.

283. Hamid, A. R., M. J. Pfeiffer, G. W. Verhaegh, E. Schaafsma, A. Brandt, F. C. Sweep, J. P. Sedelaar, and J. A. Schalken. 2013. Aldo-keto reductase family 1 member C3 (AKR1C3) is a biomarker and therapeutic target for castration-resistant prostate cancer. *Molecular medicine* 18: 1449-1455.

284. Sun, S. Q., X. Gu, X. S. Gao, Y. Li, H. Yu, W. Xiong, H. Yu, W. Wang, Y. Li, Y. Teng, and D. Zhou. 2016. Overexpression of AKR1C3 significantly enhances human prostate cancer cells resistance to radiation. *Oncotarget* 7: 48050-48058.

285. Vainio, P., S. Gupta, K. Ketola, T. Mirtti, J. P. Mpindi, P. Kohonen, V. Fey, M. Perala, F. Smit, G. Verhaegh, J. Schalken, K. A. Alanen, O. Kallioniemi, and K. Iljin. 2011. Arachidonic acid pathway members PLA2G7, HPGD, EPHX2, and CYP4F8 identified as putative novel therapeutic targets in prostate cancer. *The American journal of pathology* 178: 525-536.

286. Vainio, P., L. Lehtinen, T. Mirtti, M. Hilvo, T. Seppanen-Laakso, J. Virtanen, A. Sankila, S. Nordling, J. Lundin, A. Rannikko, M. Oresic, O. Kallioniemi, and K. Iljin. 2011. Phospholipase PLA2G7, associated with aggressive prostate cancer, promotes prostate cancer cell migration and invasion and is inhibited by statins. *Oncotarget* 2: 1176-1190.

287. Tsujii, M., S. Kawano, S. Tsuji, H. Sawaoka, M. Hori, and R. N. DuBois. 1998. Cyclooxygenase regulates angiogenesis induced by colon cancer cells. *Cell* 93: 705-716.

288. Daikoku, T., D. Wang, S. Tranguch, J. D. Morrow, S. Orsulic, R. N. DuBois, and S. K. Dey. 2005. Cyclooxygenase-1 is a potential target for prevention and treatment of ovarian epithelial cancer. *Cancer research* 65: 3735-3744.

289. Farivar-Mohseni, H., S. J. Kandzari, S. Zaslau, D. R. Riggs, B. J. Jackson, and D. W. McFadden. 2004. Synergistic effects of Cox-1 and -2 inhibition on bladder and prostate cancer in vitro. *American journal of surgery* 188: 505-510.

290. Bluemn, E. G., I. M. Coleman, J. M. Lucas, R. T. Coleman, S. Hernandez-Lopez, R. Tharakan, D. Bianchi-Frias, R. F. Dumpit, A. Kaipainen, A. N. Corella, Y. C. Yang, M. D. Nyquist, E. Mostaghel, A. C. Hsieh, X. Zhang, E. Corey, L. G. Brown, H. M. Nguyen, K. Pienta, M. Ittmann, M. Schweizer, L. D. True, D. Wise, P. S. Rennie, R. L. Vessella, C. Morrissey, and P. S. Nelson. 2017. Androgen Receptor Pathway-Independent Prostate Cancer Is Sustained through FGF Signaling. *Cancer cell* 32: 474-489 e476.

291. Ylitalo, E. B., E. Thysell, E. Jernberg, M. Lundholm, S. Crnalic, L. Egevad, P. Stattin, A. Widmark, A. Bergh, and P. Wikstrom. 2017. Subgroups of Castration-resistant Prostate Cancer Bone Metastases Defined Through an Inverse Relationship Between Androgen Receptor Activity and Immune Response. *European urology* 71: 776-787.

292. Beltran, H., S. Tomlins, A. Aparicio, V. Arora, D. Rickman, G. Ayala, J. Huang, L. True, M. E. Gleave, H. Soule, C. Logothetis, and M. A. Rubin. 2014.

Aggressive variants of castration-resistant prostate cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* 20: 2846-2850.

293. Yuan, T. C., S. Veeramani, and M. F. Lin. 2007. Neuroendocrine-like prostate cancer cells: neuroendocrine transdifferentiation of prostate adenocarcinoma cells. *Endocrine-related cancer* 14: 531-547.

294. Aparicio, A., C. J. Logothetis, and S. N. Maity. 2011. Understanding the lethal variant of prostate cancer: power of examining extremes. *Cancer discovery* 1: 466-468.

295. Magnon, C., S. J. Hall, J. Lin, X. Xue, L. Gerber, S. J. Freedland, and P. S. Frenette. 2013. Autonomic nerve development contributes to prostate cancer progression. *Science* 341: 1236361.

296. Crea, F., L. Sun, A. Mai, Y. T. Chiang, W. L. Farrar, R. Danesi, and C. D. Helgason. 2012. The emerging role of histone lysine demethylases in prostate cancer. *Molecular cancer* 11: 52.

297. Xiang, Y., Z. Zhu, G. Han, H. Lin, L. Xu, and C. D. Chen. 2007. JMJD3 is a histone H3K27 demethylase. *Cell research* 17: 850-857.

298. Costa, V., M. Aprile, R. Esposito, and A. Ciccodicola. 2013. RNA-Seq and human complex diseases: recent accomplishments and future perspectives. *European journal of human genetics : EJHG* 21: 134-142.

299. Bruning, O., W. Rodenburg, P. F. Wackers, C. van Oostrom, M. J. Jonker, R. J. Dekker, H. Rauwerda, W. A. Ensink, A. de Vries, and T. M. Breit. 2016.

Confounding Factors in the Transcriptome Analysis of an In-Vivo Exposure Experiment. *PloS one* 11: e0145252.

300. Shalek, A. K., and M. Benson. 2017. Single-cell analyses to tailor treatments. *Science translational medicine* 9.

301. Chan, I. S., and G. S. Ginsburg. 2011. Personalized medicine: progress and promise. *Annual review of genomics and human genetics* 12: 217-244.

302. Shin, S. H., A. M. Bode, and Z. Dong. 2017. Precision medicine: the foundation of future cancer therapeutics. *NPJ precision oncology* 1: 12.

303. Slamon, D. J., B. Leyland-Jones, S. Shak, H. Fuchs, V. Paton, A. Bajamonde, T. Fleming, W. Eiermann, J. Wolter, M. Pegram, J. Baselga, and L. Norton. 2001. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *The New England journal of medicine* 344: 783-792.

304. Kyriakopoulos, C. E., Y. H. Chen, M. A. Carducci, G. Liu, D. F. Jarrard, N. M. Hahn, D. H. Shevrin, R. Dreicer, M. Hussain, M. Eisenberger, M. Kohli, E. R. Plimack, N. J. Vogelzang, J. Picus, M. M. Cooney, J. A. Garcia, R. S. DiPaola, and C. J. Sweeney. 2018. Chemohormonal Therapy in Metastatic Hormone-Sensitive Prostate Cancer: Long-Term Survival Analysis of the Randomized Phase III E3805 CHAARTED Trial. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 36: 1080-1087.

305. Sweeney, C. J., and D. Chamberlain. 2015. Insights into E3805: the CHAARTED trial. *Future oncology* 11: 897-899.

306. James, N. D., M. R. Sydes, N. W. Clarke, M. D. Mason, D. P. Dearnaley, M. R. Spears, A. W. Ritchie, C. C. Parker, J. M. Russell, G. Attard, J. de Bono, W. Cross, R. J. Jones, G. Thalmann, C. Amos, D. Matheson, R. Millman, M. Alzouebi, S. Beesley, A. J. Birtle, S. Brock, R. Cathomas, P. Chakraborti, S. Chowdhury, A. Cook, T. Elliott, J. Gale, S. Gibbs, J. D. Graham, J. Hetherington, R. Hughes, R. Laing, F. McKinna, D. B. McLaren, J. M. O'Sullivan, O. Parikh, C. Peedell, A. Protheroe, A. J. Robinson, N. Srihari, R. Srinivasan, J. Staffurth, S. Sundar, S. Tolan, D. Tsang, J. Wagstaff, M. K. Parmar, and S. investigators. 2016. Addition of docetaxel, zoledronic acid, or both to first-line long-term hormone therapy in prostate cancer (STAMPEDE): survival results from an adaptive, multiarm, multistage, platform randomised controlled trial. *Lancet* 387: 1163-1177.

307. Fizazi, K., N. Tran, L. Fein, N. Matsubara, A. Rodriguez-Antolin, B. Y. Alekseev, M. Ozguroglu, D. Ye, S. Feyerabend, A. Protheroe, P. De Porre, T. Kheoh, Y. C. Park, M. B. Todd, K. N. Chi, and L. Investigators. 2017. Abiraterone plus Prednisone in Metastatic, Castration-Sensitive Prostate Cancer. *The New England journal of medicine* 377: 352-360.

308. James, N. D., J. S. de Bono, M. R. Spears, N. W. Clarke, M. D. Mason, D. P. Dearnaley, A. W. S. Ritchie, C. L. Amos, C. Gilson, R. J. Jones, D. Matheson, R. Millman, G. Attard, S. Chowdhury, W. R. Cross, S. Gillessen, C. C. Parker, J. M. Russell, D. R. Berthold, C. Brawley, F. Adab, S. Aung, A. J. Birtle, J. Bowen, S. Brock, P. Chakraborti, C. Ferguson, J. Gale, E. Gray, M. Hingorani, P. J. Hoskin, J. F. Lester, Z. I. Malik, F. McKinna, N. McPhail, J. Money-Kyrle,

J. O'Sullivan, O. Parikh, A. Protheroe, A. Robinson, N. N. Srihari, C. Thomas, J. Wagstaff, J. Wylie, A. Zarkar, M. K. B. Parmar, M. R. Sydes, and S. Investigators. 2017. Abiraterone for Prostate Cancer Not Previously Treated with Hormone Therapy. *The New England journal of medicine* 377: 338-351.

309. Colak, S., and J. P. Medema. 2014. Cancer stem cells--important players in tumor therapy resistance. *The FEBS journal* 281: 4779-4791.

310. Chen, K., Y. H. Huang, and J. L. Chen. 2013. Understanding and targeting cancer stem cells: therapeutic implications and challenges. *Acta pharmacologica Sinica* 34: 732-740.

Vita

Hsueh-Ping Chao was born in Pingtung, Taiwan, the daughter of Chia-Chu Chao and Lina Shiang. After completing her work at National Chung Hsing University, Taichung, Taiwan in 2005, she entered Nation Yang-Ming University, Taipei, Taiwan. She received the degree of Master of Sience in July 2007. During the following years, she worked as a research assistant in Taipei Veterans General Hospital, Taipei, Taiwan and Mackay Memorial Hospital, Taipei, Taiwan. In August 2012, Hsueh-Ping came to the United States and was admitted to the Graduate School of Biomedical Sciences (GSBS) at the University of Texas Health Science Center at Houston and the University of Texas MD Anderson Cancer Center. Hsueh-Ping pursued her PhD thesis research work in the laboratory of Dr. Dean Tang at the University of Texas MD Anderson Cancer Center, Science Park, Department of Epigenetics and Molecular Carcinogenesis in Smithville, Texas.

Permanent address:

No.20, Kaifeng St.,

Pingtung City, Pingtung County, 900, Taiwan