



HAL
open science

A Pleiotropic Role of the SWI/SNF Complex in Cancer – Insights From Two Tumor Types: Small Cell Carcinoma of the Ovary, Hypercalcemic Type and Prostatic Carcinoma

Joanna Cyrta

► **To cite this version:**

Joanna Cyrta. A Pleiotropic Role of the SWI/SNF Complex in Cancer – Insights From Two Tumor Types: Small Cell Carcinoma of the Ovary, Hypercalcemic Type and Prostatic Carcinoma. Cancer. Université Paris-Saclay, 2020. English. NNT : 2020UPASL045 . tel-03158745

HAL Id: tel-03158745

<https://theses.hal.science/tel-03158745>

Submitted on 4 Mar 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A pleiotropic role of the SWI/SNF complex in cancer - insights from two tumor types: small cell carcinoma of the ovary, hypercalcemic type and prostatic carcinoma

Thèse de doctorat de l'université Paris-Saclay

Ecole doctorale n°582 : Cancérologie : Biologie - Médecine - Santé (CBMS) Spécialité de doctorat : Sciences de la vie et de la santé
Unité de recherche : Université Paris-Saclay, Institut Gustave Roussy, Inserm, Prédicteurs moléculaires et nouvelles cibles en oncologie, 94800, Villejuif, France
Rétèrent: Faculté de médecine

Thèse soutenue en visioconférence totale le 02/12/2020, par

Joanna CYRTA

Composition du Jury

Gaëlle FROMONT PU-PH, CHRU de Tours, Université de Tours	Présidente, Rapporteur
Jean-Yves PIERGA PU-PH, Institut Curie, Université Paris Sciences Lettres	Rapporteur & Examineur
Franck BOURDEAUT Praticien Spécialiste des CLCC, Institut Curie, Université Paris Sciences Lettres	Examineur
Yohann LORIOT Médecin Chercheur, Institut Gustave Roussy, Université Paris-Saclay	Examineur
Jean-Yves SCOAZEC PU-PH, Institut Gustave Roussy, Université Paris-Saclay	Directeur de thèse
Alexandra LEARY Praticien Spécialiste des CLCC, Institut Gustave Roussy, Université Paris-Saclay	Co-Encadrante
Mark A. RUBIN Professeur, Université de Berne	Invité

Remerciements

Cette thèse a été rendue possible par l'aide et par le soutien de plusieurs personnes, à qui j'adresse ma plus grande reconnaissance.

A mon Directeur de thèse **Monsieur le Professeur Jean-Yves Scoazec**, merci pour votre soutien pédagogique et votre grande bienveillance au cours de ce projet. Je suis également très heureuse de notre collaboration sur les carcinomes neuroendocrines, qui nous permettra je l'espère de continuer à étendre nos explorations à d'autres types tumoraux.

A ma Co-Directrice de thèse **Madame le Docteur Alexandra Leary** - merci de m'avoir fait découvrir la thématique du complexe SWI/SNF au travers la recherche sur les SCCOHT, de m'avoir soutenue toute au long de ce projet de thèse avec la plus grande patience, et pour les commentaires très pertinents sur le contenu de cette dissertation.

I am indebted to **Professor Mark A. Rubin** for giving me the opportunity to join his team, to be involved in high-end research in the field of prostate cancer, and most of all, for allowing me to develop the present study focusing on SWI/SNF. I am very grateful for your outstanding mentorship, both during my time at Weill Cornell and well beyond.

Je remercie Madame le **Professeur Gaëlle Fromont-Hankard** d'avoir accepté de juger cette thèse en tant que rapporteur. Votre expertise de pathologiste et de chercheuse dans le domaine du cancer de la prostate font de votre présence dans ce jury un privilège. Veuillez trouver ici l'expression de ma plus grande reconnaissance.

Je remercie Monsieur le **Professeur Jean-Yves Pierga** d'avoir accepté de juger cette thèse en tant que rapporteur. Face à votre dévouement aux patients, à l'enseignement et à la recherche, je suis honorée de pouvoir vous présenter mon travail. Veuillez trouver ici l'expression de mes meilleurs remerciements.

Je remercie Monsieur le **Docteur Franck Bourdeaut** d'avoir accepté de siéger à mon jury de thèse. J'aimerais également vous remercier de m'avoir donné l'opportunité de collaborer avec votre équipe. Etant très impressionnée par votre recherche sur le complexe SWI/SNF, c'est un énorme plaisir de pouvoir collaborer avec vous et votre équipe.

Je remercie Monsieur le **Docteur Yohann Lorient** d'avoir accepté de siéger à mon jury de thèse en tant qu'examineur. Face à votre expertise en matière des cancers génito-urinaires en tant que clinicien et chercheur et votre engagement dans la science, c'est un privilège de pouvoir vous présenter mon travail.

Je remercie Monsieur le **Docteur François Le Loarer** d'avoir accepté de faire partie de mon Jury de thèse, en regrettant que cela n'a pas pu être possible au final. Connaissant votre travail sur les entités rares et sur le complexe SWI/SNF, je serais ravie de pouvoir discuter mon travail avec vous dans l'avenir.

I am very grateful to Rubin lab members, and in particular to **Anke Augspach** and **Phillip Thienger**, for their priceless help with the SWI/SNF study. I have always been impressed with your enthusiasm, creativity and scientific rigor. I wish you best of luck with your research and many exciting discoveries.

Je suis reconnaissante à ma Cheffe de service, le **Docteur Anne Vincent-Salomon**, pour son soutien et ses encouragements lors de la finalisation de cette thèse.

Je tiens à remercier Monsieur **Patrick Graber** pour son remarquable travail au sein du bureau des internes de l'AP-HP ; ce n'est que grâce à son aide que mon projet de séjour de recherche aux Etats-Unis a pu aboutir.

Je remercie Madame **Léa Poisot** de l'Ecole doctorale pour son aide avec les différentes étapes administratives de mon parcours.

A ma famille et à mes amis,

To my mom - thank you for your unconditional love and support. I admire you greatly and I love you beyond words.

To my aunt Ewa and my cousins, Agata and Magda. I feel grateful and proud to have you as my family.

To my dear grandparents, thank you for giving me a beautiful childhood.

To my godfather Tom - thank you for your precious support throughout all those years.

To Pr and Mrs Thomson, thank you for your unmatched kindness and generosity. I'm humbled to have you as my friends.

To my Białystok friends: Bogunia and Rudy, Sabina and Piotrek, Kama and Gary, Matyśka, Maryś and Konada, Ulaka and Błażej, Ada and Michał, Dzięgiel, Druh - thank you for all the moments we have shared, and for always welcoming me back. I cherish our friendship and I hope we meet again soon.

A Chacha, ma « coloc éternelle », merci d'avoir toujours été là pour moi et pour ton soutien en toutes circonstances. Je t'admire énormément. A Léo - merci pour ta bienveillance et pour m'avoir tenu compagnie lors d'un certain après-midi vélo.

A Joanna et Romain, merci pour votre précieuse amitié et pour m'avoir confié l'immense privilège d'être la marraine d'Ada. Et bien sûr, pour l'esprit surf. A Mathieu, merci pour notre complicité et pour tes remarques avisées - avec Romain - sur le complexe Polycomb. A Chloé, merci pour ton énergie positive.

A Nico, mon INTJ préféré - merci pour ta sincère amitié, qui est très précieuse pour moi et bien sûr, pour le grand intérêt que tu as toujours porté à la biologie cellulaire.

To the Tram/Roosevelt Island/Night Owl tribe: Coraline, Angela and Cesar, Darinka and Hemanth, Sabrina and Boaz, and Hratch: thank you for being such beautiful souls. I was blessed to get to know you during these New York years and I hope we meet again soon.

To Chris and Ajsza - thank you for your friendship, and for sharing pearls of life wisdom during our Central Park runs - and beyond.

A Nadia et Guillermo - merci pour votre gentillesse et votre bienveillance.

A ma coloc Justine, merci pour ton énergie positive hors-pair.

To Moritz, my Mate. Thank you for believing in me (and for your interest in "Swiss knife"). I feel truly grateful for having you in my life.

Table of Contents

Part I. Introduction	1
1. The SWI/SNF complex: a general description.	1
2. Different forms of mSWI/SNF exist through combinatorial subunit assembly.....	4
3. Role of SWI/SNF in embryonic development, cell differentiation and cell reprogramming.	8
4. The SWI/SNF complex and cancer	11
5. Aims of this research.....	23
6. Related publications	24
Part II. Small cell carcinoma of the ovary of the hypercalcemic type (SCCOHT): comprehensive molecular characterization of a malignancy driven by loss-of-function SWI/SNF alterations..	26
1. Background.....	26
2. Results	27
3. Discussion	53
4. Methods	69
Part III. The SWI/SNF complex has tumor-promoting functions in prostate cancer.....	75
1. Background.....	75
2. Results	79
3. Discussion	121
4. Methods	137
Part IV. Discussion and conclusions: a pleiotropic role of SWI/SNF in cancer.	149
References	153
Appendix I: Publications	177
Appendix II: Other paper contributions during this PhD.....	178

Part I. Introduction

1. The SWI/SNF complex: a general description.

Gene expression and chromosome functions are tightly regulated by chromatin structure. Chromatin remodeling processes include the sliding or removal of histones, and incorporation of histone variants (Clapier et al., 2017). These actions are carried out by multi-protein chromatin remodeling complexes; it has been proposed that because chromatin remodeling requires a series of events that must be carefully coordinated in space and in time, protein complexes allow to accomplish such processes in a much more efficient way than single proteins would (Kadoch & Crabtree, 2015). Chromatin remodeling complexes are grouped into four major families: SWI/SNF, INO80, ISWI (Imitation SWI) and CHD/M-2 (Chromodomain Helicase DNA-binding)(Mashtalir et al., 2018). All of these four families contain Snf2-like ATP-ase subunits, which are capable of ATP-dependent chromatin remodeling leading to changes in chromatin accessibility (Kadoch & Crabtree, 2015), but they differ in their structures and their specific activities (Clapier et al., 2017).

The SWI/SNF chromatin remodeling proteins were first discovered in yeast, when two independent mutation screens showed that mutation of genes involved in mating-type switching (Switch, i.e. SWI genes) or growth on sucrose (Sucrosis Non-Fermenting, i.e. SNF genes) resulted in redundant phenotypes (Neugeborn & Carlson, 1984; Peterson & Herskowitz, 1992). The SWI/SNF family is conserved throughout evolution, and homologues exist in plants, flies and mammals (Hargreaves & Crabtree, 2011). In *Drosophila*, the SWI/SNF complex corresponds to the Trithorax-group proteins (TrxG), which control correct body segmentation by activating Hox genes (Ingham, 1983).

Mammalian Switch Sucrose Non-Fermenting (mSWI/SNF) complexes, also known as Brg/Brahma-associated factor (BAF) complexes, represent approximately 1-1.5 MDa entities composed of ~11-15 protein subunits (although this number is variable, as discussed hereafter)

(Kadoch & Crabtree, 2015). Most complexes comprise the following highly conserved subunits, termed “core subunits”: one of the two possible catalytic subunits with ATP-ase activity - SMARCA4 (BRG1) or SMARCA2 (BRM); SMARCB1 (BAF47/SNF5/INI-1); SMARCC1 (BAF155); and SMARCC2 (BAF170). The latter two have been shown to form dimers: SMARCC1/SMARCC2 heterodimers in differentiated cells and SMARCC1/SMARCC1 homodimers in embryonic stem cells (please see below)(Mashtalir et al., 2018). In addition, the complexes incorporate accessory subunits. Importantly, most subunits have one or sometimes more possible paralog(s), encoded by different genes; such paralogs are mutually exclusive within the complex (**Fig.1**). For example, each complex will only incorporate one of the two possible catalytic subunits: either SMARCA4 or SMARCA2.

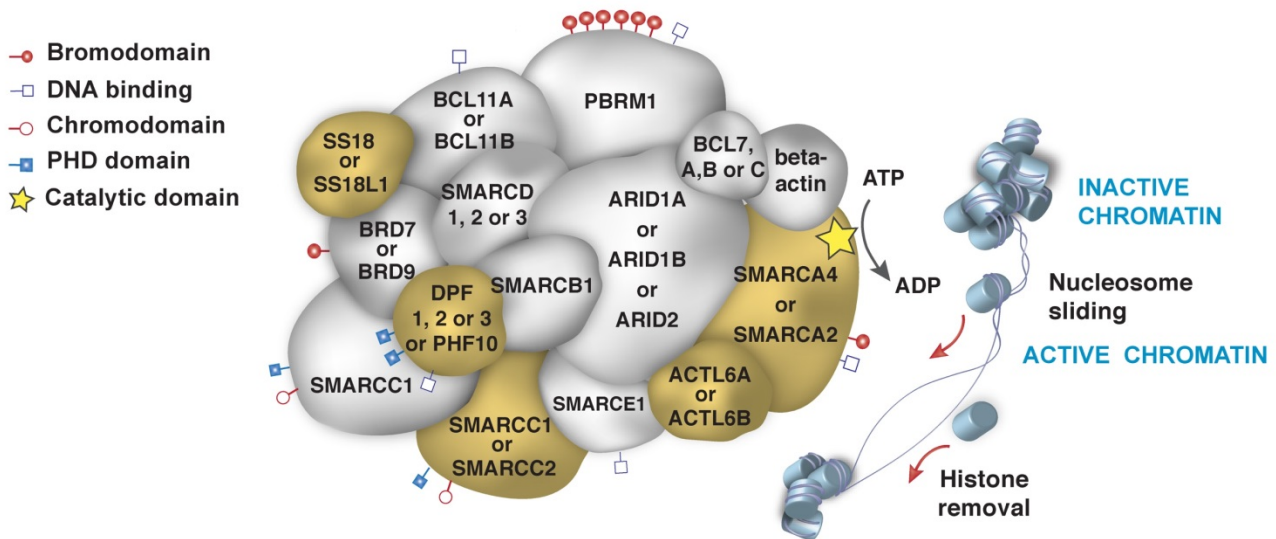


Figure 1. Schematic representation of the mSWI/SNF complex composition.

Multiple names indicated for the same subunit in this schematic representation reflect the existence of mutually exclusive paralogues. Subunit sizes indicate molecular weights in a semi-quantitative manner. Subunit localization within the complex is shown in an arbitrary manner. Some positions are constant (or “core”), while others can be occupied by one of two (or more) possible, mutually exclusive subunits. This offers a high diversity of combinations. Each subunit carries specific functional domains and a different molecular weight. The complex uses ATP-ase energy to remodel chromatin, although other possible mechanisms of action are increasingly recognized (Kadoch & Crabtree, 2015). Subunits highlighted in yellow are specifically discussed in the results sections of this study.

SWI/SNF subunits are described as "dedicated", meaning that they have not been found to exist "freely" outside of the complex (with rare exceptions) and do not exhibit functions outside of the complex (Kadoch & Crabtree, 2015). Given that that some SWI/SNF proteins go by many names, the same subunit is often designated by different terms across the literature. **Table 1** provides disambiguation of gene and protein names, and some basic information about mSWI/SNF subunits.

HUGO gene symbol	Aliases / protein names	Functional domain(s)	Role(s)	Predicted molecular weight (kDa)
SMARCA4	BRG1	ATP-ase, bromodomain	Catalytic (ATP-ase), helicase	184.5
SMARCA2	BRM	ATP-ase, bromodomain	Catalytic (ATP-ase), helicase	181
SMARCB1	BAF47, hSNF5, INI-1	SNF5		44
SMARCC1	BAF155	Chromodomain SANT, BRCT	DNA-binding, histone binding, protein-protein interaction?	123
SMARCC2	BAF170	Chromodomain, SANT, BRCT	DNA-binding, histone binding, protein-protein interaction?	133
SMARCD1	BAF60A	SWIB, MDM2	Interaction with nuclear receptors	58
SMARCD2	BAF60B	SWIB, MDM2		59
SMARCD3	BAF60C	SWIB, MDM2		55
SMARCE1	BAF57	HMG	DNA binding?	47
ARID1A	BAF250A, SMARCF1	ARID	DNA binding	242
ARID1B	BAF250B	ARID	DNA binding	236
ARID2	BAF200	ARID	DNA binding	197
PBRM1	BAF180, Polybromo	Bromodomain		193
ACTL6A	BAF53A	Actin	Chromatin association, enhancing ATP-ase activity	47.5
ACTL6B	BAF53B	Actin	Chromatin association, enhancing ATP-ase activity	47
PHF10	BAF45A	Zinc finger, RING-type		56
DPF1	BAF45B	Zinc finger, RING-type		42.5
DPF2	BAF45C	Zinc finger, RING-type		43
DPF3	BAF45D	Zinc finger, RING-type		44
BRD7	BRD7	Bromodomain		74
BRD9	BRD9	Bromodomain	"Reader" of acetylated lysines	67
SS18	SYT		Transcriptional coactivator	46

SS18L1	CREST			43
GLTSCR1	BICRA			158.5
GLTSCR1L	BICRAL			115
ACTB	Beta-actin, BAF42	Actin		41.5
BCL11A	BCL11A	Zinc finger, C2H2-type		91
BCL11B	BCL11B	Zinc finger, C2H2-type		95.5

Table 1. Disambiguation of mSWI/SNF subunit names and basic information regarding those subunits. (Kadoch & Crabtree, 2015; Masliah-Planchon et al., 2015).

2. Different forms of mSWI/SNF exist through combinatorial subunit assembly.

In contrast to the SWI/SNF complex found in yeast, mammalian SWI/SNF (mSWI/SNF) complexes display a diversity of compositions, achieved through combinatorial assembly of various subunits, encoded by 29 genes (Kadoch & Crabtree, 2015)(**Fig.2**). This likely mirrors higher complexity of these organisms, including a larger genome, an intricate body design and the presence of a central nervous system.

The two main forms of the mSWI/SNF complex are BAF (Brg1/Brm-Associated Factor complex, also termed the “canonical BAF complex”, or cBAF) and PBAF (Polybromo-associated BAF). Each of these two forms is characterized by the presence of specific subunits: in particular, BAF incorporates ARID1A/B (BAF250/250B) and DPF2, while PBAF is characterized by incorporation of BAF180 (polybromo/PBRM1), SMARCA4, BAF200 (ARID2), BAF45B (PHF10) and BRD7 (Kadoch & Crabtree, 2015; Mashtalir et al., 2018; Michel et al., 2018).

Most recently, a third “main” variant of the mSWI/SNF complex was described, and called the “non-canonical” BAF complex, or ncBAF (also termed “GBAF” by some authors) (Gatchalian et al., 2018; Mashtalir et al., 2018; Michel et al., 2018; X. Wang et al., 2019). It is defined by the presence of BRD9 and of one of the two recently identified SWI/SNF subunits, GLTSCR1 or GLTSCR1L (BICRA/BICRAL) (Gatchalian et al., 2018; Mashtalir et al., 2018;

Michel et al., 2018; X. Wang et al., 2019). Remarkably, this form of mSWI/SNF also seems to lack some other subunits: SMARCB1, the ARID core subunits, SMARCE1, and DPF/PHF subunits; this makes it a particularly interesting form of the complex to study in tumors that display deficiency of these subunits, as residual ncBAF complexes could represent a potential therapeutic target in such tumors (as discussed hereafter). While the differences between these three main forms of the complex have not been elucidated to date, cBAF has been reported to localize predominantly to enhancers, while PBAF and ncBAF appear to be more enriched at promoters (Mittal & Roberts, 2020).

In addition, other specialized forms of mSWI/SNF complexes have been identified at different stages of embryogenesis and of tissue maturation (Ho, Jothi, et al., 2009; Ho et al., 2011; Lessard et al., 2007; J. I. Wu et al., 2007; Yoo et al., 2009). A very notable example is provided by changes in SWI/SNF composition during neuronal differentiation. Cells committed to the neural lineage initially express a neural progenitor form of the complex (termed npBAF), which incorporates among others the BAF53A, BAF45a/d and SS18 subunits (Lessard et al., 2007; Staahl et al., 2013; J. I. Wu et al., 2007; Yoo et al., 2009). However, upon differentiation to post-mitotic neurons, the complex undergoes a switch to the neural variant (nBAF) and incorporates respective paralogs of the previously present subunits (i.e., BAF53B, BAF45B/C and SS18L1). This switch is mediated by the repression of BAF53A by micro-RNAs (miR-9* and miR-124) in response to downregulation of REST, a repressor of neurogenesis (Yoo et al., 2009).

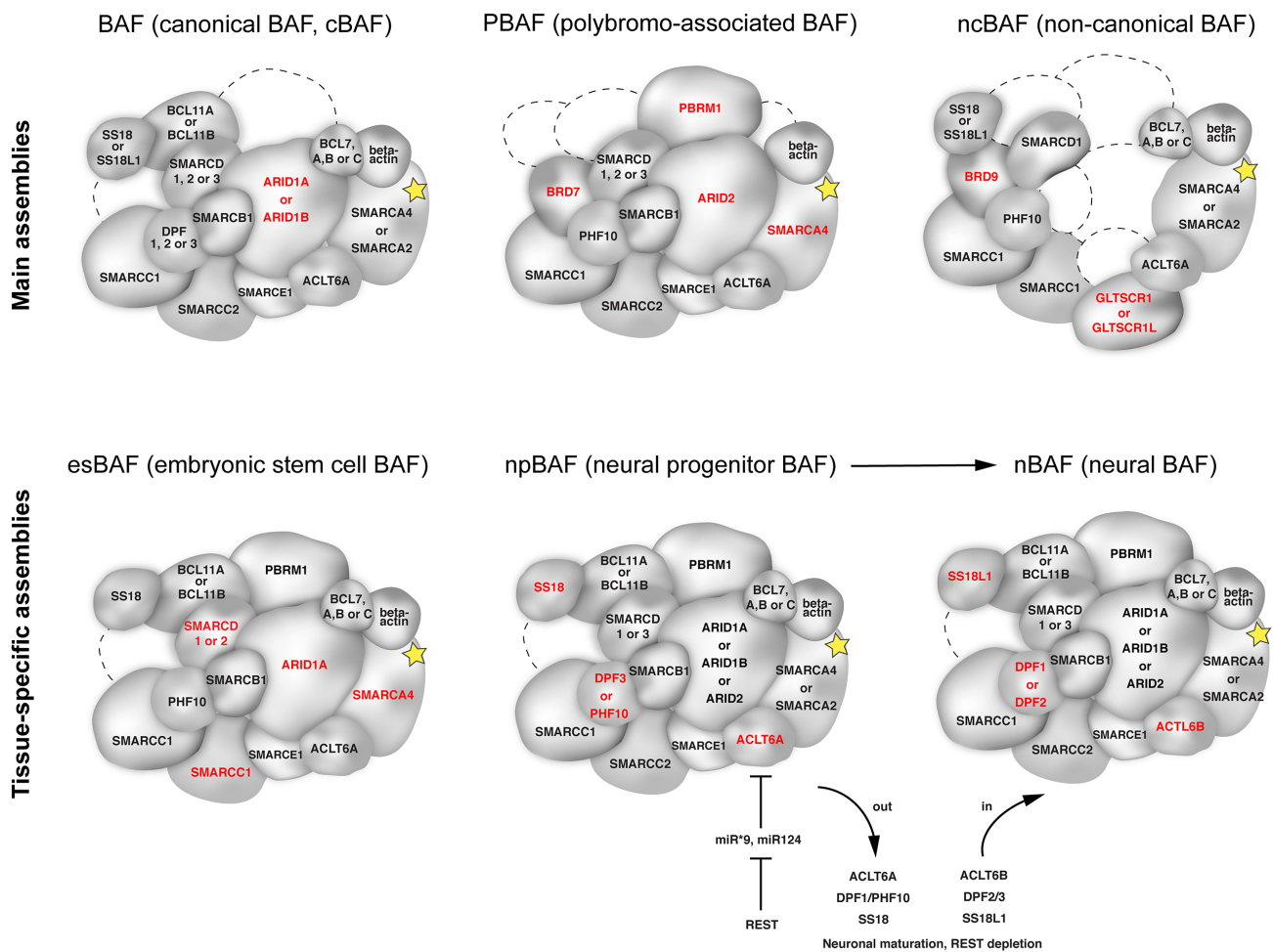


Figure 2. Schematic representations of combinatorial assemblies of the mSWI/SNF complex. Subunits which have been specifically linked to the specialized composition of each complex type are highlighted in red.

Another distinct version of the complex is the embryonic stem cell complex, termed esBAF, which is characterized by the presence of SMARCA4, SMARCC1 and SMARCD1, and the absence of their somatic cell homologs: SMARCA2, SMARCC2 and SMARCD3 (Ho, Jothi, et al., 2009; Ho et al., 2011; Ho, Ronan, et al., 2009; Yan et al., 2008). This specialized composition seems critical for embryonic stem cell functions: for example, forced expression of SMARCC2 (BAF170) impairs their ability of self-renewal (Ho, Ronan, et al., 2009).

Of note, other cell type-related specificities of SWI/SNF composition have also been described: in cardiac progenitors, the complexes incorporate exclusively SMARCD3 and not its paralogs (Lickert et al., 2004); in hematopoietic stem cells, they were shown to incorporate SMARCA2, SMARCD1/2, and Bcl7b/c, but not SMARCA4, SMARCD3, or Bcl7a (Kadoch & Crabtree, 2015; Prasad et al., 2014).

SWI/SNF complexes usually localize to active transcription sites, marked with the active chromatin mark H3K27ac. Among the many functional relationships of SWI/SNF with other proteins or complexes, one of the best studied is the one with the Polycomb Repressor Complex 2 (PRC2). PRC2 can oppose SWI/SNF activity and silence gene expression via its catalytic subunit EZH2 (Enhancer of zeste homologue 2), which places the inactive histone mark H3K27me3 on the chromatin (Mittal & Roberts, 2020). Disruption of the SWI/SNF-PRC2 balance (**Fig.3**) can lead to marked epigenetic perturbations with profound effects on the embryonic development and a role in oncogenesis, as mentioned hereafter.

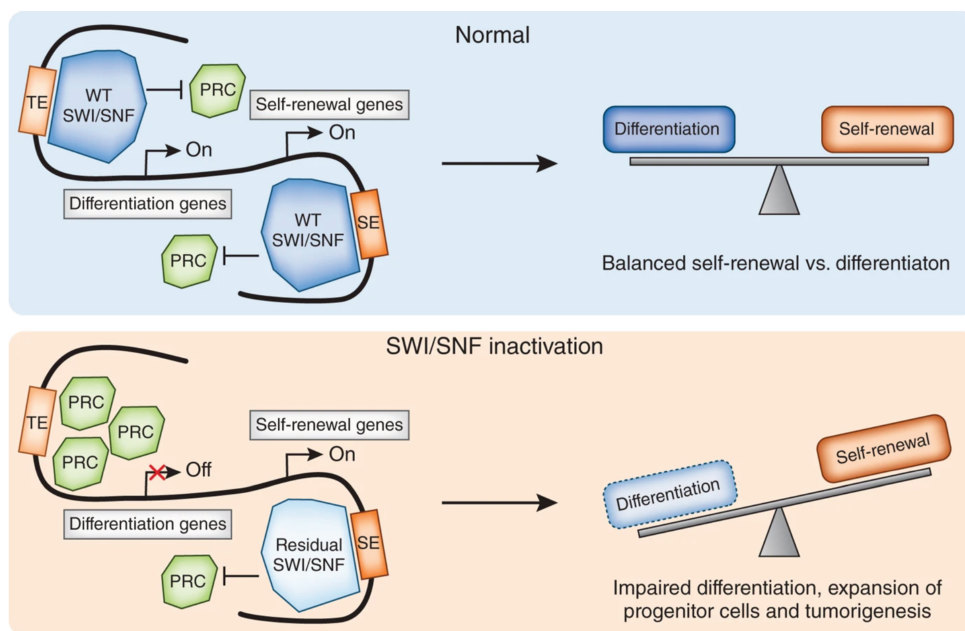


Figure 3. The SWI/SNF versus PRC2 antagonism. Figure reproduced from: (C. Lu & Allis, 2017). Due to loss of function alterations, SWI/SNF is unable to oppose PRC2 functions at gene promoters and typical enhancers (TE), which leads to aberrant silencing of these genes

(for example, differentiation genes) by PRC2. SE indicates super-enhancers.

3. Role of SWI/SNF in embryonic development, cell differentiation and cell reprogramming.

Several mouse knock-out models support a crucial role of the SWI/SNF complex in embryonic development. For example, mice with homozygous *Smarca1* inactivation experience early embryonic lethality (day 7) (Roberts et al., 2000). Mouse embryos with homozygous *Smarca4* inactivation have been shown to die during pre-implantation stage (Bultman et al., 2000), and *Smarca4* heterozygotes exhibit increased risk of exencephaly in late embryonic stages, although this phenomenon has not been described in humans. Conversely, mice with homozygous *Smarca2* inactivation were shown to develop normally, but exhibited signs suggestive of increased proliferation in various tissues, such as an increased body weight in adulthood and a higher proliferation rate of embryonic fibroblasts (Reyes et al., 1998). However, a subsequent study suggested that this mouse model may represent an incomplete knockout, whereby a truncated, but still functional form of the protein is still expressed in tissues (Thompson et al., 2015). Thus, the *in vivo* effects of *Smarca2* knock-out remain to be validated in additional studies. Other subunits have also been implicated in embryonic development in mouse knock-out models: for example, *Arid1a* depletion results in the absence of mesoderm and early embryonic arrest (Gao et al., 2008), and *Smarcd3* depletion induces defects in heart development (Lickert et al., 2004).

Several developmental syndromes in humans have also been linked to SWI/SNF mutations. The Coffin-Siris syndrome is characterized by mild to severe intellectual disability, difficulties in language acquisition, coarse facial features and abnormalities of the fifth fingers and toes; it has been linked to the presence of pathogenic *ARID1B* mutations (Santen et al., 2012), but also to mutations in other SWI/SNF subunits, including *SMARCB1* and *SMARCA4* (Sekiguchi et al., 2019; Tsurusaki et al., 2012). The Nicolaides-Baraitser includes

developmental delays of variable severity, absent or limited speech, seizures, short stature, facial dimorphism, and finger dysmorphism with prominent distal phalanges and interphalangeal joints; it has been linked to the presence of *SMARCA2* mutations or in rare cases, to *ARID1B* mutations (Van Houdt et al., 2012). Notably, mutations observed in these syndromes are heterozygous, suggesting a dose-dependent or a dominant-negative mechanism (Kadoch & Crabtree, 2015). In addition, mutations found in *SMARCA4* and *SMARCB1* are missense mutations or in-frame indels, rather than truncating mutations, as opposed to those found in the Rhabdoid Tumor Predisposition Syndrome (discussed below); this could explain why these developmental syndromes are not associated with a predisposition to SWI/SNF-deficient tumors, which usually involve deleterious alterations resulting in a complete loss of expression of a given subunit (Masliah-Planchon et al., 2015).

The esBAF complexes are crucial for the maintenance of pluripotency and proliferation of embryonic stem cells, and they co-occupy targets with master regulators (SOX2, Oct4 and Nanog (Ho, Jothi, et al., 2009; Ho et al., 2011; Ho, Ronan, et al., 2009; Yan et al., 2008). Knock-down of critical components of esBAF, such as *SMARCA4* or *ARID1A*, impairs self-renewal and pluripotency of these cells, and leads to a decrease in expression of master stem cell factors (Ho, Ronan, et al., 2009; Kidder et al., 2009; Yan et al., 2008). Interestingly, overexpression of *Smarcc1* and *Smarca4*, in addition to an ectopic expression of the “four factors” (*Oct4*, *Sox2*, *Klf4* and *c-Myc*), was shown to increase the efficacy of reprogramming of mouse somatic cells into induced pluripotent stem cells (iPS) (Singhal et al., 2010); this result was also observed in the absence of *c-Myc*. Overexpression of *Smarcc1* alone had a more potent effect than overexpression of *Smarca4* alone, and combined overexpression of *Smarcc1* and *Smarca4* was shown to be synergistic rather than additive. Likewise, in another study, knock-down of the respective paralogs of these subunits, i.e. *Smarca2* and *Smarcc2*, also improved reprogramming efficacy (Jiang et al., 2015). Such changes in SWI/SNF subunit levels may favor the assembly

of esBAF-like complexes in lieu of somatic-type BAF complexes, which in turn enhances cell reprogramming.

From a functional standpoint, in mouse embryonic stem cells (mESC), SWI/SNF complexes oppose the action of the Polycomb Repressive Complex (PRC2)(Ho, Jothi, et al., 2009; Ho et al., 2011). Depletion of SMARCA4 in mESC leads to an increased deposition of the inhibitory histone mark H3K27me3 by PRC2 at SMARCA4-activated sites, which allows to silence these genes. Homeobox genes (Hox genes), responsible for body plan specification, are the exception, as both SWI/SNF and PRC2 complexes seem to act synergistically on those genes, whereby SWI/SNF could facilitate the placement of the H3K27me3 mark by PRC2 (Ho et al., 2011; Kadoch & Crabtree, 2015). Conversely, in *Drosophila*, SWI/SNF opposes PRC at all loci, including Hox genes (Tamkun et al., 1992).

The recently identified non-canonical BAF was also found to be present in mouse embryonic stem cells, and ncBAF and esBAF seem to exert distinct and complementary roles in maintaining cell pluripotency (Gatchalian et al., 2018). This could be due in part to interaction of esBAF and ncBAF with different pluripotency factors, which target the complexes to distinct areas of the chromatin (for example, ncBAF interacts with BRD4 in a bromodomain-dependent manner) (Gatchalian et al., 2018).

As seen above, some forms of SWI/SNF, such as nBAF, are expressed in a tissue-specific manner to serve specific functions. *Baf53b*^{-/-} mice show high mortality after birth (attributed to a failure to nurse), and mice that survive display behavioral problems and severe defects of long-term memory and of synaptic plasticity; these can be explained, at least in part, by impaired dendritic growth (Vogel-Ciernia et al., 2013; Wu et al., 2007). Remarkably, in human fibroblasts, forced expression of the micro-RNAs miR-9* and miR-124, which have been shown to repress the expression of BAF53A and thus allow incorporation of BAF53B as part of

the nBAF complex in neuronal maturation, results in conversion of fibroblasts to neurons (Yoo et al., 2011).

In summary, distinct SWI/SNF subunits and specialized SWI/SNF assemblies appear to play critical roles in embryonic development and in cell fate specification.

4. The SWI/SNF complex and cancer

4.1 Alterations in SWI/SNF genes are frequently encountered in cancer

Next generation sequencing studies from The Cancer Genome Atlas and International Cancer Genome Consortium have shown that SWI/SNF genes are among the most commonly mutated in cancer. Over 20% of human cancers have been estimated to carry a genomic alteration in at least one of the SWI/SNF subunits (Kadoch & Crabtree, 2015; Kadoch et al., 2013; Shain & Pollack, 2013). **Table 2** summarizes established examples of such malignancies, as well as the approximate frequencies at which genomic alterations in SWI/SNF subunits are observed. Based on these results, the SWI/SNF complex is often regarded as a tumor suppressor.

Cancer type	Organ system	Gene	Protein	Type of alteration	% of cases	Additional comments	References
Malignant rhabdoid tumor (MRT), AT/RT if intra-cranial	Kidney, soft tissue, viscera, brain	<i>SMARCB1</i>	BAF47/INI1	Biallelic loss-of-function alterations	95%	Possible germline alterations (~25%)	(Versteeg et al., 1998)
	Brain > other sites	<i>SMARCA4</i>	BRG1	Biallelic loss-of-function alterations	<5%	Possible germline alterations (at least 40%)	(Hasselblatt et al., 2011)
SCCOHT	Ovary	<i>SMARCA4</i>	BRG1	Biallelic loss-of-function alterations, concomitant loss of <i>SMARCA2</i> expression	>90%	Possible germline alterations (~40%)	(Jelinic et al., 2014; Kupryjanczyk et al., 2013; Ramos et al., 2014; Witkowski et al., 2014)
SMARCA4-deficient thoracic sarcoma	Thorax	<i>SMARCA4</i>	BRG1	Biallelic loss-of-function alterations, concomitant loss of <i>SMARCA2</i> expression	~100%	Possible overlap with <i>SMARCA4</i> -deficient lung adenocarcinomas	(Le Loarer et al., 2015)
Lung adenocarcinoma	Lung	<i>SMARCA4</i>	BRG1	Truncating, sometimes missense, usually homozygous	5-10%	Usually TTF1 negative, often poorly differentiated	(Cancer Genome Atlas Research, 2014c)
Clear cell carcinoma of the ovary or endometrium	Ovary or endometrium	<i>ARID1A</i>	BAF250A	Truncating, sometimes missense, usually homozygous	~50%		(S. Jones et al., 2010; Wiegand et al., 2010)
Endometrioid carcinoma	Ovary or endometrium	<i>ARID1A</i>	BAF250A	Truncating, sometimes missense, usually homozygous	~30%		(Wiegand et al., 2010)
Dedifferentiated/undifferentiated carcinoma of the ovary or endometrium**	Ovary or endometrium	<i>ARID1A</i> and <i>ARID1B</i> *	BAF250A and BAF250B	Loss of protein expression, genomic correlates remain to be determined	~25%		(Coatham et al., 2016) (Ramalingam, Croce, & McCluggage, 2017)
Dedifferentiated/undifferentiated carcinoma of the endometrium**	Endometrium	<i>SMARCA4</i> and <i>SMARCA2</i> *	BRG1 and BRM	Loss of protein expression, genomic correlates remain to be determined	30-40%	Usually PAX8 negative, possible germline alterations (case reports)	(Karnezis, Hoang, et al., 2016)
Urothelial carcinoma	Bladder, upper urothelial tract	<i>ARID1A</i>	BAF250A	Truncating, sometimes missense, usually homozygous	20-25%		(Cancer Genome Atlas Research, 2014b)
Gastric adenocarcinoma	Stomach	<i>ARID1A</i>	BAF250A	Truncating	10% to >70%	Frequency depends on molecular subtype (highest in MSI or EBV+ cancers)	(Cancer Genome Atlas Research, 2014a)
Clear cell renal cell carcinoma	Kidney	<i>PBRM1</i>	BAF180	Usually truncating, often homozygous	35-45%		(Varela et al., 2011)
Medullary renal carcinoma	Kidney	<i>SMARCB1</i>	BAF47/INI1	Deletion, often as a result of gene rearrangement events	~100%	Associated with sickle cell trait	(Calderaro et al., 2016)
SMARCB1-deficient sinonasal carcinoma	Sinonasal	<i>SMARCB1</i>	BAF47/INI1	Deletions	~100%		(Agaimy et al., 2014)
Synovial sarcoma	Soft tissue	<i>SS18</i>	SS18/SYT	<i>SS18-SSX</i> gene fusion***	~100%	Associated with attenuated expression of BAF47/INI1 by IHC	(Clark et al., 1994)

Poorly differentiated pediatric chordoma	Soft tissue	<i>SMARCB1</i>	BAF47/INI1	Deletions	~100%	Associated with poor prognosis	(Mobley et al., 2010)
Myoepithelial carcinoma of the soft tissue	Soft tissue	<i>SMARCB1</i>	BAF47/INI1	Deletions	20-40%		(Le Loarer et al., 2014)
Epithelioid sarcoma	Soft tissue	<i>SMARCB1</i>	BAF47/INI1	Deletion (90%) or mutation (10%)	~95%		(Le Loarer et al., 2014)
Epithelioid MPNST	Nerve sheath	<i>SMARCB1</i>	BAF47/INI1	Loss of protein expression, genomic correlates remain to be determined	~70%		(Jo & Fletcher, 2015)
Epithelioid schwannoma	Nerve sheath	<i>SMARCB1</i>	BAF47/INI1	Usually missense or splice site mutations	40%	Possible germline alterations (schwannomatosis)	(Jo & Fletcher, 2017)
Extraskeletal myxoid chondrosarcoma	Soft tissue	<i>SMARCB1</i>	BAF47/INI1	Deletions	17%		(Kohashi et al., 2008)
Medulloblastoma	CNS	<i>SMARCA4</i>	BRG1	Heterozygous missense mutations or in-frame indels in the helicase domain**	4% overall	Frequency of mutations by molecular subgroup: WNT > Group 3 > Shh	(D. T. Jones et al., 2012)
Meningioma	CNS	<i>SMARCB1</i>	BAF47/INI1	Usually missense or splice site mutations		Possible germline alterations (with schwannomatosis)	(Bacci et al., 2010)
	CNS	<i>SMARCE1</i>	BAF57	Germline mutations, localized to the HMG domain; LOH in tumors		~100% in non <i>NF2</i> -driven multiple spinal meningiomas; often clear cell histology; poor prognosis	(Smith et al., 2013)
Burkitt lymphoma	Lymphoid	<i>SMARCA4</i>	BRG1		5-10%		(Love et al., 2012)
Undifferentiated carcinoma (various sites)	Pancreas, colorectum, urinary tract...	<i>SMARCA4</i> , <i>SMARCA2</i> , <i>SMARCB1</i> <i>SMARCC1</i>	BRG1, BRM, BAF47, BAF155	Loss of expression of at least one of these proteins by IHC; genomic correlates remain to be determined	25-90%		(Agaimy, Bertz, et al., 2016; Agaimy, Daum, et al., 2016; Tessier-Cloutier et al., 2020)

Table 2: Cancer types with alterations in SWI/SNF subunits (non-exhaustive). MPNST: malignant peripheral nerve sheath tumor; CNS: central nervous system; LOH: loss of heterozygosity; *indicates concomitant alterations in both genes listed; ** histologic/morphologic overlap; ***not a loss-of-function event (see paragraph 4.2).

The best-known example of a malignancy characterized by genomic SWI/SNF loss-of-function alterations are malignant rhabdoid tumors (MRT), an extremely aggressive form of childhood cancer (including its intra-cranial form, Atypical Teratoid-Rhabdoid Tumor, or AT/RT). These tumors are characterized by the presence of highly recurrent biallelic loss-of-function alterations in *SMARCB1*, found in nearly all cases (>95%) (Versteeg et al., 1998). This results in a loss of nuclear expression of SMARCB1 (also termed BAF47, hSNF5 or INI-1), which can be evidenced by immunohistochemistry and represents an important element in pathology for MRT diagnosis. In addition, germline *SMARCB1* mutations are present in about 25% of MRT patients (Bourdeaut et al., 2011; Eaton et al., 2011). The presence of germline *SMARCB1* mutations is responsible for the Rhabdoid Tumor Predisposition Syndrome 1 (RTPS1), which is characterized by extremely high penetrance (~100%) (Bourdeaut et al., 2011) and represents a bona fide scenario of a tumor suppressor gene, with loss of the wild-type allele ("second hit") following the Knudson model.

Functional confirmation for the role of SMARCB1 alterations in MRT oncogenesis was obtained through a genetically engineered mouse model with heterozygous *Smarcb1* inactivation, which demonstrated predisposition to soft tissue tumors of the neck and the cervical area with a histology highly reminiscent of human MRTs (Roberts et al., 2000). However, the penetrance of this model is relatively low (~12%) and tumors tend to occur in older animals. Conversely, in an inducible model of homozygous *Smarcb1* inactivation, the tumor onset is remarkably rapid even with respect to other tumor suppressor knock-out models (median delay, 11 week), and the penetrance is complete (100%) (Roberts et al., 2002). Most malignancies arising in this model are mature CD8+ T cell lymphomas, and a subset are MRTs. In a subsequent study, it was shown that the phenotype was actually highly dependent on the temporality of *SMARCB1* inactivation: in an inducible mouse model, *Smarcb1* inactivation between E6 and E10 led to the formation of mainly intra-cranial tumors, with high penetrance and short delay (median, 3 months) (Han et al., 2016).

Very rare MRT do not harbor *SMARCB1* mutations but instead, display loss-of-function *SMARCA4* alterations (Hasselblatt et al., 2011; Hasselblatt et al., 2014; Schneppenheim et al., 2010). Nearly all cases reported to date were intra-cranial tumors (AT/RT), contrarily to *SMARCB1*-mutated MRT. However, extra-cerebral *SMARCA4*-mutated MRT can also be encountered (Chun et al., 2016; Fruhwald et al., 2006; Saunders et al., 2020); and Dr. Franck Bourdeaut, *unpublished*). Germline *SMARCA4* mutations, seen in at least 40% of patients with *SMARCA4*-mutated MRT/ATRT, are responsible for the Rhabdoid Tumor Predisposition Syndrome 2 (RTPS2). Interestingly, *SMARCA4*-mutated MRT may occasionally display missense mutations (Masliah-Planchon et al., 2016).

In recent years, sequencing of ovarian small cell carcinomas of the hypercalcemic type (SCCOHT) has enabled the first molecular characterization of this rare tumor, by demonstrating that virtually all cases harbor biallelic, deleterious *SMARCA4* alterations with loss of expression of the *SMARCA4* (BRG1) protein (Jelinic et al., 2014; Kupryjanczyk et al., 2013; Ramos et al., 2014; Witkowski et al., 2014). Remarkably, this is accompanied by a concomitant complete loss of expression of the *SMARCA2* paralog (Karnezis, Wang, et al., 2016). Thus, SCCOHT were included among tumors in which deleterious SWI/SNF alteration represents the initiating driver event, and it was suggested by some authors that they should be considered as the ovarian counterpart of MRT (Witkowski et al., 2014). In addition, germline *SMARCA4* alterations are found in up to 43% of SCCOHT (Witkowski et al., 2016), and one family case study reported a mother diagnosed with SCCOHT and a daughter diagnosed with an AT/RT, with both tumors carrying the identical loss-of-function *SMARCA4* mutation (Witkowski et al., 2013). Overall, SCCOHT are now considered part of the RTPS2 spectrum, alongside *SMARCA4*-mutated MRTs.

In mice, heterozygous *Smarca4* inactivation was shown to increase susceptibility to apocrine subcutaneous tumors (Bultman et al., 2000). However, these tumors occurred with low penetrance (15%), did not exhibit loss of heterozygosity for *Smarca4*, and did not resemble a

known *SMARCA4*-deficient tumor type seen in humans, suggesting that this mouse model does not adequately recapitulate RTPS2. Regarding the *SMARCA2* paralog, heterozygous or homozygous inactivation of this gene in mice was not sufficient to cause tumor susceptibility on its own (Reyes et al., 1998), but it did increase the frequency of lung tumors (interpreted as “adenomas”) after administration of a lung-specific carcinogen (Glaros et al., 2007); however, as mentioned above, this mouse model may not represent a complete knockout (Thompson et al., 2015).

While MRT and SCCOHT are remarkable examples of SWI/SNF-deficient tumors, these two malignancies are exceedingly rare. As shown in **Table 2**, the notoriety of SWI/SNF as a “tumor suppressor” has been established through examples of numerous other tumor types, both rare and relatively frequent ones, with documented recurrent SWI/SNF alterations. However, this overview of SWI/SNF alterations across tumor types suggests that their effects are highly context- and tissue-dependent. First, these alterations are observed at variable frequencies depending on the tumor type, suggesting that while they may represent a bona fide initial oncogenic event in some tumors, in other tumors they may instead be related to tumor progression, act in conjunction with other genomic alterations, or represent passenger alterations. Second, the tumorigenic potential is highly dependent on the alteration type: for example, while germline loss-of-function *SMARCB1* mutations lead to an MRT predisposition syndrome with 100% penetrance, germline alterations associated with schwannomatosis are usually hypomorphic (e.g. missense, splice site). In addition, heterozygous SWI/SNF alterations can also be seen in cancers, similarly to what is observed in intellectual disability syndromes, suggesting the possibility of dosage-sensitive or dominant-negative mechanisms of the tumor suppressive activity (Kadoch & Crabtree, 2015).

Some phenotypic patterns have been noted across malignancies with SWI/SNF alterations, although this genotype-phenotype correlation is not absolute. First, many tumors from the spectrum of SWI/SNF-deficient malignancies show a “rhabdoid cytomorphology”,

exemplified by round or polygonal, poorly cohesive cells with characteristic perinuclear cytoplasmic inclusions, which represent bundles of intermediate filament (Itakura et al., 2001); they also display large, often eccentric nuclei and prominent nucleoli (**Fig.4**). This morphology is characteristic of the eponymous MRTs and of SCCOHT, but it can also be seen in other SWI/SNF-deficient tumors including, but is not limited to, SMARCA4-deficient thoracic sarcomas, undifferentiated/dedifferentiated carcinomas of various anatomic sites (endometrium, ovary, colorectum, pancreas, urinary tract), SMARCB1-deficient sinonasal carcinomas, and various types of SMARCB1-deficient soft tissue neoplasms (Schaefer & Hornick, 2020). While the presence of a rhabdoid morphology is strongly suggestive of a tumor with a SWI/SNF alteration, the opposite is not true: for example, most lung adenocarcinomas with *SMARCA4* mutations, clear cell carcinomas of the ovary or of the endometrium with *ARID1A* mutations, clear cell renal cell carcinomas with *PBRM1* mutations, and other tumors, do not display a rhabdoid morphology (**Fig.5**).

Some authors have also pointed out that several other tumor types with SWI/SNF deficiency share a clear-cell histomorphology (namely, clear cell renal cell carcinoma, clear cell carcinoma of the ovary/endometrium, and clear cell meningioma)(Masliah-Planchon et al., 2015; Smith et al., 2013). While clear cell morphology can be due to cytoplasmic accumulation of glycogen or other substances, a causal relationship between this morphology type and SWI/SNF alterations has not been established to date. In addition, the presence of SWI/SNF mutations does not seem mandatory for clear cell morphology to occur in these tumors types (for example, >50% of clear cell renal cell carcinomas do not harbor SWI/SNF alterations).

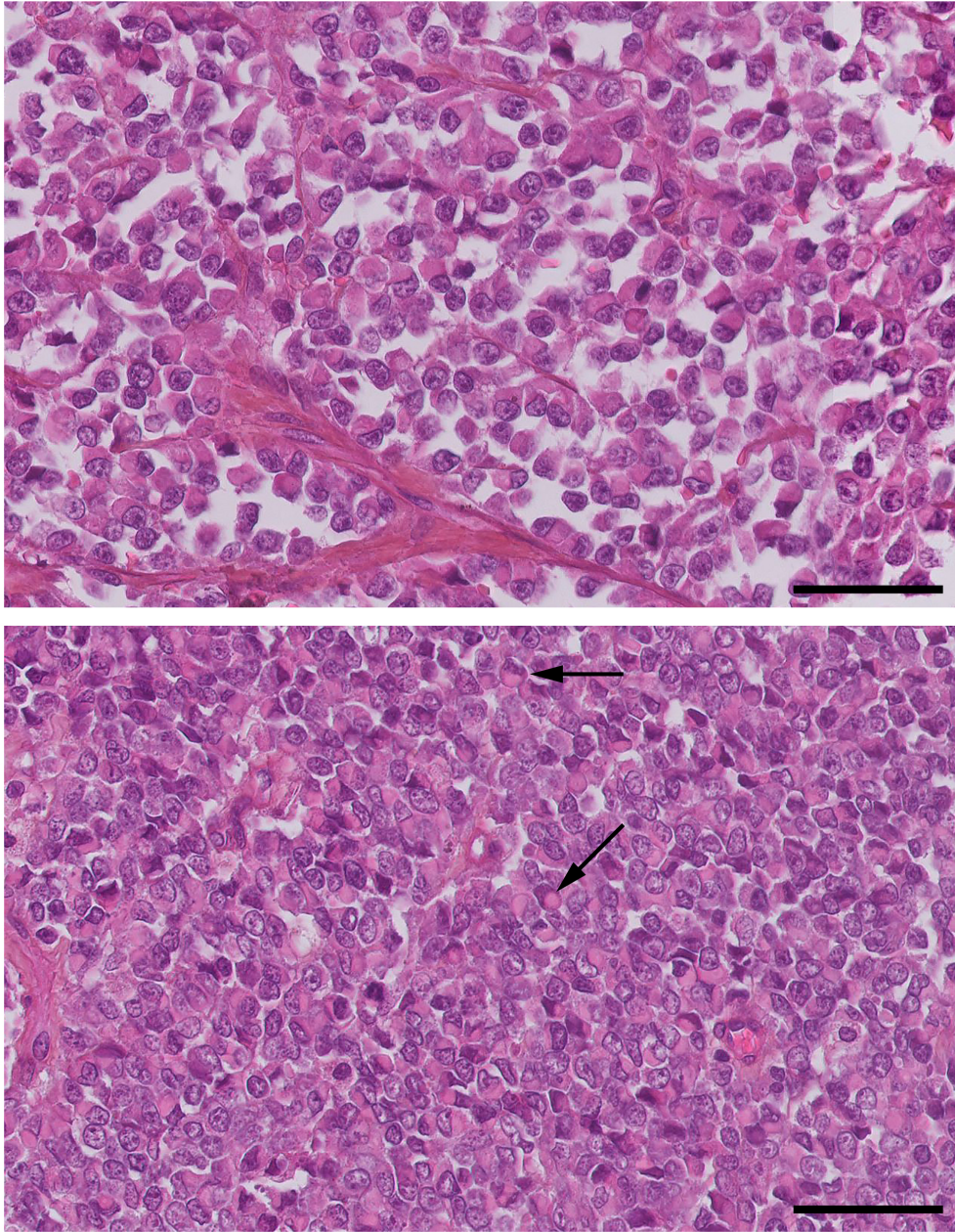


Figure 4. Examples of a rhabdoid cytomorphology in a MRT case. The tumor is composed of poorly cohesive medium to large cells showing large nuclei, prominent nucleoli and characteristic eosinophilic intra-cytoplasmic inclusions (arrows). Hematoxylin-eosin-saffron stain, 20x magnification, scale bars: 50 μ m.

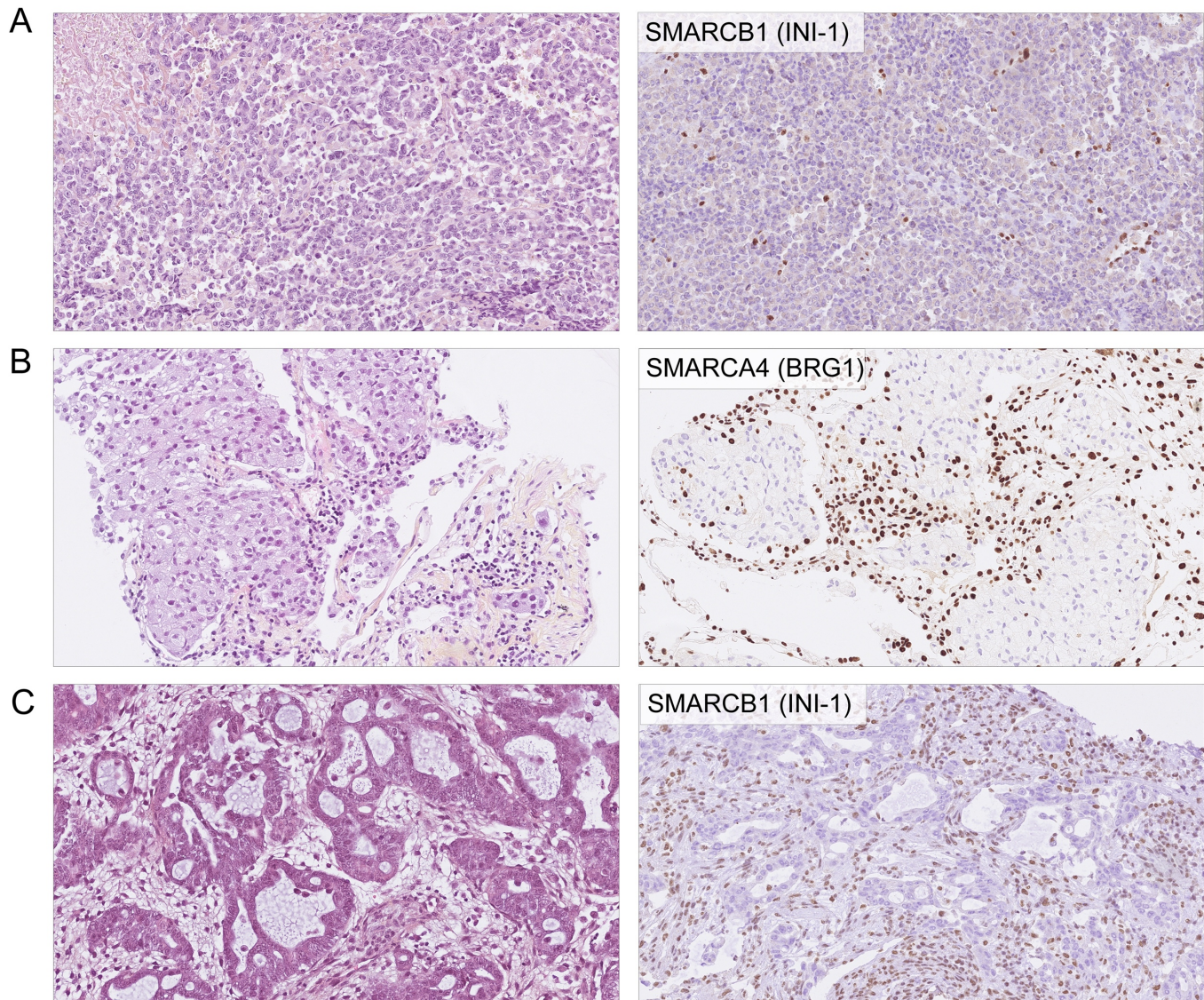


Figure 5. Examples of SWI/SNF-deficient malignancies (histomorphology and relevant immunostainings). A. Malignant rhabdoid tumor, SMARCB1-deficient; **B.** Lung adenocarcinoma, SMARCA4-deficient; **C.** Medullary carcinoma of the kidney, SMARCB1-deficient.

In addition to insights from tumor genomics, several functional properties of SWI/SNF support its tumor-suppressive functions. For example, SWI/SNF is known to bind to the tumor suppressor protein Rb, and cooperates with Rb to repress E2F1 and to control the cell cycle (Dunaief et al., 1994; Trouche et al., 1997). Functional studies in *Drosophila* and in mice have also shown that one of SWI/SNF functions consists in antagonizing the Polycomb Repressive Complex (PRC2) at distinct areas of the genome (Ho et al., 2011; Tamkun et al., 1992).

Accordingly, following SMARCB1 depletion in MRT, the SWI/SNF complex loses the ability to evict Polycomb from such sites, which leads to aberrant repression of tumor-suppressor genes, such as *CDKN2A* (the cyclin P16/INK4A), by PRC2, through deposition of the H3K27me3 histone mark (Wilson et al., 2010).

4.2 Evidence for dependency of some cancers on SWI/SNF

In parallel to the above examples where SWI/SNF seems to behave like as tumor suppressor, there is also some evidence for potential tumor-promoting functions of SWI/SNF. First, oncogenic genomic alterations in SWI/SNF have been identified in some tumor types. The best-known example is the *SS18-SSX* gene fusion, found in nearly 100% of synovial sarcomas. The *SS18-SSX* fusion product is incorporated in the complex in the place of the dedicated *SS18* subunit. The *SS18-SSX* fusion oncoprotein retargets the complex towards areas of the genome where it would normally not be present, which allows the aberrant complex to oppose polycomb-mediated gene repression at those sites. This in turn leads to an abnormal expression of oncogenic genes (e.g. *SOX2*, *PAX6*) and gene programs, inducing cell proliferation (Clark et al., 1994; Kadoch & Crabtree, 2013; McBride et al., 2018). In addition, perturbed composition of the complex with the *SS18-SSX* fusion oncoprotein leads to exclusion of the *SMARCB1* subunit, which subsequently gets degraded, resulting in low *SMARCB1* protein levels in tumor cells, which can be evidenced by IHC (**Fig.6**). Of note, a fusion-specific antibody against the *SS18-SSX* fusion oncoprotein has recently been developed, and immunostaining with this antibody appears to be exquisitely specific for synovial sarcoma, making it a valuable diagnostic tool (Baranov et al., 2020).

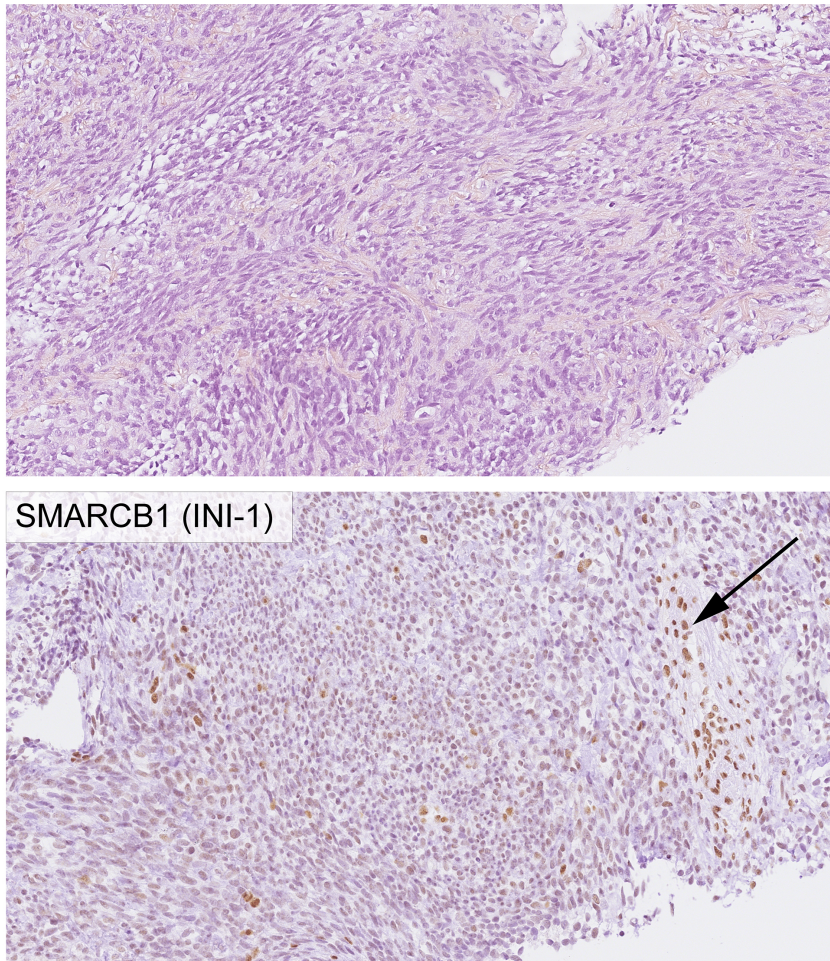


Figure 6. Morphology (top) and SMARCB1 (BAF47/INI-1) immunostaining (bottom) in monophasic synovial sarcoma. The nuclear staining in tumor cells is weaker than in neighboring benign structures, such as vessels (arrow).

Second, some tumors types show dependency on specific SWI/SNF subunits and/or overexpression of these subunits. High expression of SMARCA4 (BRG1) was shown to correlate with worse prognosis in various cancer types (Bai et al., 2013; Guerrero-Martinez & Reyes, 2018; Muthuswami et al., 2019; Q. Wu, Sharma, et al., 2016). In functional studies, SMARCA4 was crucial to maintain cell proliferation of various malignancies, including leukemia, breast cancer, neuroblastoma, melanoma and glioma (Bai et al., 2012; Buscarlet et al., 2014; Jubierre et al., 2016; Laurette et al., 2020; Q. Wu, Madany, et al., 2016). Other SWI/SNF subunits were also implicated in cancer cell proliferation or survival. For example, the BAF45B

(DPF1) subunit was shown to contribute to maintenance of “cancer stem cell” features in glioblastoma cells (Hiramatsu et al., 2017). Such vulnerabilities are extremely interesting to explore, as they could offer potential therapeutic targets, even though they may turn out to be due to synthetic lethality mechanisms (similar to the example of BRCA1/2 and PARP1) rather than to intrinsic oncogenic functions of SWI/SNF.

Studies of synthetic lethality approaches indeed revealed that in tumors with a deficiency of a SWI/SNF subunit, tumor cells are dependent on the residual aberrant complexes for their survival and growth. Functional studies showed that SWI/SNF complexes can still assemble in the absence of SMARCB1, SMARCA4 or ARID1A (Doan et al., 2004; Helming et al., 2014; Hoffman et al., 2014). In line with this hypothesis, SMARCB1-deficient tumors in a mouse MRT model depend on SMARCA4 for their survival (X. Wang et al., 2009). Large-scale dependency screens in cell lines, such as Project Achilles, revealed that in *SMARCA4* and *ARID1A* mutant cell lines, the top identified vulnerability was a paralogue SWI/SNF subunit in each case (Helming et al., 2014). Accordingly, targeting the catalytic subunit BRM (*SMARCA2*) induces cell death in some *SMARCA4*-mutant cancers, such as *SMARCA4*-deficient lung adenocarcinomas (Hoffman et al., 2014). Similarly, in *ARID1A*-deficient cancer cell lines, depletion of *ARID1B* leads to destabilization of the residual complex and impairs cell growth (Helming et al., 2014). Most recently, synovial sarcoma and MRT cell lines were shown to be exquisitely sensitive to the loss of *BRD9* or *GLTSCR1* subunits, suggesting their dependency on the non-canonical BAF complex in the absence of a functional canonical BAF (Michel et al., 2018; X. Wang et al., 2019).

Dependency of cancer cells on selected SWI/SNF subunits may ultimately lead to the development of SWI/SNF-targeting therapeutics. Accordingly, treatment with a biological inhibitor of *SMARCA4* (ADAADi) was shown to decrease the growth of triple-negative breast cancer cell line MDA-MB-231 (Wu et al., 2016), and more recently of prostate cancer cells

(Muthuswami et al., 2019). Similarly, small molecule inhibitors of BRD9 inhibit the growth of mouse and human acute myeloid leukemia (AML) cells (Hohmann et al., 2016).

However, these synthetic lethality studies point to an intriguing and somewhat paradoxical observation. If some SMARCA4-deficient tumors are very sensitive to the loss of SMARCA2, why do other tumor types (namely, MRT and SCCOHT) tolerate, or are indeed driven by, the concomitant loss of both these subunits (Hasselblatt et al., 2011; Witkowski et al., 2014)? Likewise, if targeting ARID1B is an efficient approach in ARID1A-deficient tumors, why do some tumor types (e.g. dedifferentiated carcinoma of the endometrium or ovary) display combined loss of ARID1A and ARID1B (Coatham et al., 2016)? Collectively, these observations suggest a tumor and tissue type-dependent relationship between SWI/SNF and cancer biology.

5. Aims of this research

The overarching goal of this research was to expand the knowledge on tumor-suppressor and tumor-promoting roles of SWI/SNF in cancer.

In the first part of this dissertation, I report molecular findings in an independent cohort of small cell carcinomas of the ovary, hypercalcemic type (SCCOHT), as an example of a tumor driven by SWI/SNF loss-of-function alterations. This very rare, but highly lethal malignancy affecting young women, is driven by deleterious mutations in the *SMARCA4* gene. Our goal was to perform a comprehensive molecular characterization of SCCOHT cases from our institution, in order to expand the knowledge about the biology of these rare tumors.

In the second part of this research, I aimed to better characterize the role of the SWI/SNF complex in prostate cancer. Contrarily to some above-cited malignancies in which SWI/SNF has been extensively studied, the potential place of SWI/SNF in prostate cancer progression has not been sufficiently elucidated. One of our specific aims was also to include

neuroendocrine prostatic carcinoma in our analyses, given that epigenetic mechanisms are known to play an important role in the development of this highly aggressive form of prostate cancer.

6. Related publications

Please see Appendix I and II.

- Most results from Part II were published in:

Small Cell Carcinoma of the Ovary, Hypercalcemic Type (SCCOHT) beyond *SMARCA4* Mutations: A Comprehensive Genomic Analysis.

Aurélie Auguste, Félix Blanc-Durand, Marc Deloger, Audrey Le Formal, Rohan Bareja, David C. Wilkes, Catherine Richon, Béatrice Brunn, Olivier Caron, Mojgan Devouassoux-Shisheboran, Sébastien Gouy, Philippe Morice, Enrica Bentivegna, Andrea Sboner, Olivier Elemento, Mark A. Rubin, Patricia Pautier, Catherine Genestie, Joanna Cyrta and Alexandra Leary. *Cells* **2020. 9(6), 1496.**

- Results from paragraph 2.8 in Part II are part of a manuscript in preparation (work with Dr Franck Bourdeaut's team, Institut Curie).
- Most results from Part III were published in:

Role of Specialized Composition of SWI/SNF Complexes in Prostate Cancer Lineage Plasticity.

Joanna Cyrta*, Anke Augspach*, Maria Rosaria de Filippo, Davide Prandi, Phillip Thienger, Matteo Benelli, Victoria Cooley, Rohan Bareja, David Wilkes, Sung-Suk Chae, Paola Cavaliere, Noah Dephoure, Anne-Christine Uldry, Sophie Braga Lagache, Sandra Cohen, Muriel Jaquet, Laura P. Brandt, Mohammed Alshalalfa, Andrea Sboner, Felix Feng, Shangqian Wang, Himisha Beltran, Tamara Lotan, Martin Spahn, Marianna Kruithof-de Julio, Yu Chen, Karla V. Ballman, Francesca Demichelis, Salvatore Piscuoglio and Mark A. Rubin.
* *first co-authorship. Nature Communications. 2020 Nov 3;11(1):5549.*

- Work on clinicopathological and molecular classification of castration-resistant prostate cancer (referenced in Part III) has been published in:

Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer.

Himisha Beltran, Davide Prandi, Juan Miguel Mosquera, Matteo Benelli, Loredana Puca, Joanna Cyrta, Clarisse Marotz, Eugenia Giannopoulou, Balabhadrapatruni V.S.K. Chakravarthi, Sooryanarayana Varambally, Scott A. Tomlins, David M. Nanus, Scott T. Tagawa, Eliezer M. Van Allen, Olivier Elemento, Andrea Sboner, Levi A. Garraway, Mark A. Rubin, and Francesca Demichelis. *Nature Medicine. 2016 Mar; 22(3): 298–305.*; and:

Genomic correlates of clinical outcome in advanced prostate cancer.

Wassim Abida*, Joanna Cyrta*, Glenn Heller, Davide Prandi, Joshua Armenia, Ilsa Coleman, Marcin Cieslik, Matteo Benelli, Dan Robinson, Eliezer M. Van Allen, Andrea Sboner, Tarcisio Fedrizzi, Juan Miguel Mosquera, Brian D. Robinson, Navonil De

Sarkar, Lakshmi P. Kunju, Scott Tomlins, Yi Mi Wu, Daniel Nava Rodrigues, Massimo Loda, Anuradha Gopalan, Victor E. Reuter, Colin C. Pritchard, Joaquin Mateo, Diletta Bianchini, Susana Miranda, Suzanne Carreira, Pasquale Rescigno, Julie Filipenko, Jacob Vinson, Robert B. Montgomery, Himisha Beltran, Elisabeth I. Heath, Howard I. Scher, Philip W. Kantoff, Mary-Ellen Taplin, Nikolaus Schultz, Johann S. deBono, Francesca Demichelis, Peter S. Nelson, Mark A. Rubin, Arul M. Chinnaiyan, and Charles L. Sawyers
* *first co-authorship*. ***PNAS***. **June 4, 2019**. **116 (23) 11428-11436**

- Results from Paragraph 2.11 in Part III are part of a manuscript in preparation.

Part II. Small cell carcinoma of the ovary of the hypercalcemic type (SCCOHT): comprehensive molecular characterization of a malignancy driven by loss-of-function SWI/SNF alterations.

1. Background

Small cell carcinoma of the ovary of the hypercalcemic type (SCCOHT) is a very rare cancer, with fewer than 500 cases reported to date in the literature (Tischkowitz et al., 2020). SCCOHT occur in young women, with an average age at diagnosis of 24 years (Young, Oliva, & Scully, 1994) and carry dismal prognosis, with an overall long-term survival of 10-20% (Tischkowitz et al., 2020). A phase II trial conducted at Institut Gustave Roussy reported a 3-year overall survival of 49%, despite an aggressive multimodal treatment combining chemotherapy with the PAVEP regimen (cisplatin, adriamycin, vepeside and cyclophosphamide), radical surgery and high dose chemotherapy with autologous stem cell transplant (Pautier et al., 2007). In a recent multicentric prospective study, intensive treatment with cytoreductive surgery, multidrug therapy, high-dose chemotherapy and pelvic radiotherapy led to a median overall survival of 25.7 months, with a plateau around 38 months, and outcomes were better in patients who underwent intensification with hematological transplant; however, significant toxicity was noted for these regimens (Blanc-Durand et al., 2020).

The genomic characteristics of SCCOHT remained poorly understood until 2014, when four teams independently reported the existence of recurrent loss-of-function mutations in the *SMARCA4* gene in SCCOHT, with germline *SMARCA4* mutations in about 40% of cases (Jelinic et al., 2014; Kupryjanczyk et al., 2013; Ramos et al., 2014; Witkowski et al., 2014). This finding has since been confirmed by other studies. It was also demonstrated that in addition to deleterious *SMARCA4* alterations, SCCOHT are characterized by concomitant complete loss of expression of the paralog *SMARCA2* subunit (Karnezis et al., 2016). Pathological diagnosis combines clinical presentation (ovarian tumor in a young female, usually advanced stage),

histomorphology (poorly differentiated malignant neoplasm composed of poorly cohesive, round or polygonal cells with a rhabdoid morphology) and IHC (combined, complete loss of SMARCA4 and SMARCA2 expression).

Nevertheless, many questions remain unanswered regarding the pathogenesis of SCCOHT. First, their histological nature has not been elucidated to date. The 2014 WHO Classification of Tumours of Female Reproductive Organs (4th edition) places SCCOHT among “miscellaneous” tumors (Kurman RJ, 2014). Given that SCCOHT can show focal expression of various markers, including cytokeratins and neuroendocrine/neural markers, an epithelial nature for this tumor was initially proposed, hence their current name (McCluggage et al., 2017). However, this has never been confirmed by ultrastructural or functional studies, and the observed IHC patterns may be a result of aberrant multi-lineage differentiation, similar to what has been reported in MRT (Panwalkar et al., 2020). Conversely, there is some evidence for a possible germ cell origin of SCCOHT. Indeed, association of an ovarian germ cell neoplasms (such as immature teratoma, yolk sac tumor or dermoid cyst) with SCCOHT has been reported in several extensively sampled cases, and focal expression of germ cell markers (such as SALL4) can be seen in SCCOHT cells (Kupryjanczyk et al., 2013; McCluggage et al., 2017).

Most importantly, effective treatment options against SCCOHT are still lacking. Despite initial chemosensitivity, these tumors invariably relapse. Thus, identifying potential therapeutic targets or strategies represents an unmet clinical need.

The aim of this study was to conduct a comprehensive molecular characterization in an independent cohort of SCCOHT patients treated at Institut Gustave Roussy, with the overarching goal of broadening knowledge about this rare malignancy.

2. Results

2.1 Clinical and pathology findings

We identified 8 patients diagnosed with SCCOHT between 2002 and 2011, for whom fresh-frozen tumor material was available. The mean age at diagnosis was 31 years (range, 14-40). All patients for whom disease stage information was available were diagnosed with FIGO stage III or IV tumors. No patient showed objective hypercalcemia at diagnosis. Three patients had received neoadjuvant chemotherapy. After a median follow-up of 10 months (range, 3-36), seven patients had died of disease and one patient was in remission (alive and disease-free) 36 months from initial diagnosis. **Table 3** summarizes the available clinical findings and treatments.

Patient study ID	Age at Dx	Sample type	Sample status regarding chemo	Personal hx	Family cancer hx	Stage at Dx	Hyper-Ca at Dx?	Initial surgery type	Medical treatments	RXT	Follow-up* [months]	Outcome
IGR01	14	Resection	chemo-exposed	-	-	III	No	Debulking + CHIP	VIP (5 cycles) + Doxorubicin	NA	6	DOD
IGR02	26	Resection	chemo-naïve	-	-	NA	No	Debulking	PAVEP x 4, CARBOPEC + autologous SCT	-	36	ADF
IGR03	30	Resection	chemo-naïve	-	-	III	No	Debulking	PAVEP 6 cycles (C4-6 without Doxorubicin)	-	7	DOD
IGR04	40	Resection	chemo-exposed	-	-	III	No	Annexectomy	BEP x 1 and PAVEP x 5	-	10	DOD
IGR05	29	Resection	chemo-naïve	NA	NA	NA	NA	NA	NA	NA	12	DOD
IGR06	35	Resection	chemo-exposed	Endometriosis	Hodgkin disease (grandfather)	IV	No	Annexectomy	EP x 1 and PAVEP x 2 Debulking PAVEP x3 CARBOPEC + autologous SCT Topotecan/Taxol	Yes	12	DOD
IGR07	34	CT-guided biopsy	chemo-naïve	-	Pancreatic cancer (grandmother)	III	NA	No surgery (poor PS)	PAVEP	NA	3	DOD
IGR08	39	Resection	chemo-naïve	NA	NA	NA	NA	NA	NA	NA	NA	DOD

Table 3. A summary of the available clinical data for the cohort. Dx: diagnosis; Hx: history; HyperCa: hypercalcemia; DOD: died of disease; ADF: alive, disease free; CT: computer tomography; SCT: stem cell transplant; PS: performance status; RXT: radiation therapy; NA: not available. *time from initial diagnosis to most recent clinical information.

Histomorphology of all cases showed features characteristic of SCCOHT, which included: sheets of poorly differentiated cells with a rhabdoid morphology (large nuclei with open chromatin, prominent nucleoli, perinuclear hyaline cytoplasmic inclusions) and/or a “small round blue cell” morphology, occasional formation of pseudo-follicles, conspicuous mitotic activity, and foci of tumor necrosis. **Fig.7** shows examples of SCCOHT cases from this cohort.

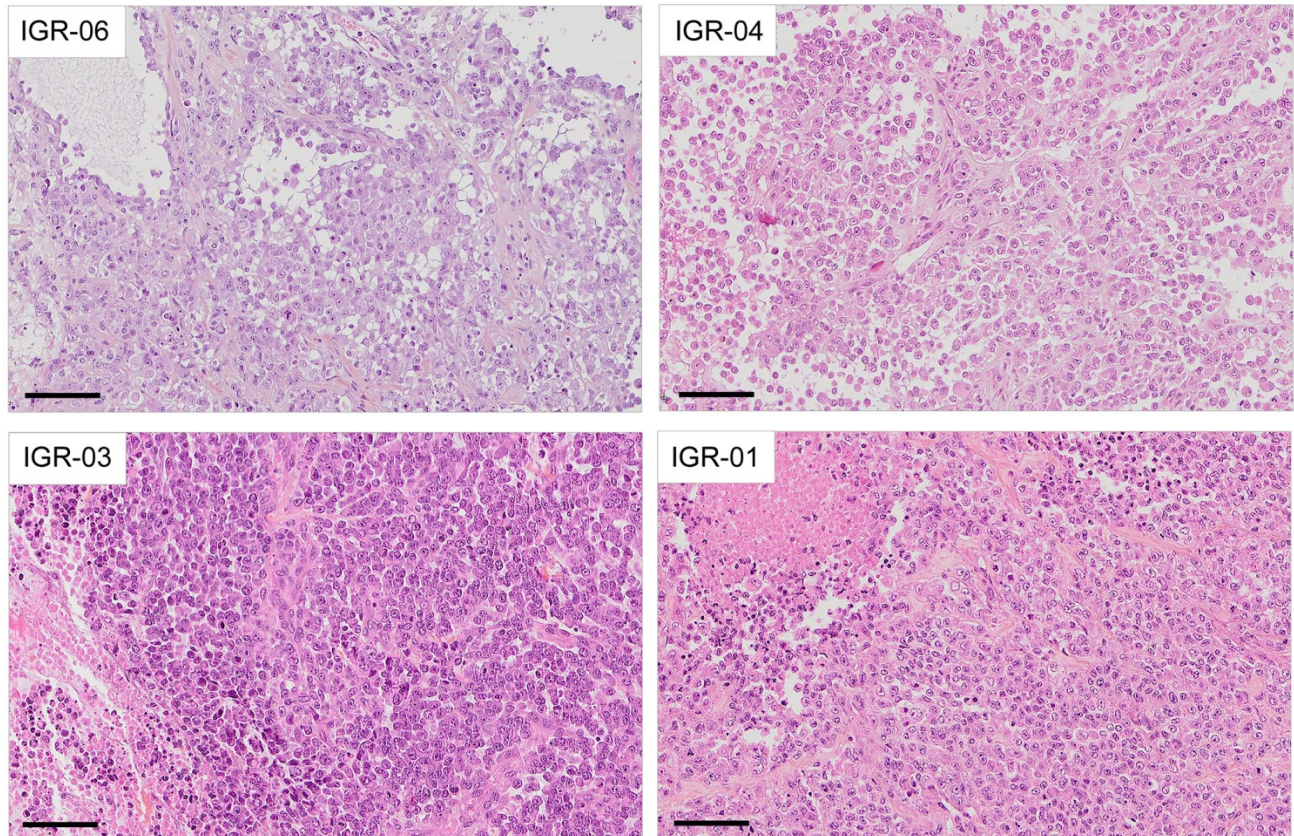


Figure 7. Histomorphology of representative SCCOHT cases from this cohort. HES stain, scale bar: 100 μ m.

2.2 Genomic profiles of SCCOHT: general findings

WES was performed on 6 tumor-normal pairs. The mean depth of coverage was 109X, with at least 98% of the targeted exome covered by at least 10 reads and 95% showing a read quality score (QC) ≥ 30 .

The first notable finding was a relatively low mutation burden (TMB) in SCCOHT. The TMB, calculated based on the number of non-synonymous, somatic-only mutations in coding regions per Mb of DNA, showed a median of 5.60 mutations/Mb (mean, 5.43 mutations/Mb; range: 3.56-6.42) using the parameters specified in Methods. Among these genes, very few showed somatic-only SNVs in more than one sample. To include alterations potentially related to a Loss-Of-Heterozygosity (LOH) event in our analysis, the filters applied were: variant allele frequency (VAF) higher in tumor than in normal tissue; somatic p-value<0.001 (Fisher's exact test); and location in a coding region. Using these cut-offs, 500 variants in 335 genes were retained. Fourteen genes were altered in at least 3 of the 6 samples: *SMARCA4* (5 samples); *PLK5*, *ANKRD24* and *FBN3* (4 samples); *ABCA7*, *ACTL9*, *CACTIN*, *EMR1*, *FUT5*, *GRIN3B*, *KANK3*, *KRI1*, *LINGO3* and *PLIN4* (3 samples). Remarkably, all of these genes were located in the 19p13.2-3 locus and carried variants with high tumor VAF (mean, 0.87), strongly suggesting a recurrent LOH event in 19p13.2-3.

To further characterize these alterations, we queried their frequency in the general population (using the gnomAD v2.1.1 database) and their potential functional impact (using the Polyphen2 prediction tool) (**Fig.8**). The majority (56/64) of these variants were known polymorphisms (variant frequency in the general population $\geq 1\%$) and 45/64 were classified as benign by the Polyphen-2 classifier. Variants that were not polymorphisms and that were classified as potentially or probably damaging, or for which functional prediction scores were not available, included: four different variants in *SMARCA4* in 5 patients (detailed below), the p.G223V variant in *PLK5* (1 patient), the p.R220H variant in *ACTL9* (1 patient) and the c.4208delT frameshift in *ABCA7* (1 patient). All of these variants were Sanger-verified.

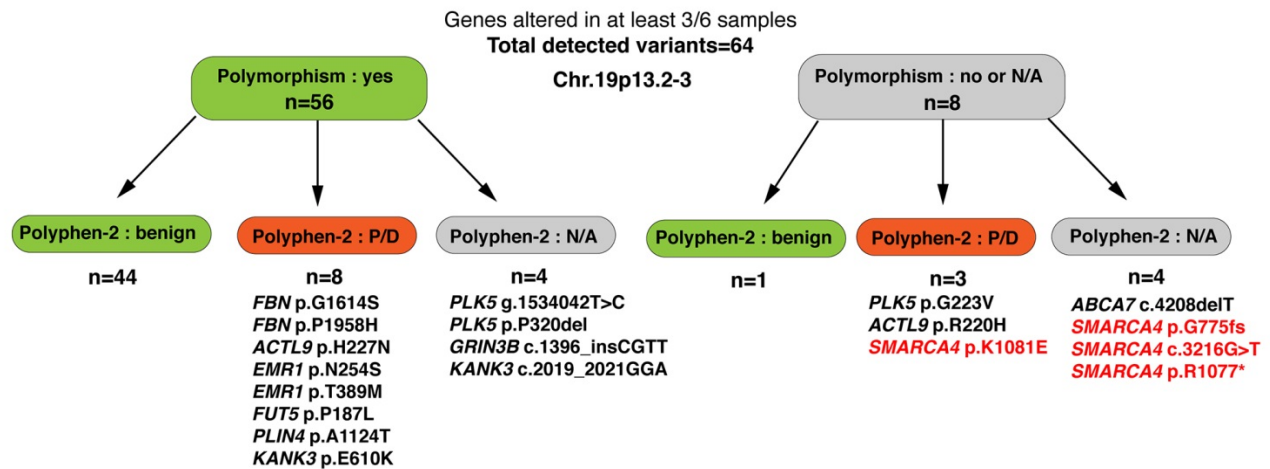


Figure 8. A breakdown of variants detected in SCCOHT by WES. Variants classified as potentially or probably damaging, or for which functional prediction scores were not available, are explicitly listed. Variants in the *SMARCA4* gene are highlighted in red.

Several known cancer-related genes were additionally queried to exclude the presence of alterations at low allele frequencies: *TP53*, *KRAS*, *PIK3CA*, *PTEN*, *BRAF*, *EGFR*, *AKT1*, *CDKN1A* (*p21*) and *ERBB2*. No alterations in those genes were found.

2.3 *SMARCA4* loss-of-function mutations in SCCOHT and related findings

In line with previous studies, *SMARCA4* was mutated by WES in 5/6 cases from our cohort, with a high mean allele frequency (0.86), consistent with biallelic alterations. The encountered *SMARCA4* mutations are shown in **Table 4**. In two patients, these alterations were associated with the presence of a germline variant and a loss of heterozygosity (LOH) in the tumor. The observed *SMARCA4* alterations included: truncating alterations in three cases (including an identical p.Arg1077* variant in two different patients), one splice site mutation, and one missense mutation (p.Lys1081Glu, predicted as probably deleterious by the Polyphen-2 tool). Truncating *SMARCA4* alterations were located upstream of the SNF2-ATP coupling domain of *SMARCA4* is in favor of a deleterious functional impact (**Fig.9**).

Interestingly, in one tumor diagnosed as SCCOHT (IGR-03), no *SMARCA4* mutation was detected, but instead, concomitant and potentially biallelic loss-of-function alterations in two other SWI/SNF genes were found: *ARID1A* (2 frameshifts: p.Gln555fs and p.Thr1004fs) and *ARID1B* (stop gained Arg1944*).

Of note, in addition to a deleterious *SMARCA4* mutation, case IGR-01 showed a p.Arg635* stop gain in the *SMARCA1* gene, which encodes the ATP-ase of another chromatin remodeling complex, ISWI (VAF=0.37).

Case	SWI/SNF gene(s)	Mutation(s)	Germline VAF	Somatic VAF
IGR-01	<i>SMARCA4</i>	p.Asn775fs	0.42	0.9
IGR-02	<i>SMARCA4</i>	p.Lys1081Glu	0	0.76
IGR-03 ^a	<i>ARID1A</i> <i>ARID1A</i> <i>ARID1B</i>	p.Gln555fs p.Thr1004fs p.Arg1944*	0 0 0	0.16 0.18 0.28
IGR-04	<i>SMARCA4</i>	p.Arg1077*	0.45	0.86
IGR-05	<i>SMARCA4</i>	c.3216-1G>T splice	0	0.98
IGR-06	<i>SMARCA4</i>	p.Arg1077*	0	0.73

Table 4: *SMARCA4* mutations found in SCCOHT from this cohort. ^aindicates the case in which no *SMARCA4* alteration was found. VAF: variant allele frequency, * indicates stop gain.

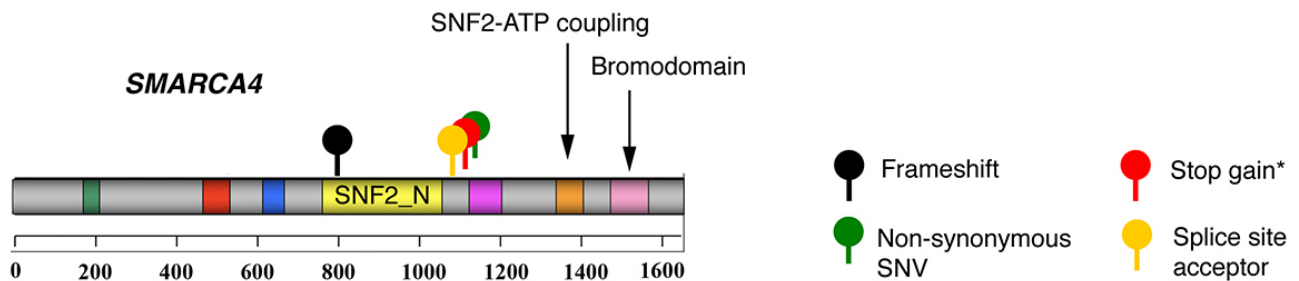


Figure 9. Schematic representation of *SMARCA4* mutations found in this cohort. *indicates that this identical mutation was found in two different patients.

SMARCA4 (BRG1) protein expression in these 6 cases by IHC was previously reported in (Genestie et al., 2020) and was not reviewed as part of the current study. Genestie et al. reported loss of SMARCA4 (BRG1) protein expression in all cases that displayed *SMARCA4* mutations (5 cases), while the one tumor with no *SMARCA4* mutation detected (IGR-03) showed retained SMARCA4 expression. The pathology features of this tumor were consistent with SCCOHT diagnosis and in particular, tumor cells displayed a prominent rhabdoid morphology (**Fig.7**).

2.4 SMARCA2 loss of expression in SCCOHT and related findings

Several studies have previously shown that *SMARCA4* loss in SCCOHT is accompanied by complete loss of *SMARCA2* (BRM) expression at the protein level, which can be evidenced by immunohistochemistry (Karnezis, Wang, et al., 2016). Accordingly, the 5 SCCOHT cases from our cohort with proven *SMARCA4* mutation by WES also displayed a complete loss of *SMARCA2* (BRM) expression by IHC (not shown). In line with previous studies, WES did not reveal loss-of-function mutations in the *SMARCA2* gene in our patients that could explain this loss of expression, suggesting that *SMARCA2* loss of expression may be due to epigenetic or post-transcriptional silencing of the gene.

We also validated loss of *SMARCA2* (BRM) expression by Western blotting in the two currently available SCCOHT cell lines, BIN67 and SCCOHT-1 (**Fig.10**).

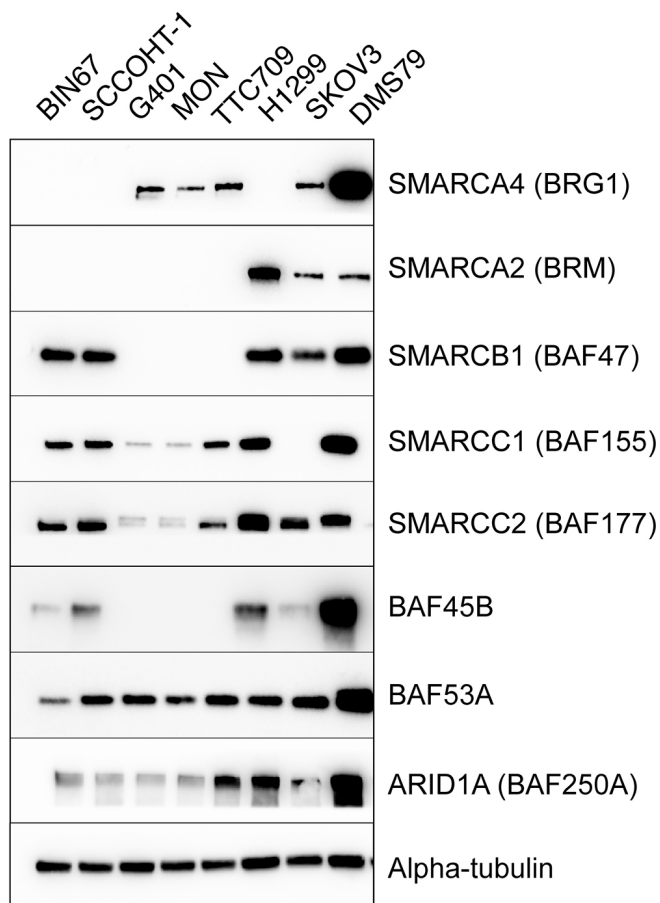


Figure 10. Western blot showing the expression of selected SWI/SNF subunits in human cell lines. The cell lines are SCCOHT (BIN-67, SCCOHT-1), MRT (G401, MON, TTC709), SMARCA4-deficient lung adenocarcinoma (H1299), high-grade endometrioid adenocarcinoma of the ovary (SKOV3) and neuroendocrine small cell lung cancer (DMS79).

SMARCA2 silencing has previously been linked to the existence of insertional polymorphisms in the *SMARCA2* promoter, located -741 bp and -1321 bp, respectively, from the transcription start site, in lung cancers (Liu et al., 2011) and in MRT cells (Kahali et al., 2014). This silencing could possibly occur through increased binding to negative regulators of transcription, such as HDAC9 or MEF2D, to these polymorphic sites (Kahali et al., 2014). However, the status of these promoter polymorphism sites in SCCOHT has not been investigated to date. We performed Sanger sequencing of these sites in 8 patient SCCOHT samples (DNA extracted from fresh-frozen tumor tissue) and in two SCCOHT cell lines (**Table 5**). We did not observe any cases with a homozygous polymorphism at both sites. One tumor and one cell line (SCCOHT-1) harbored a homozygous -741 polymorphism, another tumor was homozygous for the -1321 promoter site polymorphism, and all other cases displayed a

heterozygous -741 and -1321 polymorphism status (**Fig.11**). The G401 MRT cell line, used as control, showed a homozygous polymorphism at both sites, in keeping with previously published results (Kahali et al., 2014). Taken together, our results suggest that *SMARCA2* loss of expression in SCCOHT cannot be solely explained by the presence of previously described homozygous polymorphisms in the *SMARCA2* promoter.

Sample	-741 bp polymorphism site status	-1321 bp polymorphism site status
IGR-01	Heterozygous	Heterozygous
IGR-02	Heterozygous	Heterozygous
IGR-03	Heterozygous	Homozygous
IGR-04	Heterozygous	Homozygous
IGR-05	Heterozygous	Heterozygous
IGR-06	Heterozygous	Heterozygous
IGR-07	Homozygous	Heterozygous
BIN-67 cell line	Heterozygous	Heterozygous
SCCOHT-1 cell line	Homozygous	Heterozygous
G401 cell line	Homozygous	Homozygous

Table 5. Results of Sanger sequencing of the two promoter polymorphism sites.

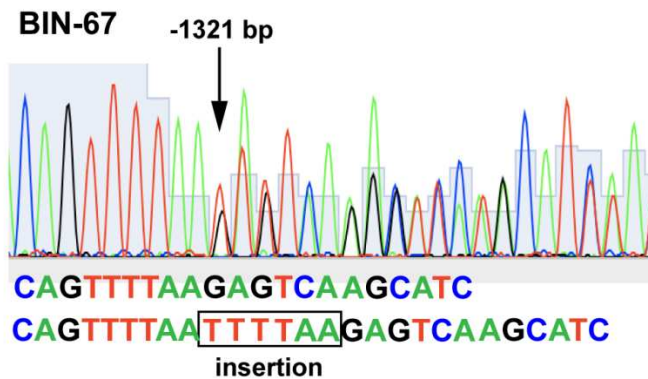


Figure 11. An example of Sanger sequencing results at the -1321 bp *SMARCA2* promoter polymorphism site in the BIN67 cell line (result: presence of a heterogeneous 6 bp insertion polymorphism).

Combined *SMARCA4* and *SMARCA2* loss of expression is also seen in a subset of MRTs and AT/RTs (Hasselblatt et al., 2011) and in *SMARCA4*-deficient thoracic sarcomas (*SMARCA4*-DTS)(Le Loarer et al., 2015). On IHC, *SMARCA4*-DTS have been reported to show strong and diffuse positivity for SOX2 in nearly all cases. Thus, SOX2 was suggested as a

potentially useful diagnostic marker for differentiating SMARCA4-DTS from *SMARCA4*-mutated lung adenocarcinoma (Le Loarer et al., 2015; Perret et al., 2019; Yoshida et al., 2017). The transcription factor SOX2 is one of the “four factors” (“Yamanaka factors”) involved in reprogramming of somatic cells into induced pluripotent stem cells (Takahashi & Yamanaka, 2006); in cancer, it has been shown to promote dedifferentiation and to be associated with aggressive phenotypes (Ben-Porath et al., 2008; Leis et al., 2012; Mu et al., 2017). In MRT and AT/RT, SOX2 expression can be variably encountered (Okuno et al., 2010; Venneti et al., 2011). To our knowledge, SOX2 expression in SCCOHT has not been queried to date. We performed SOX2 IHC on 10 SCCOHT patient tumor samples and observed complete lack of expression in 6 cases (**Fig.12**), positivity in scattered single tumor cells in 2 cases, and focal areas of positivity (accounting for <10% of the tumor surface) in 2 cases. Thus, recurrent and diffuse SOX2 expression does not seem to be a feature of SCCOHT, contrarily to SMARCA4-DTS. This further suggests the existence of potential biological differences between various tumor types with SWI/SNF alterations, even when considering tumors with seemingly identical alterations (here: combined SMARCA4 and SMARCA2 loss), which could be in line with context-dependent roles of these alterations.

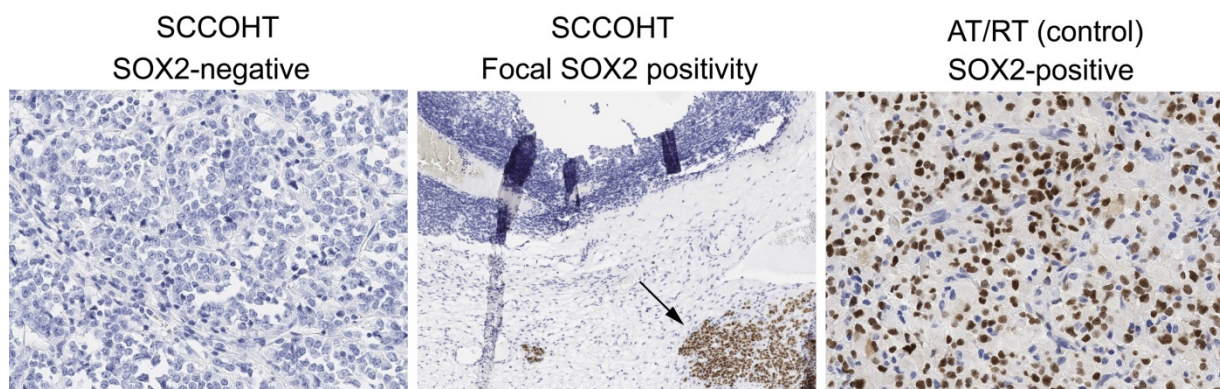


Figure 12. Representative results of SOX2 immunohistochemistry in SCCOHT. SOX2 staining was completely negative in the majority of cases (first image). Two cases showed focal tumor areas of limited size with SOX2 positivity (second image, arrow). A case of AT/RT positive for SOX2 is shown as control (third image).

2.5 Copy number profiles of SCCOHT

Eight fresh-frozen tumors were subjected to aCGH analysis (**Fig.13**). The aCGH profiles of seven tumors showed remarkable genomic stability, with few SCNAs, and only one tumor exhibited some genomic instability (IGR-07) (no WES results were available for this tumor).

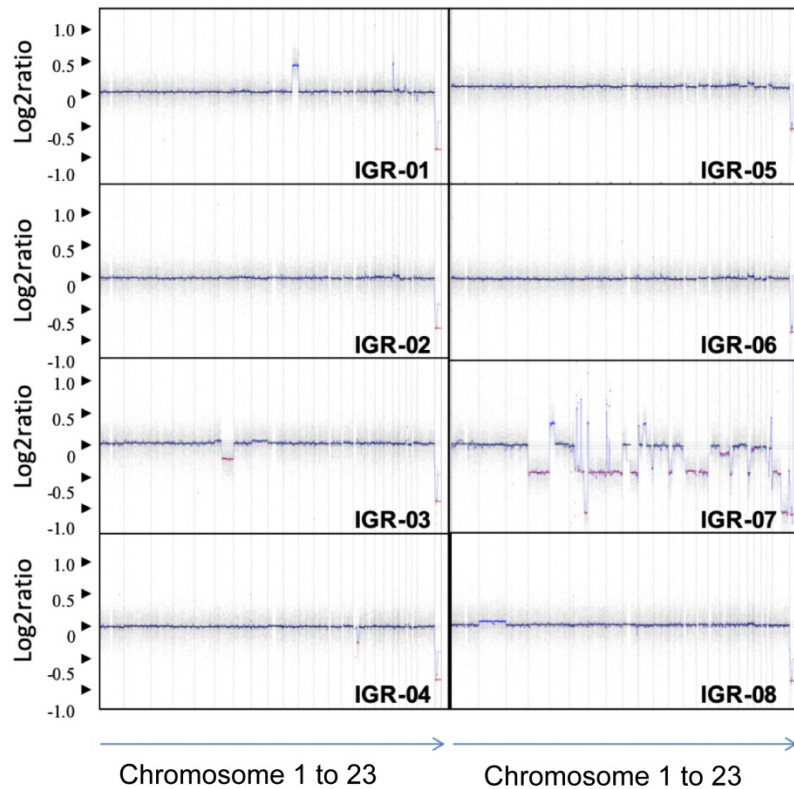


Figure 13. Genome-wide aCGH profiles of the 8 SCCOHT cases from this study.

WES confirmed recurrent loss-of-heterozygosity (LOH) at the 19p13.2-3 region in five of six tumors, and the smallest common LOH region was Chr19:373.916-11.465.316 (**Fig.14A and B**). On aCGH, no copy number losses were present in this region, supporting the presence of a recurrent copy neutral-LOH (CN-LOH) event (**Fig.14C**).

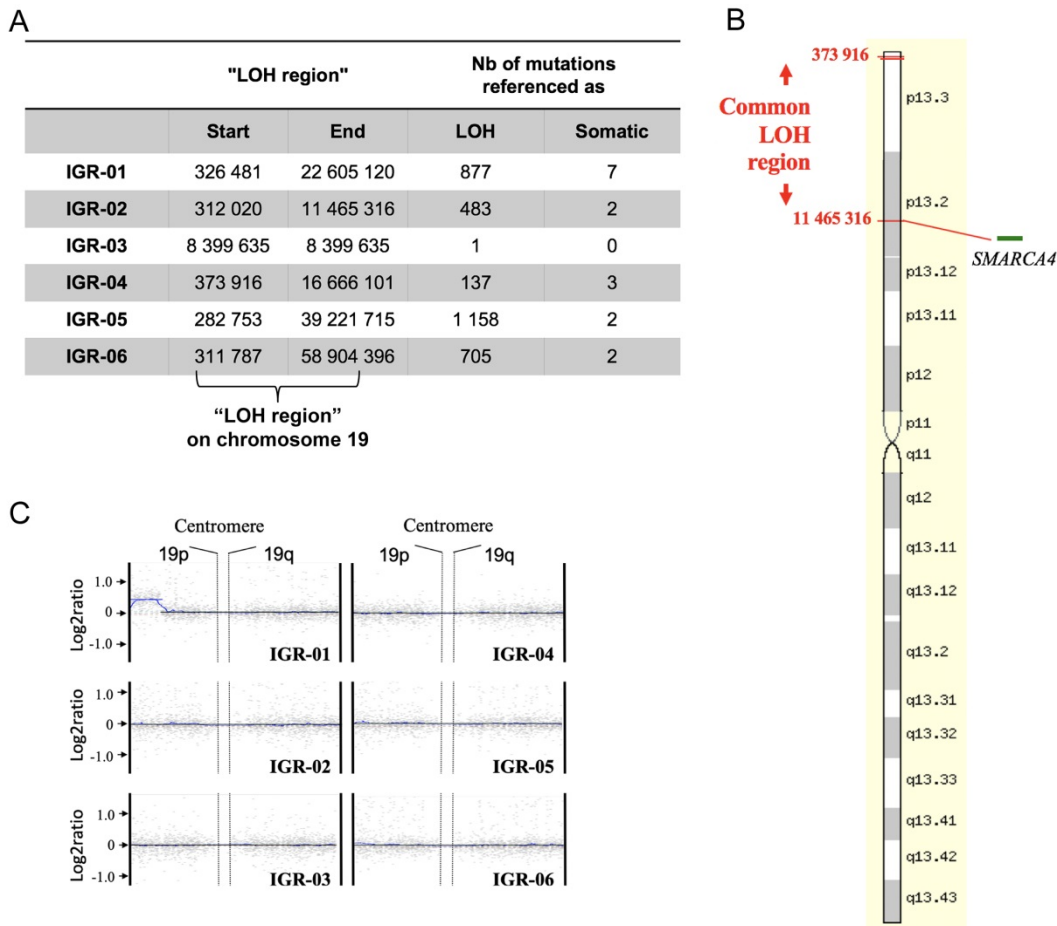


Figure 14. Identification of a common LOH region among 6 SCCOHT cases. A. Coordinates of the LOH region in 19p13.2-3 seen by WES in each tumor. **B.** Schematic representation of chromosome 19 and the smallest common LOH region. **C.** aCGH profiles centered on chromosome 19 objectify copy-neutral profiles for the 6 tumors.

Neither aCGH nor WES identified recurrent copy number alterations (arbitrarily defined as present in at least half of the tumors). However, by aCGH, sixteen genes showed copy number gains in at least three of the eight tumors. These genes were located at 12q13.3, 12q32.2, and 16q24.3. Four of these genes showed a $\log_2(\text{ratio}) > 2.3$, which represents a five-fold increase in copy number, suggestive of amplification: *SHMT2*, *NDUFA4L2*, *LRP1* and *NXPH4*.

2.6 Transcriptomic profiles of SCCOHT

Given that, by definition, tumor suppressor loss is not directly druggable, we decided to use the RNA-seq data generated in our cohort to perform a differential expression analysis and try to identify some highly overexpressed genes, which could represent putative therapeutic targets. We compared RNA-seq based gene expression levels in 6 SCCOHT samples *versus* 5 samples of benign ovarian tissue from the GTEx database. Samples IGR-03 (no *SMARCA4* mutation) and IGR-07 (showing some degree of genomic instability on CGH array) were excluded from the analysis, to ensure a more homogenous nature of the SCCOHT cohort. The five GTEx samples were selected from donors <35 years of age, to try to match the age of SCCOHT patients. To minimize batch effect, the analysis was conducted on rank-normalized gene expression values.

Using this approach, ~1900 genes were found to be significantly differentially expressed (\log_2 fold change >1 or <-1 and p_{adj} <0.05). Selected top and bottom significantly deregulated genes are shown in **Fig.15A**. Among the top significantly upregulated genes, we identified some putative therapeutic targets, such as *MAGEA4* (which encodes a cancer-testis antigen and which was the most significantly overexpressed gene: mean rank-normalized expression values 0.74 and 0.06, respectively, \log_2 fold change=3.59, p_{adj} =0.015), *AURKB* (which encodes the Aurora B kinase: mean rank-normalized expression values 0.73 and 0.19, respectively, \log_2 fold change=1.92, p_{adj} =0.015), and *ERBB4* (mean rank-normalized expression values 0.76 and 0.21, respectively, \log_2 fold change=1.87, p_{adj} =0.045). Other overexpressed genes included genes encoding metalloproteinases (*MMP10*, *MMP9*); genes related to neural development (e.g. *NCAM2*, *NTS*, *ATCAY*, *CBLN2*); genes related to osteogenic differentiation (e.g. *BMP7*) and embryonic genes (e.g. *CLDN6*, which encodes an embryonic cell junction protein). Conversely, and as a means of quality control of the analysis, the expression of some genes known to be highly expressed in benign ovarian tissue (e.g. *INHA*, *FOXL2*, *AMHR2*) was significantly lower in SCCOHT.

By Gene Set Enrichment Analysis (GSEA)(**Fig.15B**), gene sets significantly enriched in SCCOHT were related to E2F targets and cell cycle progression and activation of oncogenic pathways (KRAS, MYC, mTORC1). There was also a significant enrichment of the gene set representing “genes downregulated upon SMARCB1 knock-down” (SNF5_DN.V1_UP), consistent with a deregulated SWI/SNF complex in SCCOHT.

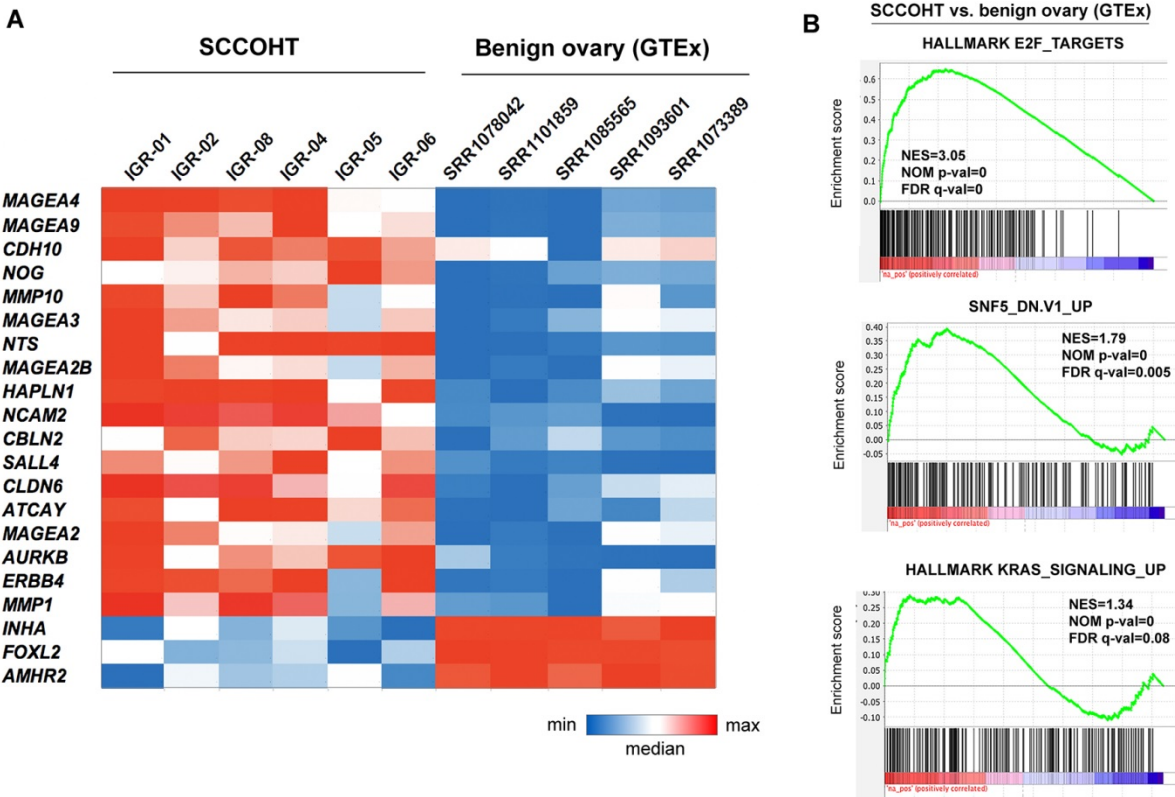


Figure 15. Transcriptomic data (RNA-seq) in 6 SCCOHT cases from this cohort and in 5 samples of benign ovarian tissue from GTEx. A. Heatmap representation of gene expression levels for selected genes, significantly deregulated in the differential expression analysis. **B.** Gene Set Enrichment Analysis - selected gene sets showing a significant positive enrichment in SCCOHT.

SMARCA2 expression was significantly lower in SCCOHT than in benign ovarian tissue (mean rank-normalized expression values: 0.50 vs. 0.95, log2 fold change=-0.92, padj=0.015), although with a log2 fold change at -0.92. There was no statistically significant difference in SMARCA4 expression, likely due to low expression in two samples of benign ovarian tissue.

Of note, the expression of most other SWI/SNF subunits was maintained at the mRNA level and was higher in SCCOHT than in benign ovarian tissue (**Fig.16**), with the exception of DPF3, which was significantly downregulated (mean rank-normalized expression values: 0.17 in SCCOHT vs. 0.71 in benign ovarian tissue, \log_2 fold change=-2.07, $\text{padj}=0.015$). This retained expression of other SWI/SNF subunits in SCCOHT is consistent with our Western blot results from **Fig.10** and in line with findings by Pan et al. suggesting that a residual, aberrant complex exists in SCCOHT despite the absence of the catalytic subunits (Pan et al., 2019).

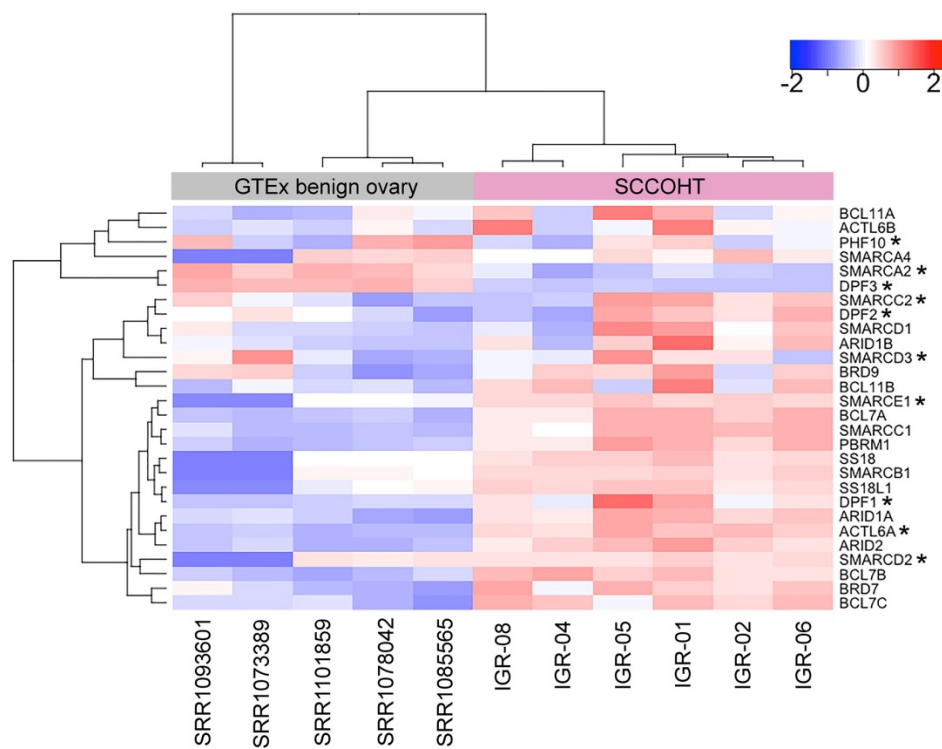


Figure 12. Heatmap representation of gene expression levels for SWI/SNF genes in 6 cases of SCCOHT from this cohort and 5 samples of benign ovarian tissue (GTEx). Genes for which the differential expression reached statistical significance ($\text{padj}<0.05$) are marked with an asterisk.

None of the four genes that showed recurrent amplifications in SCCOHT (*NDUFA4L2*, *SHMT2*, *NXPH4*, *LRP*) was significantly overexpressed. *PLK5* showed very low expression values in both groups (mean rank-normalized expression values 0.18 and 0.17, respectively).

Lastly, no significant overexpression of *SOX2* was noted in SCCOHT at the mRNA level, consistent with IHC data. No significant overexpression of *PTH1L*, which encodes Parathyroid Hormone-Related Protein (which has been postulated to cause the hypercalcemia observed in a subset of SCCOHT, although this has never been confirmed), was observed.

To verify whether these mRNA results could have correlates at the protein level, we performed a proof-of-principle anti-Claudin-6 IHC in 13 FFPE patient samples of SCCOHT. Five cases (39%) showed membranous positivity for Claudin-6, including weak expression in 3 cases and moderate expression in 2 cases, suggesting that Claudin-6 is indeed expressed at the protein level in a subset of SCCOHT patients (**Fig.17**).

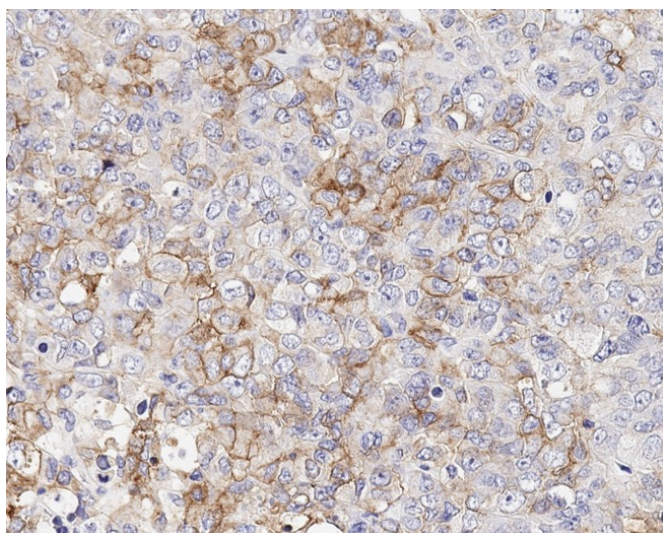


Figure 17. An example of anti-Claudin-6 IHC in a case of SCCOHT (IGR-06), showing moderate expression at the cell membrane.

2.7 Epigenetic vulnerabilities in SCCOHT.

In a recent study, Pan et al. have shown that the loss of catalytic SWI/SNF activity in SCCOHT largely alters SWI/SNF functions as an epigenetic regulator (Pan et al., 2019). To assess the putative sensitivity of SCCOHT to currently available epigenetic treatments, we used the histone deacetylase inhibitor trichostatin A (TSA) and the DNA methyltransferase inhibitor 5'-azacytidine (5'-dAZAC), to treat cell lines with differing *SMARCA4* genotypes for 72h (**Fig.18**). The SCCOHT cell line BIN-67, which harbors an inactivating *SMARCA4* mutation and shows complete loss of *SMARCA2* expression (**Fig.10**), was exquisitely sensitive to 5'-dAZAC

and TSA at subnanomolar concentrations (**Fig.18A,B**). Conversely, the H1299 lung adenocarcinoma cell line, which carries a *SMARCA4* mutation, but shows retained *SMARCA2* expression, and the ovarian high-grade endometrioid adenocarcinoma cell line SKOV3 (*SMARCA4+*/*SMARCA2+*) were completely resistant to 5'-dAZAC and 100-fold less sensitive to TSA than BIN-67.

Many studies suggest an antagonistic relationship between the SWI/SNF complex and Polycomb proteins, such as Enhancer of Zest 2 (EZH2) (Wilson et al., 2010). Combined loss of *SMARCA4* and *SMARCA2* in SCCOHT cell lines may induce an oncogenic dependency on EZH2 activation (Kim et al., 2015) and confer extreme sensitivity to EZH2 inhibitors *in vitro* and *in vivo* (Y. Wang et al., 2017). Based on this rationale, a patient with *SMARCA4*-mutated SCCOHT was enrolled in a phase I trial of tazemetostat (EPZ-6438), a highly selective EZH2 inhibitor (Italiano et al., 2018). This 25-year-old patient initially presented with stage IV SCCOHT treated with surgery, combination platinum-based chemotherapy followed by high dose consolidation and autologous stem cell rescue (**Fig.18C**). Unfortunately, she relapsed within eight weeks and was, therefore, enrolled in the EZH2 inhibitor clinical trial. She presented partial response (RECIST 1.1) after four months of treatment and remained progression-free for eight months. Although the clinical benefit was relatively short, the degree of response in this patient with highly chemo-resistant disease supports further investigation of epigenetic strategies in SCCOHT.

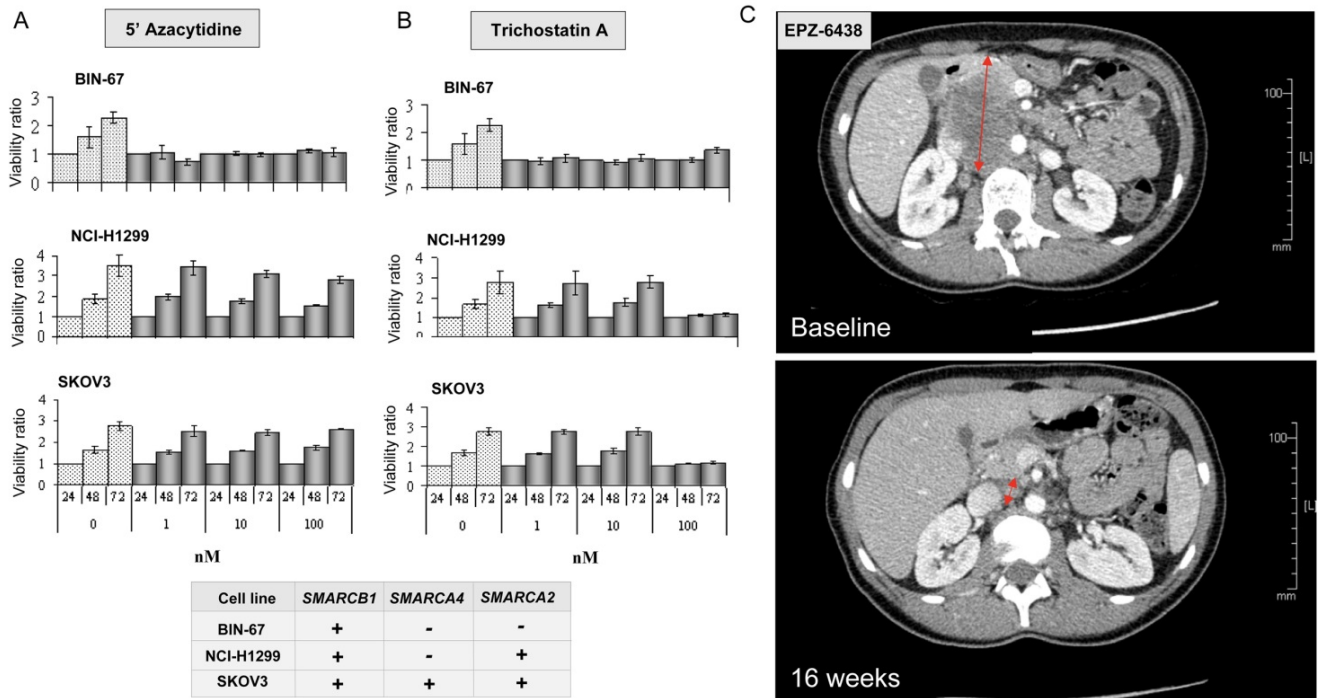


Figure 18. Epigenetic vulnerabilities in SCCOHT. A. Effects of treatment with the DNMT inhibitor 5'-AZAC and the HDAC inhibitor Trichostatin A in cell lines with different SWI/SNF phenotypes. **B.** Effects of treatment with the EZH2 inhibitor EPZ-6438 in a SCCOHT patient as part of a clinical trial.

2.8 SMARCA4 and SMARCA2 expression in extra-cranial MRTs, and relationships between tumors from the “rhabdoid tumor” spectrum.

Malignant rhabdoid tumors and SCCOHT share many similarities at the morphologic and genomic level, including a remarkably stable genome, low tumor mutation burden and a rhabdoid morphology. In addition, several studies using DNA methylation or transcriptomic profiles showed that SCCOHT are more closely related to MRT than to some other tumor types (Fahiminiya et al., 2016; Le Loarer et al., 2015). Thus, SCCOHT have been proposed by some authors to represent the ovarian counterpart of MRT (Witkowski et al., 2014), and the term “MRT of the ovary” has even been used interchangeably with “SCCOHT” by some authors. However, the molecular similarities and differences between SCCOHT and MRT have not been extensively investigated.

While most MRTs show loss-of-function alterations in *SMARCB1*, a subset displays deleterious *SMARCA4* alterations as an alternative event (Hasselblatt et al., 2011; Hasselblatt et al., 2014; Schneppenheim et al., 2010); this situation is more frequent in AT/RTs than in extra-cranial MRTs (ECRT). As part of an ongoing study investigating the molecular characteristics of *SMARCA4*-mutated ECRT led by Dr. Franck Bourdeaut (Institut Curie), we queried the expression of *SMARCA4*, *SMARCA2* and *SMARCB1* in a series of ECRT by IHC (**Fig.19**).

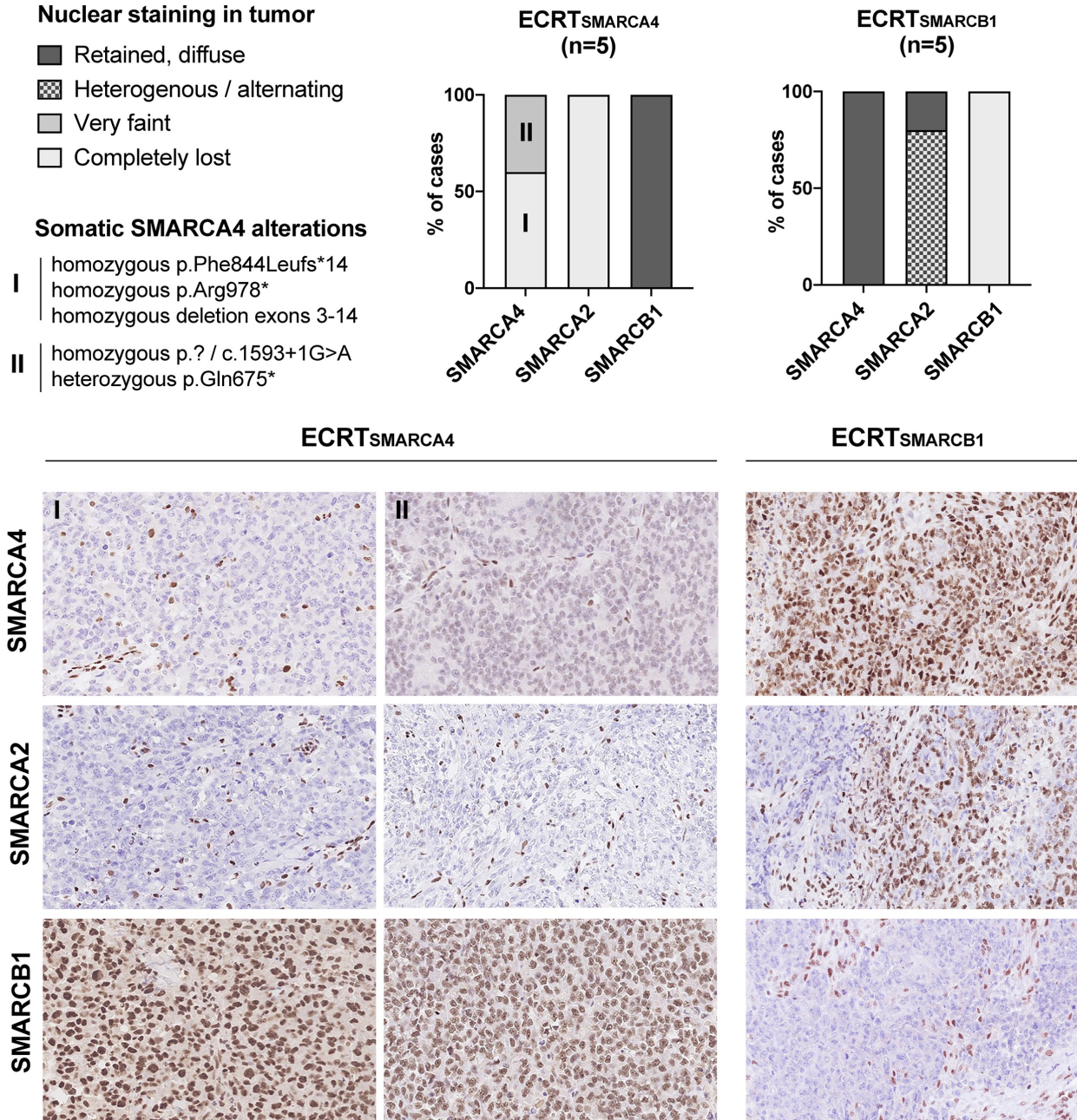


Figure 19. Results of IHC for SMARCA4, SMARCA2 and SMARCB1 in *SMARCA4*-mutated extra-cranial MRTs (ECRT_{SMARCA4}) and *SMARCB1*-mutated extra-cranial MRTs (ECRT_{SMARCB1}).

As expected, SMARCA4 (BRG1) protein expression was abolished in all five cases of *SMARCA4*-mutated ECRT (ECRT_{SMARCA4}). Surprisingly, in two cases of ECRT_{SMARCA4}, we identified very faint residual nuclear staining for SMARCA4 (Fig.20). One of these cases

showed a homozygous splice site mutation (c.1593+1G>A) and the other showed a heterozygous stop gain mutation (p.Gln675*) with no “second hit” identified by the used methods (next generation amplicon-based sequencing using a gene panel, on a NextSeq Illumina® sequencer, minimum depth 100X). This suggests that some genetic alterations (here: a splice site mutation and a truncating alteration with no second hit identified) may be accompanied by low residual levels of the protein.

We also demonstrated that all ECRT_{SMARCA4} from our series displayed a complete and uniform loss of SMARCA2 (BRM) protein expression, which again is consistent with what has been described in SCCOHT and in SMARCA4-DTS (**Fig.19**). Conversely, in the five tested cases of ECRT_{SMARCB1}, BRM expression was retained, albeit heterogeneous, alternating tumor areas with strong expression and areas with complete lack of expression (**Fig.21**).

SMARCA4 negative (n=3)

SMARCA4 weak (n=2)

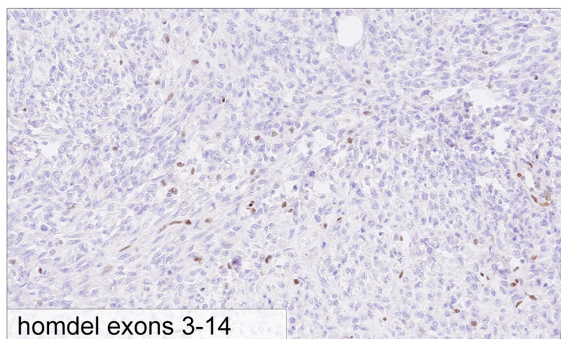
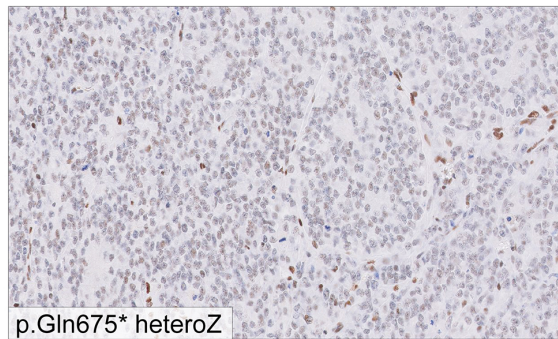
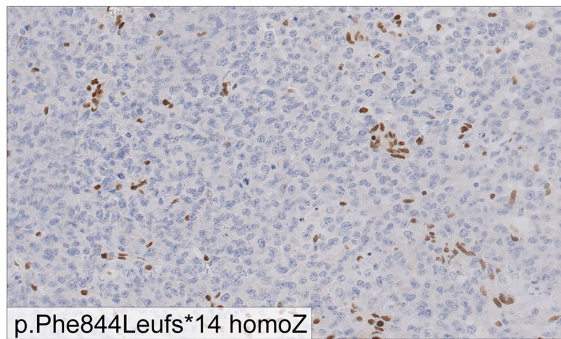
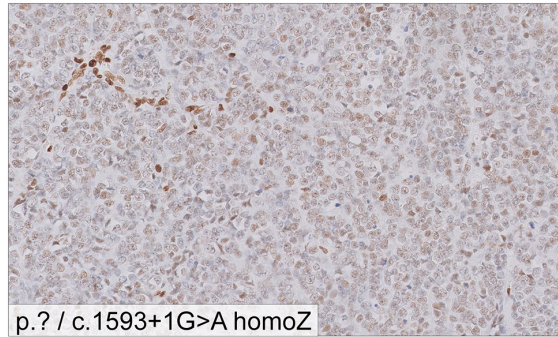
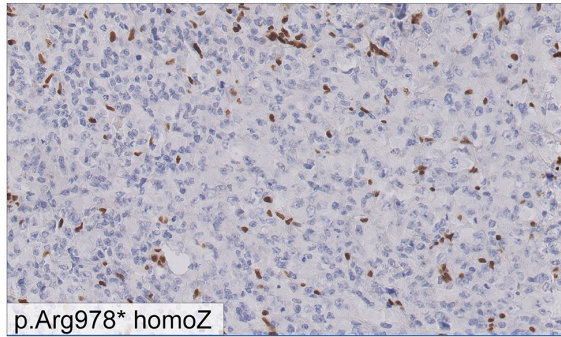


Figure 20. The two SMARCA4 staining patterns encountered in ECRT_{SMARCA4}.

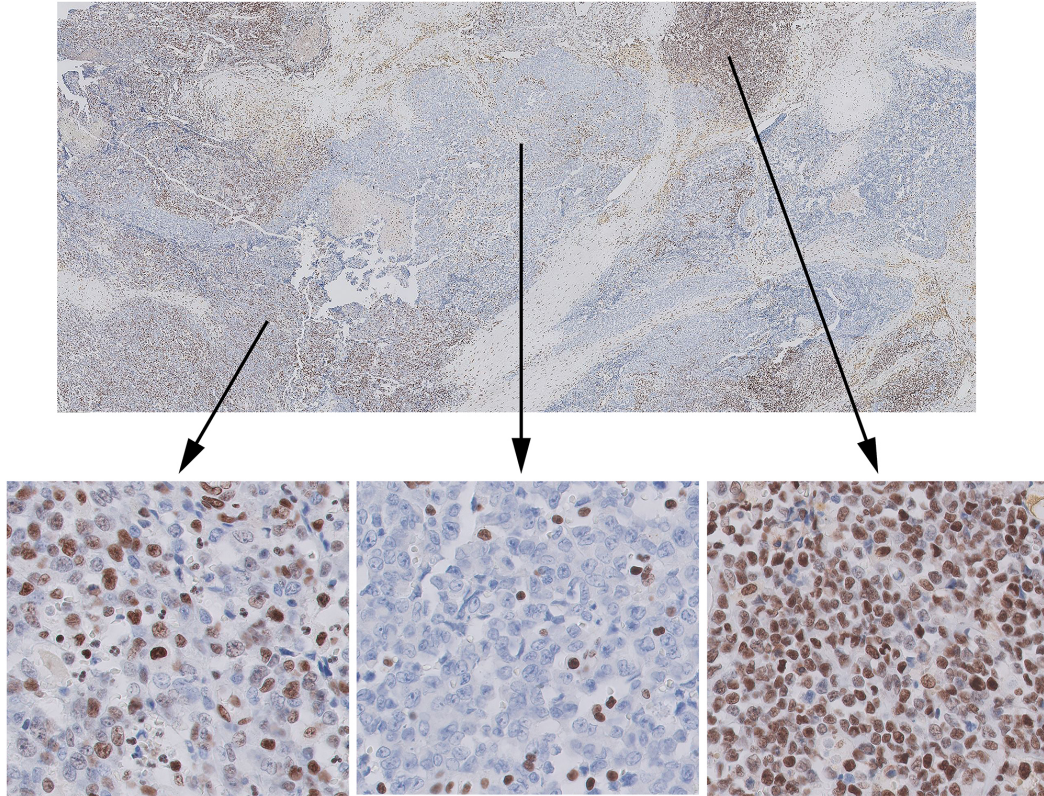


Figure 21. SMARCA2 (BRM) immunostaining in a case of ECRT_{SMARCB1}. A heterogeneous “alternating” expression pattern is seen.

One important question is whether AT/RT, ECRT_{SMARCA4}, ECRT_{SMARCB1} and SCCOHT represent distinct entities, or whether they are identical from a molecular standpoint. To address this, Dr. Bourdeaut’s team performed UMAP (Uniform Manual Approximation and Projection), unsupervised clustering and consensus clustering analyses using RNA-seq data and DNA methylation data (EPIC array) from these tumor types (**Fig.22**).

Within the ATRT group, MYC, SHH and TYR sub-entities were clearly separated both based on RNAseq and on DNA methylation, as expected based on previously published studies (Johann et al., 2016). Both classifications clearly distinguished ATRTs from extracranial tumors (ECRT and SCCOHT), with the exception of ATRT-MYC, which were more closely related to ECRT.

Among extracranial tumors, DNA-methylation and RNA-seq data clearly distinguished the SCCOHT group from the ECRT_{SMARCB1} group. Conversely, ECRT_{SMARCA4} samples showed features somewhat intermediate between SCCOHT and ECRT_{SMARCB1}.

Consensus clustering showed that based on DNA methylation profiles, ECRT_{SMARCA4} (n=9) were more closely related to SCCOHT than to ECRT_{SMARCB1}. Conversely, based on RNA-seq data, ECRT_{SMARCA4} (n=4) appeared to be more closely related to ECRT_{SMARCB1} than to SCCOHT (with the caveat that the ECRT_{SMARCA4} sample size for the transcriptome-based analysis was small).

In addition, overall DNA methylation levels at promoters were highest in ATRT_{TYR}, followed in decreasing order by ATRT_{SHH}, ATRT_{MYC}, ECRT_{SMARCB1}, ECRT_{SMARCA4} and, lastly, by SCCOHT, which showed the lowest overall methylation levels.

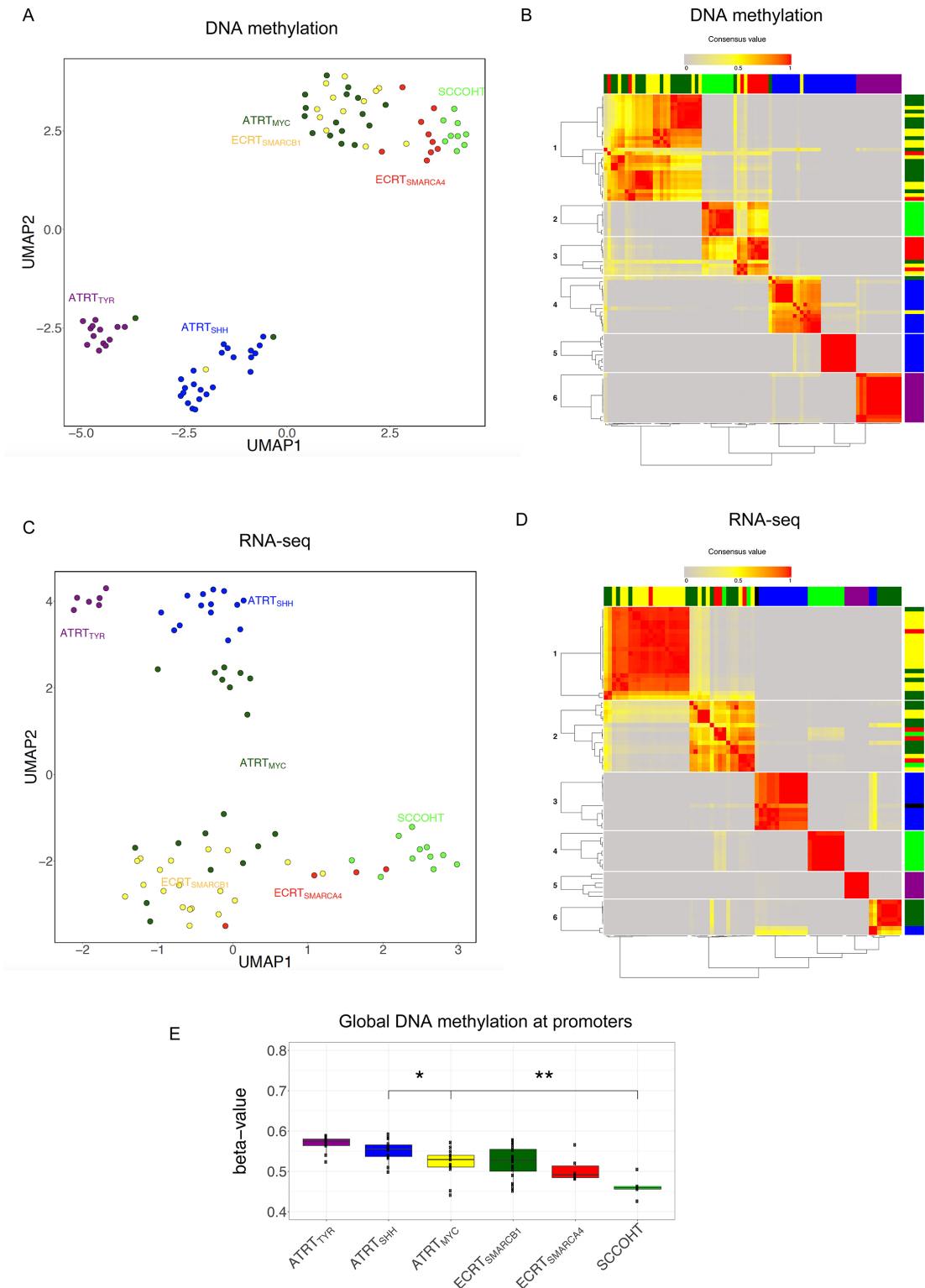


Figure 22. Molecular relationships between AT/RT, ECRT_{SMARCA4}, ECRT_{SMARCB1} and SCCOHT. DNA methylation array (EPIC): A- UMAP analysis, B- consensus clustering. Transcriptomic (RNA-seq) data: C-UMAP analysis, D- consensus clustering. E: global genomic DNA methylation levels. Analysis and figures credits: Mamy Andrianteranagna, Institut Curie.

3. Discussion

3.1 The genomic landscape of SCCOHT.

In keeping with previous studies, our findings in an independent cohort of SCCOHT confirm that these aggressive tumors show remarkable genomic stability, with a diploid DNA content (Eichhorn et al., 1992; Forster et al., 1997; Gamwell et al., 2013). In addition, we found SCCOHT to have a low mutation load (mean, 5.43 mut/Mb). Nevertheless, the numbers are a bit higher than what we may have expected (accordingly, the mean TMB in the series by Lin et al. was <1 mut/Mb) (Lin et al., 2017). These discrepancies are due to a difference in the methods used for TMB calculation across studies, and the lack of universal consensus on how WES-based TMB should be calculated. The low TMB in SCCOHT is in keeping with what has been described in MRT. Indeed, MRT show one of the lowest TMBs among all cancer types (Lawrence et al., 2013). SCCOHT from our cohort also lacked mutations in other genes most commonly altered across various cancer types.

Collectively, such genomic profiles are surprising for a highly lethal malignancy, since “genome instability and mutations” are considered to be some of the main “hallmarks of cancer” (Hanahan & Weinberg, 2011). They support the hypothesis that SCCOHT belong to a very special group of tumors, which are driven by epigenetic deregulation and not by cumulative genomic alterations (except for the initiating mutations in *SMARCA4*, as outlined hereafter).

In line with previous studies, we identified *SMARCA4* mutations in all but one SCCOHT (5/6) from our series. All of these mutations were present at a high allele frequency (VAF 0.73-0.98), in line with an early oncogenic event. Three of these alterations were truncating mutations (one frameshift mutation, and one nonsense mutation that was identical in two different patients); one was a splice site mutation; and one was a missense mutation, which is an uncommon event. Indeed, in a recent review article, Lu et al. summarized the 96 unique *SMARCA4* pathogenic variants reported in SCCOHT to date (including the 6 cases presented

herein): 36.4% were frameshift mutations, 32.2% were nonsense/stop mutations, 20.3% were splice site mutations, 5.9% were missense mutations, and 5.1% were in-frame deletions/insertions (Lu et al., 2019). The missense mutation p.Lys1081Glu in one of our SCCOHT cases was predicted as deleterious by bioinformatics tools, it was homozygous (VAF=0.76) and located near another missense mutation site (p.Gly1080Asp) previously reported in a case of SCCOHT (Witkowski et al., 2014).

Interestingly, our combined WES and CGH-array data show that loss of heterozygosity (LOH) at the 19p13.3 locus, which contains *SMARCA4*, is not associated with a copy number loss (deletion), but instead represents a copy number neutral LOH event (CN-LOH). CN-LOH can account for inactivation of tumor suppressor genes and likely implicates the loss of the normal allele and duplication of the mutated copy (O'Keefe et al., 2010). 19p LOH has previously been detected in SCCOHT by WES (Witkowski et al., 2014), and our WES and aCGH results provide additional evidence for a copy-neutral nature of this event. Of note, telomeric CN-LOH has been linked to meiotic errors occurring during cross-over (O'Keefe et al., 2010), which could be in line with the postulated germ cell origin of SCCOHT tumors (Kupryjanczyk et al., 2013; McCluggage et al., 2017), although further studies are needed to support this hypothesis. 19p CN-LOH associated with inactivating *SMARCA4* mutations have also been reported in non-small cell lung cancer (Medina et al., 2004).

One case (IGR-03) from our series exhibited concomitant inactivating mutations in *ARID1A* and *ARID1B*, two paralog DNA-binding subunits of SWI/SNF, but did not show *SMARCA4* mutations. This tumor was also the only case in which 19p CN-LOH was not present. Concomitant *ARID1A/B* alterations are seen in ~25% of dedifferentiated endometrial and ovarian carcinomas (Coatham et al., 2016). While case IGR-03 could illustrate the challenges of differential diagnosis between SCCOHT and dedifferentiated ovarian carcinoma, another possibility is the existence of a molecular and morphologic overlap between those two

entities, both of which are characterized by a poorly differentiated, aggressive tumor and a critically deregulated SWI/SNF complex.

Of note, one case (IGR-01) showed a p.Arg635* stop gain in the *SMARCA1* gene in addition to a deleterious *SMARCA4* mutation. *SMARCA1* encodes the ATP-ase of another chromatin remodeling complex, ISWI, and is located on the X chromosome, suggesting that this alteration, which was seen at an allele frequency of 0.37, could potentially have a deleterious impact if the second X chromosome carrying the wild type allele was silenced through lyonization, although this would need to be confirmed.

As a complement to the genomic findings, we also show for the first time that SCCOHT are not characterized by SOX2 protein overexpression, contrarily to another aggressive *SMARCA4/SMARCA2* double-negative malignancy: *SMARCA4*-deficient thoracic sarcoma (*SMARCA4*-DTS). This emphasizes the existence of biological differences between *SMARCA4*-DTS and SCCOHT. In addition to SOX2 overexpression, *SMARCA4*-DTS also show greater genomic instability, and tend to occur in older patients with a male predominance and a history of smoking, all of which further suggest that they represent a distinct entity at the clinical and molecular level (Le Loarer et al., 2015).

In addition to *SMARCA4* loss-of-function alterations, a few variants were seen in genes other than *SMARCA4*, all localized in the 19p13.2-3 locus and subject to the CN-LOH event, for which a functional impact could not be ruled out. In particular, the *PLK5* p.G223V variant, predicted as potentially damaging by the Polyphen-2 classifier, was found in 3/33 SCCOHT samples of the extended cohort. The protein kinase domain of *PLK5* is truncated in humans compared to mice, but the residual protein containing the polo-box binding domain may act as a stress inducible tumor suppressor regulating G1 arrest (Andrysiak et al., 2010). Nevertheless, the relevance of these variants remains to be validated functionally.

3.2 Loss of SMARCA4 protein expression in SCCOHT.

Following the discovery of recurrent *SMARCA4* mutations in SCCOHT, IHC for SMARCA4 has been established as a diagnostic tool in the pathology workup of SCCOHT. In a cohort of 44 SCCOHT and 94 rare ovarian neoplasms that represent potential morphologic mimics of SCCOHT, loss of SMARCA4 expression by IHC had a sensitivity and a specificity of 88% and 94%, respectively, for the diagnosis of SCCOHT (Genestie et al., 2020). Despite these results, some non-SCCOHT gynecologic tumors may occasionally show loss of SMARCA4 expression (Karnezis, Wang, et al., 2016). However, dual loss of SMARCA4 and SMARCA2 was reported to be an exclusive feature of SCCOHT among ovarian tumors, and was only seen in exceedingly rare cases of uterine corpus tumors (one dedifferentiated carcinoma and two high-grade endometrial stromal sarcomas) (Karnezis, Wang, et al., 2016).

The reported (Genestie et al., 2020) loss of SMARCA4 expression in SCCOHT cases with homozygous truncating mutations from our cohort (p.Asn774 frameshift and p.Arg1077*) is in keeping with previously published studies. Likewise, retained expression of SMARCA4 in the case IGR-03 from our cohort is not surprising, given the absence of a *SMARCA4* mutation in this case (which, instead, presented *ARID1A* and *ARID1B* mutations). The fact that SMARCA4 expression was reported to be negative by IHC in the case IGR-02 (missense p.Lys1081Glu mutation) is a bit more surprising, as missense mutations are generally not expected to result in a complete loss of the protein product, although one cannot exclude the possibility that such mutations could negatively affect protein stability, or even mRNA stability (Nguyen et al., 2011). One limitation of these results is that the IHC protocol used in this study makes it difficult to compare to other published cohorts (the anti-BRG1 polyclonal antibody used, Santa Cruz sc-10768, has been discontinued, and a manual IHC staining technique was employed).

In the literature, one case of SCCOHT and one case of AT/RT with missense *SMARCA4* mutations and with available IHC results have been reported, and both showed retained SMARCA4 expression by IHC. The SCCOHT case harbored a homozygous missense

p.Gly1080Asp *SMARCA4* mutation (Witkowski et al., 2014) (**Table 6**), and the AT/RT case harbored a homozygous p.Arg1192Cys missense *SMARCA4* mutation (Masliah-Planchon et al., 2016). In addition, several cases of SCCOHT with retained *SMARCA4* expression have been reported (**Table 6**), in which no *SMARCA4* mutation was detected. While this result could be due to technical limitations of the sequencing methods used (alteration present, but not detected), an interesting possibility to explore is whether other SWI/SNF genes could be altered in these rare cases, similarly to what we observed in the case IGR-03 from our cohort (harboring dual *ARID1A/ARID1B* mutations).

References	Number of cases with <i>SMARCA4</i> loss by IHC*	Cases with retained <i>SMARCA4</i> expression by IHC:		
		<i>SMARCA4</i> mutation in cases with retained expression		<i>SMARCA4</i> IHC
Kupryjanczyk et al. 2013	2/2 (100%)	NA		
Witkowski et al. 2014	29/31 (94%)	p.Gly1080Asp (n=1)	homozygous	retained
		c.1420+1G>T (n=1)	heterozygous	weak
Jelinic et al. 2014	4/4 (100%)	NA		
Ramos et al. 2014a	7/9 (78%) ^a	None detected (n=2)		retained
Ramos et al. 2014b	9/10 (90%) ^b	None detected (n=1)		retained
Karanian-Philippe et al. 2015	12/12 (100%) ^c	NA		
Karnezis et al. 2016	28/31 ^{d,e}	Not sequenced		retained
Genestie et al. 2020	39/44 (88%)	None detected (n=5) ^f		retained

Table 6. Data from the literature comparing *SMARCA4* IHC results and mutation status in SCCOHT. *numbers cited for each paper do not include cases already published in earlier studies; ^bnumbers include one SCCOHT cell line; ^cone case showed weak granular staining in rare tumor nuclei; ^dnumbers include two patient-derived xenografts; ^estudy having specifically examined dual *BRG1* and *BRM* loss in SCCOHT; ^fone sample showed biallelic inactivating mutations in *ARID1A* and *ARID1B* (case IGR-03); NA: not applicable.

3.3 Mechanisms of *SMARCA2* silencing in SCCOHT.

Early preclinical studies suggested that *SMARCA4*-mutated tumors (such as non-small cell lung cancers) were critically reliant on the *SMARCA2* paralog (Hoffman et al., 2014). Conversely, SCCOHT do not seem amenable to this synthetic lethality strategy, given the

complete loss of *SMARCA2* expression demonstrated in our series and in previous studies (Karnezis, Wang, et al., 2016). This loss of expression is not explained by mutations in the coding sequence of *SMARCA2*, as shown in earlier studies and in our cohort, and is postulated to be a result of epigenetic and/or post-transcriptional silencing. Results in several *SMARCA2*-deficient cell lines suggested that altered processing or decreased stability of the *SMARCA2* transcript could be involved in loss of *SMARCA2* protein expression (Yamamichi et al., 2005), and it was shown that *SMARCA2* can be re-expressed in *SMARCA2*-negative cells through various treatments, such as HDAC inhibitors (Glaros et al., 2007; Yamamichi et al., 2005).

Similarly, most *SMARCB1*-mutated MRT cell lines were shown to display loss of *SMARCA2* expression, which can be restored through *SMARCB1* re-expression, or through treatment with the flavonoid Flavopiridol, a cyclin-dependent kinase inhibitor (Kahali et al., 2014). While the mechanisms by which these treatments restore *SMARCA2* expression are not entirely clear, they support the hypothesis of epigenetic or post-transcriptional silencing. Accordingly, in SCCOHT models, it was shown that treating BIN67 and SCCOHT-1 cell lines with the histone deacetylase inhibitor TSA led to re-expression of *SMARCA2* at the mRNA and at the protein level (Karnezis et al., 2016). However, treatment with other epigenetic modulators (the DNMT inhibitor d'AZAC, the EZH2 inhibitor GSK343) did not induce *SMARCA2* expression, which suggests that silencing of *SMARCA2* may implicate histone deacetylation. In addition, re-expression of *SMARCA2* through lentiviral transduction in MRT and in SCCOHT cells suppresses cell growth, further suggesting that the absence of *SMARCA2* expression is an important element in the physiopathology of these tumor types (Kahali, Yu, et al., 2014; Karnezis, Wang, et al., 2016).

It was previously shown that two sites within the promoter of the *SMARCA2* gene (-741 bp and -1321 bp) can harbor insertion polymorphisms of 7 bp and 6 bp, respectively (Liu et al., 2011) (**Fig.23**).

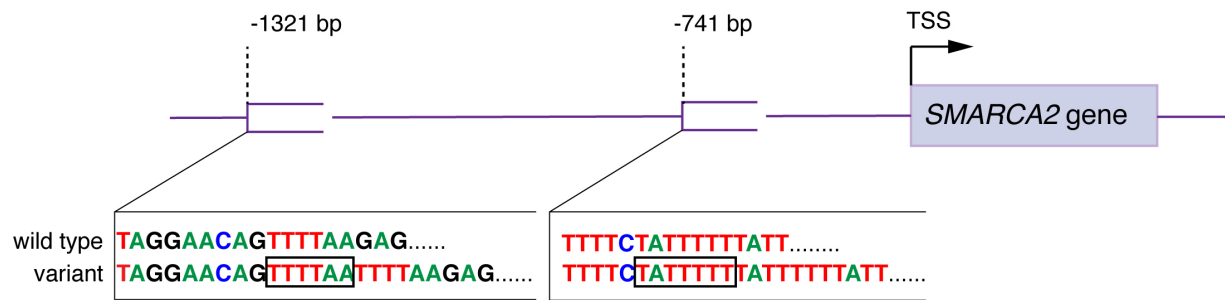


Figure 23. Schematic representation of *SMARCA2* promoter polymorphisms described in (Liu et al., 2011). Black boxes indicate the insertion of 7 bp at the -741 site and 6 bp at the -1321 site, respectively; TSS: transcription start site.

These insertions create sequences homologous to Myocyte Enhancer Factor-2 (MEF2) binding motifs, and it has been hypothesized that MEF2 could help recruit HDAC to these sites; accordingly, knock-down of MEF2 or HDAC9 in MRT cells restores *SMARCA2* expression (Kahali, Gramling, et al., 2014). The presence of homozygous polymorphisms at both sites was found to correlate with *SMARCA2* loss in human lung cancer cell lines and tumors (Liu et al., 2011), although in MRT, the results were less clear, and only the -1321 bp polymorphism appeared to be statistically linked to *SMARCA2* silencing (Kahali et al., 2014). Interestingly, the germline presence of homozygous polymorphisms at both sites in humans is associated with a higher risk of lung cancer (adjusted OR 2.19) (Liu et al., 2011) and of head and neck squamous cell carcinoma (adjusted OR 2.23) (J. R. Wang et al., 2013), and with worse overall survival in patients with non-small cell lung cancer (Liu et al., 2017).

In an effort to explore the mechanism of *SMARCA2* silencing in SCCOHT, we report for the first time that homozygous polymorphisms at these two *SMARCA2* promoter sites do not seem to be a recurrent event in SCCOHT. Nevertheless, it is intriguing that none of the SCCOHT tumors in our study showed a wild-type promoter status for either of these sites, when the known frequency of wild-type genotypes for each of these two sites in the Caucasian population is reported to be around 30% (Liu et al., 2011). Thus, further explorations are

warranted to elucidate whether heterozygous polymorphisms can contribute to *SMARCA2* silencing. In particular, in MRT cell lines, the presence of heterozygous polymorphisms was sufficient to result in increased binding of the epigenetic silencers HDAC9 and MEF2D (Kahali, Yu, et al., 2014).

3.4 SMARCA4 and SMARCA2 expression patterns in extra-cranial MRTs and relationships between tumors from the “rhabdoid tumor” spectrum.

While most MRTs exhibit loss-of-function *SMARCB1* alterations, *SMARCA4*-mutated extra-cranial MRTs (ECRT_{SMARCA4}) are exceedingly rare (Chun et al., 2016; Fruhwald et al., 2006; Saunders et al., 2020). As part of work by Dr Franck Bourdeaut’s team, we report 5 cases of ECRT_{SMARCA4} with available *SMARCA4* mutational status and IHC results for SWI/SNF proteins.

First, while we confirmed abolished *SMARCA4* expression by IHC in all five cases of ECRT_{SMARCA4}, we identified two cases with persistent faint nuclear staining for *SMARCA4* (both were renal tumors). Although these observations remain anecdotal and need to be confirmed in additional cases, they suggest that *SMARCA4* IHC may not always follow an “all or nothing” pattern as initially proposed; instead, a third possible staining pattern appears possible, suggesting that some genetic alterations (here: a splice site mutation and a truncating alteration with no second hit detected) may be accompanied by low residual levels of the protein. These IHC results are unlikely to be due to a technical error. The antibody used for this IHC (Abcam, EPR3912) is a monoclonal, KO validated antibody, directed against an immunogen corresponding to human *SMARCA4* aa 250-350, previously used in pathology research (Dagogo-Jack et al., 2020; D. I. Lin et al., 2019) and used in the diagnostic setting. All stainings were performed in parallel on an automated immunostainer using an established IHC protocol, and slides were evaluated jointly with a pathologist specialized in pediatric pathology (Dr. Paul Fréneaux). We propose that this novel staining pattern should be interpreted as aberrant, and

not as retained (normal) expression, given that it is much weaker than in the surrounding benign cells, that the presence of underlying *SMARCA4* mutations in those cases was confirmed by next generation sequencing, and that clinical and pathology features of these cases were consistent with MRT. Of note, the initial SCCOHT series reported by Witkowski et al. in 2014 comprised one case of SCCOHT with weak residual *SMARCA4* expression (Witkowski et al., 2014), and a case of a renal ECRT_{SMARCA4} with “markedly reduced” *SMARCA4* expression and retained *SMARCB1* expression has also been reported (Yoshida et al., 2017).

In addition, we show that abolished *SMARCA4* expression in ECRT_{SMARCA4} is accompanied by a complete loss of expression of the *SMARCA2* paralog, which was seen in all 5 cases, consistently to what is seen in SCCOHT and *SMARCA4*-DTS. Conversely, in ECRT_{SMARCB1}, *SMARCA2* staining followed an alternating, heterogeneous pattern in the 5 cases tested. In one previous study, *SMARCA2* staining was heterogeneous in 4/10 tested cases of *SMARCB1*-negative MRTs (ECRT and AT/RT), and completely lost in the remaining 6 cases (Yoshida et al., 2017). While our results could be due to sampling bias (because only 5 cases of ECRT_{SMARCB1} were tested), the existence of a heterogeneous *SMARCA2* expression pattern in a subset of MRTs is interesting to explore further, as it could suggest dynamic changes in SWI/SNF composition throughout the tumor, possibly through epigenetic subclonal heterogeneity. Of note, a recent study reported inter-tumor heterogeneity of expression of several other SWI/SNF subunits (*ACTL6A*, *ACTL6B* and *PBRM1*) in MRTs (Panwalkar et al., 2020) and demonstrated that expression of these proteins correlated with polyphenotypic differentiation along different lineages, which is a known feature of MRTs, and with immune environment composition. Thus, further studies are warranted to verify whether heterogeneity of *SMARCA2* expression in ECRT_{SMARCB1} also relates to lineage differentiation, and analyses at a single-cell level could help better understand the molecular differences between *SMARCA2*-positive and *SMARCA2*-negative ECRT_{SMARCB1} subpopulations.

An important question is whether SCCOHT and MRT represent strictly identical entities which differ only in their clinical presentation, or whether some molecular differences exist between these tumor types. Work by Dr. Bourdeaut's team has shown that based on transcriptomic and DNA methylation data, SCCOHT and ECRT_{SMARCB1} represent distinct tumor groups, while ECRT_{SMARCA4} show somewhat intermediate features between the two groups. Thus, although AT/RT, ECRT_{SMARCB1}, ECRT_{SMARCA4} and SCCOHT clearly represent a family of tumors with similar genomic and pathologic characteristics, the existence of molecular differences between those entities supports their continued separate classification.

3.5 The phenotype of SCCOHT and putative therapeutic targets: insights from the transcriptomic analysis.

Our differential expression analysis identified genes that are overexpressed in SCCOHT as compared to benign ovarian tissue. We observed overexpression of genes related to various types of cell differentiation in SCCOHT, including neural genes (e.g. *NCAM2*, *NTS*, *ATCAY*, *CBLN2*), genes related to osteogenic differentiation (e.g. *BMP7*) and embryonic genes (e.g. *CLDN6*). This is consistent with the expression of multi-lineage markers in SCCOHT by IHC, which has been known for a long time and which has accounted for difficulties in classifying these tumors from a pathology perspective. It is also in keeping with observations in MRT cells, which show expression of markers from different lineages, including stem cell, neuronal and mesenchymal markers (Deisch, Raisanen, & Rakheja, 2011; Fahiminiya et al., 2016; Muscat et al., 2016; Okuno et al., 2010), and are capable of multi-lineage differentiation under specific conditions (Muscat et al., 2016).

In 2016, integrated molecular analysis showed that AT/RT could be subdivided into three distinct subgroups: ATRT-TYR (overexpressing melanosomal genes, including the enzyme tyrosinase), ATRT-SHH (overexpressing genes from the SHH pathway) and ATRT-MYC (overexpressing the MYC and HOX cluster)(Johann et al., 2016). Molecular heterogeneity was

also demonstrated in extra-cranial MRT (Chun et al., 2016), and it was shown that extra-cranial MRT more closely related to ATRT-MYC than to the other subgroups of AT/RT (Chun et al., 2019). In future work, it would be interesting to perform multi-omics analyses in larger cohorts of SCCOHT and to assess whether SCCOHT also display some degree of transcriptomic heterogeneity and can be subdivided into molecular groups.

We also observed that genes encoding other SWI/SNF subunits are still expressed in SCCOHT, and confirmed the expression of some of these subunits at the protein level. Previous studies demonstrated that SWI/SNF complexes can still assemble in the absence of function SMARCA4 and SMARCA2 (Wilson et al., 2014), and residual complexes were recently evidenced in SCCOHT cells (Pan et al., 2019). Such aberrant residual complexes are of particular interest, as they could offer a rationale for therapeutic targeting, similarly to what has been proposed in MRT with the non-canonical BAF complex: indeed, synovial sarcoma and MRT cell lines showed exquisite sensitivity to the loss of BRD9 or GLTSCR1 subunits, which are specific of non-canonical BAF (Michel et al., 2018; X. Wang et al., 2019).

The transcriptomic analysis also described candidate putative therapeutic targets. In particular, the following proteins could warrant further investigation:

- Melanoma-associated Antigens: we found several genes from this group (*MAGEA4*, *MAGEA9*, *MAGEA2*, *MAGEA3*) to be overexpressed in SCCOHT. These proteins belong to the “cancer/testis antigen” family, which is of interest for therapeutic targeting given their specific expression patterns (restricted to male germs cells and to cancer cells) and high immunogenicity *in vivo*. MAGE-A4 is currently being investigated in phase I trials as a target for genetically modified autologous T cells in advanced or metastatic solid tumors expressing MAGE-A4. Given the remarkable genomic stability and low mutation burden of SCCOHT, Chimeric Antigen Receptor T cell (CAR-T cell) approaches could be of particular interest in these tumors, because one could hypothesize that they represent a relatively homogenous

disease (at least at the genomic level). As a future step, it could be interesting to validate overexpression of MAGE-A4 in SCCOHT at the protein level by IHC.

- Aurora B kinase (AURKB) is a member of the Aurora family of serine/threonine kinases involved in several steps of cell division, including centromere duplication, mitotic spindle formation, chromosome segregation, mitotic checkpoint and cytokinesis; it is overexpressed in many cancer types, and therefore by no means specific of SCCOHT (Carvajal, Tse, & Schwartz, 2006; Tang et al., 2017). Pharmaceutical inhibitors against multiple Aurora kinases, as well as selective inhibitors of Aurora B kinase, have been developed, including VX-680/MK0457 (Merck) and barasertib/AZD1152 (Astra Zeneca). While upregulation of the Aurora A kinase (AURKA) gene was not significant in our study, it is worth noting that the Aurora A kinase inhibitor alisertib/MLN8237 is currently being tested in a phase II study in children and young adults with AT/RT or extra-renal MRT.

- Claudin-6 (CLDN6) is a tight junction protein mainly found in embryonic stem cells, embryonic epithelia and germ cell tumors (L. Wang et al., 2012), and its overexpression in other cancer types (including gastric cancer, ovarian serous carcinoma, basal-like breast carcinoma) has also been described, and was correlated with worse prognosis in some studies (Heerma van Voss et al., 2014; Kohmoto et al., 2020; L. Wang et al., 2013). A monoclonal antibody against Claudin-6, IMAB027 (Ganymed Pharmaceuticals) has been developed and tested in a phase I trial in adult patients with advanced “ovarian cancer of any histology” expressing Claudin-6 (NCT02054351)(Ugur Sahin, 2015), but the results of this study have not been published to date. Our IHC results (**Fig.17**) suggest that Claudin-6 is expressed at the protein level in a subset of SCCOHT cases. Depending on future developments regarding the anti-Claudin-6 antibody, this target could warrant further investigation in a subset of SCCOHT patients identified by IHC.

- HER4 (ERBB4) is a receptor tyrosine kinase (RTK), which binds the ligand heuregulin (HRG) and which can form homodimers or heterodimers with HER2. The role of HER4 in cancer

remains controversial, as higher HER4 expression was associated with better prognosis in breast and bladder cancer (Memon et al., 2004; J. Wang et al., 2016). While there is currently no compelling evidence that HER4 could represent a therapeutic target in SCCOHT, it is interesting to note that an siRNA RTK dependency screen performed in a SCCOHT cell line identified signaling by HER2 and signaling by HER4 among top enriched pathways (Lang et al., 2018). In addition, the authors demonstrated promising sensitivity of SCCOHT to the multikinase inhibitor ponatinib (which targets BCR-ABL, EGFR, FGFR, PDGFR, VEGFR, SRC, RET, KIT and FLT1 kinases) *in vitro* and in xenograft models.

A few limitations of this gene expression analysis need to be mentioned. First, the benign ovarian tissue used in the differential expression analysis is not the ideal control, as it represents a mixture of cell types found in the ovary, with likely overrepresentation of the ovarian stromal cells. The “cell of origin” of SCCOHT, which would represent the ideal control for this type of analysis, remains unknown as of yet. Similarly, identifying the “cell of origin” in MRT is currently the subject of intense scientific investigations. Second, the analysis had to take into account an important batch effect, as we were comparing two datasets generated through different platforms. We did, however, take care to only use GTEx data from pre-menopausal subjects, to better match the demographics of SCCOHT patients from our cohort and thus minimize bias related to age. Lastly, an important next step will be to validate the expression of selected markers in SCCOHT at the protein level.

3.6 Future directions: exploring novel therapeutic strategies in SCCOHT

Several treatment strategies are currently under investigation for SCCOHT. In addition, despite the existence of some molecular differences between SCCOHT and MRT as shown by Dr Bourdeaut’s team, which support a continued separate classification, some therapeutic strategies tested in the context MRT could potentially be investigated in SCCOHT and *vice versa*, given their shared morphologic, genomic and molecular characteristics.

a) Epigenetic modulation

Our *in vitro* results suggest that SCCOHT are exquisitely sensitive to DNMT inhibitors and HDAC inhibitors; however, these experiments were performed over a very short period of time (72h). In human MRT cell lines and xenograft models, treatment with HDACi (LBH589/panobinostat) over a longer time period (3 weeks) and at low doses (“differentiating” dose rather than “cytotoxic” dose), promoted cell senescence and cell cycle arrest, and resulted in multi-lineage differentiation, including conspicuous intra-tumor ossification (Muscat et al., 2016). In further work, it would be important to assess how SCCOHT react to similar treatment.

SWI/SNF-deficient cancer cells were shown to be highly sensitive to the inhibition of the histone methyltransferase EZH2, which is part of the PRC complex (Kim et al., 2015). SCCOHT cells were also very sensitive to EZH2 inhibition *in vitro* and in xenograft models (Chan-Penebre et al., 2017; Y. Wang et al., 2017). In addition, some SCCOHT patients in clinical studies, including the patient described herein, showed clinically meaningful response to EZH2 inhibitors (Jones et al., 2018; Italiano et al., 2018). However, in the phase II study (NCT02601950) investigating the EZH2 inhibitor tazemetostat in adult patients with solid tumors, stage 2 futility was not passed in the arm enrolling SCCOHT patients (rhabdoid tumor arm) (Jones et al., 2018); this further suggests possible molecular heterogeneity of SCCOHT, and the need of determining predictive factors of response to such epigenetic treatments.

b) Immune checkpoint inhibitors (ICI)

In other cancer types, good response to immunotherapies has classically been linked to the abundance of tumor neoantigens (neoepitopes), which is particularly seen in tumors with a high tumor mutational burden (TMB). This has led to the remarkable FDA approval in 2017 for the use of pembrolizumab in tumors with microsatellite instability (which display high TMBs) in a histology-agnostic and anatomic site-agnostic manner. Conversely, both MRTs and SCCOHT, as shown in our cohort and in other studies, display remarkably low tumor mutation burdens (Lawrence et al., 2013; D. I. Lin et al., 2017).

Although based on the above rationale, ICI would not be expected to have an effect in SCCOHT or MRT, surprisingly, anecdotal reports exist of ICI having an effect in some patients. In a recent report, several patients with SCCOHT having recurred after surgery were given “standard treatments” (not specified) followed by anti-PD1 immunotherapy (Jelinic et al., 2018). Although these case reports remain anecdotal, they suggested a clinically meaningful effect of ICI in these patients (three patients remained disease-free for at least 1.5 years, and one patient showed sustained partial response for 6 months). Encouraging activity of immune checkpoint inhibitors (partial responses) has also been reported in several pediatric patients with MRT as part of clinical trials (Bourdeaut F., 2017; Geoerger et al., 2020). In a study by Leruste et al., PD-1 blockade was effective in decreasing tumor growth *in vivo* in syngeneic MRT graft models, and induced subsequent immunity against tumor grafts (Leruste et al., 2019).

Given these surprising results, an important step will be to try to understand why MRT and SCCOHT may respond to checkpoint inhibitor treatments, despite their very low mutation load. A subset of these tumor appears indeed to be immunogenic. In a study by Jelinic et al., most SCCOHT harbored a significant T-cell infiltrate (Jelinic et al., 2018). In MRTs, Leruste et al. found prominent T cell and myeloid cell infiltrates, associated with unexpectedly high transcriptome-based cytolytic scores and evidence for clonal expansion of CD8+ T cells (Leruste et al., 2019). Some degree of PD-L1 expression in tumor cells (range, 9-287 cells per high power field) was found in most SCCOHT from the study by Jelinic et al. (Jelinic et al., 2018), and has also been demonstrated in MRT (Abro et al., 2019). In our cohort, we did not observe significant overexpression of the *CD274* gene (which encodes PD-L1) at the transcript level in SCCOHT as compared to benign ovarian tissue, but this should be completed by an evaluation of PD-L1 expression at the protein level by IHC.

Intriguingly, work by Leruste et al. revealed that aberrant expression of endogenous retroviral elements (ERV), caused by a dysfunctional SWI/SNF complex, is in fact an important

factor underlying immunogenicity in MRTs. Thus, it would be interesting to assess whether ERVs are similarly expressed in SCCOHT.

A currently active phase II study NCT044165568 investigates the safety and effectiveness of the nivolumab + ipilimumab combination in children and young adults with SMARCB1-deficient cancers, including MRT. A French AcSé phase II clinical trial (Institut Gustave Roussy, NCT03012620), allows secured access to the anti-PD1 antibody pembrolizumab for patients with different tumor types for which the drug is not approved to date, including rare ovarian cancers (Cohort 2).

The utility of combination therapies associating immune checkpoint inhibitors with other drugs should also be considered in SCCOHT. For example, the CDK4/6 inhibitor abemaciclib, was shown to act in a synthetically lethal manner with SMARCA4 loss in lung cancer and has shown promising activity in pre-clinical models of SCCOHT, in a study conducted in collaboration with Dr Alexandra Leary's team (Xue et al., 2019). In addition, a recent case report showed marked long-term response to combined treatment with the CDK4/6 inhibitor abemaciclib and the anti-PD1 antibody nivolumab in a SCCOHT patient (Lee et al., 2020). Likewise, the lysine-specific histone demethylase 1 (LSD1/KDM1A) has been postulated as a therapeutic target in SWI/SNF-deficient tumors, because this epigenetic regulator is highly expressed in these tumors and some SWI/SNF-deficient cancer cell lines have shown sensitivity to LSD1 inhibition (Soldi et al., 2020); however, the LSD1 inhibitor SP2577/seclidemstad was also shown to stimulate T-cell infiltration and PD-L1 expression in SCCOHT models (Soldi et al., 2020). Other epidrugs, such as DNMTi, can also complement immune checkpoint inhibition, by inducing terminal differentiation and increasing expression of major histocompatibility complex (MHC) molecules in the process (Velcheti et al., 2019).

c) Genetically engineered autologous T cells or antibody-drug conjugates

Cell surface markers overexpressed in SCCOHT that would not be expressed in adult tissues could contribute to the immunogenicity of SCCOHT if they are not recognized as "self"

by the adult immune system. However, while cancer-testis antigens could represent one such candidate, Leruste et al. showed that in MRT, expression of cancer-testis antigens did not correlate with the cytolytic score, suggesting that these proteins may not be the main driver of immunogenicity in MRT (Leruste et al., 2019). However, such markers could still represent candidates for genetically engineered autologous T cells or antibody-drug conjugate approaches. Such approaches are being studied in MRT and AT/RT; for example, the B7-H3 (*CD276*) is a protein from the immunoglobulin superfamily, which is highly expressed in the prenatal, but not the postnatal, brain. CAR-T cells directed against this antigen have shown promising activity in a xenograft model of AT/RT, especially when administered into brain ventricles or intratumorally (Theruvath et al., 2020), and is currently investigated in a phase I pediatric trial in solid tumors, including MRTs (NCT04483778). *CD276* was not significantly overexpressed in SCCOHT by RNA-seq in our analysis, although it has not been queried at the protein level.

In summary, while effective treatment options for SCCOHT are still lacking, recent advances in the understanding of the biology of these tumors have suggested therapeutic strategies that warrant further investigation. Given that SCCOHT-specific research is hampered by the rarity of this disease, international collaborations and establishing a large clinical database in concert with the SCCOHT Consortium (founded in July 2018) should be particularly encouraged.

4. Methods

Patients and samples

Fresh-frozen tumor samples from 8 patients with SCCOHT were identified from the tumor banks of Institut Gustave Roussy, Cochin University Hospital, Grenoble University Hospital, Longjumeau University Hospital and Hôpital de la Croix Rousse. Central review for histological diagnosis was conducted by an expert pathologist. Matched blood samples were

available for 6 patients. All patients provided written informed consent allowing use of their tumor and non-tumor tissues for research. Approval from the hospital's institutional review board was obtained for the study and funding was obtained via an educational grant awarded by the Foundation Gustave Roussy (local IRB approval RT12014). In addition, a further 33 formalin-fixed paraffin-embedded (FFPE) SCCOHT samples were available for validation studies. Tumors were obtained with patient consent and all data anonymized.

DNA extraction

DNA was extracted from fresh-frozen tumors and matched blood using the AllPrep DNA Mini Kit (Qiagen, Valencia, California, USA) according to the manufacturer's instructions, and quantified using Qubit (Thermo Fisher Scientific, Waltham, MA, USA). DNA integrity was measured using an Agilent BioAnalyzer (Agilent, Santa Clara, CA, USA).

Whole Exome Sequencing (WES)

Exome capture and library preparation were performed using the SureSelect HumanAll Exome v5 and SureSelectXT kits, respectively (Illumina, Agilent Technologies, CA, USA). Sequencing was done on matched tumor and normal samples using HiSeq2000 (Illumina, San Diego, CA, USA) in paired-end mode with a mean target depth of 100X. Reads were mapped using BWA-MEM (V0.7.5a-r405) against reference genome hg19. Analysis of coverage was done using GATK (2.7.4-g6f46d11) DepthOfCoverage. Local realignment was performed around indels using GATK RealignerTargetCreator and GATK IndelRealigner.

Variants were called with VarScan2, using hg19 as the reference genome and requiring a minimum tumor read depth of 6, a minimum somatic read depth of 8 and a minimum tumor allelic frequency of 0.10. Results were then annotated using SnpEff (4.3t) [13] and SnpSift (4.3t) with dbSNP (v150_hg19) (<http://www.ncbi.nlm.nih.gov/SNP/>) and dbNSFP (v2.9).

Tumor mutation burden (TMB) was calculated based on the number of non-synonymous, somatic-only mutations (single nucleotide variants and small insertions/deletions) with a somatic p-value threshold at <0.05 per mega-base in coding regions considered as

having sufficient coverage (6x in tumors and 8x in matched normal samples) by the variant caller.

Oligonucleotide CGH microarrays

DNA was labeled and hybridized and CGH microarray analysis was performed as detailed in supplemental methods. Resulting $\log_2(\text{ratio})$ values were segmented using the CBS algorithm implementation from the DNACopy package for R. Aberration status calling was automatically performed for each profile according to its internal noise (absolute variation of $\log_2(\text{ratio})$ values across consecutive probes on the genome). All genomic coordinates were established on the UCSC Homo sapiens genome build hg19.

SMARCA2 promoter Sanger sequencing

Sanger sequencing of *SMARCA2* promoter polymorphism sites was performed on DNA from 8 fresh-frozen SCCOHT tumor samples and 2 cell lines (BIN67 and SCCOHT-1). The following primers were used for the -741 site: Forward: TTTGGAAGCTTGCAGTCCTT, Reverse: CCGGCTGAACTTTTTCTCC; for the -1321 site: Forward: CCCAGTTGCTCAAATGGAGT, Reverse: AGGTCGGTGTTTGGTGAGAC. After PCR, 10ul from a 50ul reaction were run on a 2% agarose gel to confirm amplification. The remaining PCR reaction was purified using the Qiagen QIAquick PCR Purification Kit, quantified and 10ng together with 25pmol of either the Forward or the Reverse primer were submitted to Genewiz (USA) for Sanger sequencing.

Immunohistochemistry

- IHC on the study cohort of 8 SCCOHT cases:

The protocol and results for the anti-SMARCA4 (BRG1) immunohistochemistry for SCCOHT presented in this study were previously published in (Genestie et al., 2020). Briefly, IHC was performed using the anti-BRG1 (Santa Cruz, sc-10768) and anti-BRM (Abcam, ab15597) antibodies at dilutions of 1/200 and 1/50, respectively. After paraffin removal and hydration, slides were immersed in 10mM citrate buffer pH 6 for 30 min for antigen retrieval, incubated with

primary antibody for one hour at room temperature, washed and incubated with biotinylated secondary antibody for 30 min at room temperature. Streptavidin-biotin amplification (VECTASTAIN Elite ABC Kit) was then performed for 30 minutes, followed by peroxidase/diaminobenzidine substrate chromogenic reaction.

- IHC for SOX2 and Claudin-6:

These IHC was performed using a Bond III automated immunostainer and the Bond Polymer Refine detection system (Leica Microsystems, IL, USA). For both stainings, slides were de-paraffinized and heat-mediated antigen retrieval was performed using the Bond Epitope Retrieval 2 solution at pH9 (H2) for 20 min. The anti-SOX2 antibody clone D6D9 (Cell Signaling Technology) was used at 1/100 dilution. The anti-Claudin 6 antibody (IBL America, catalogue number 18865) was used at 1/100 dilution with overnight incubation.

- IHC in extra-cranial MRTs:

Immunostainings presented in **Fig.19-21** were performed using a Bond III automated immunostainer and the Bond Polymer Refine detection system (Leica Microsystems, IL, USA), with the following antibodies and conditions:

Protein	Provider	Type	Clone	Retrieval	Dilution	Primary Ab incubation
SMARCA4 (BRG1)	Abcam	Rabbit monoclonal	EPR3912	pH9	1/200	60 min
SMARCA2 (BRM)	Cell Signalling	Rabbit monoclonal	D9E8B	pH9	1/1000	45 min
SMARCB1 (INI-1)	BD Biosciences	Mouse monoclonal	SNF5	pH6	1/50	30 min

RNA sequencing (RNA-seq) and differential expression analysis

RNA-seq was performed on RNA from 8 fresh-frozen SCCOHT tumors, on a HiSeq2000 sequencer, using paired-end 2 x 76bp stranded mode. Differential RNA-seq gene expression analysis between SCCOHT samples and benign ovarian tissue from the GTEx dataset was performed using rank-normalized expression values Cufflinks (2.0.2) (PMID: 22383036) was used to estimate the expression values (FPKMS), and GENCODE v19 (PMID: 22955988) GTF file for annotation. Data for benign ovarian tissue were downloaded from the GTEx portal; only GTEx

samples from patients <42 years old (n=5) were selected. All reads were independently aligned with STAR_2.4.0f1 (PMID: 26334920) for sequence alignment against the human genome sequence build hg19, downloaded via the UCSC genome browser and SAMTOOLS v0.1.19 (PMID: 19505943) for sorting and indexing reads. Cufflinks (2.0.2) was used to estimate the expression values (FPKMS), and GENCODE v19 GTF file for annotation. Since benign ovary tissue samples from GTEx were processed using different library preparation methods, and in order to minimize batch effect, we performed rank normalization of gene expression data (replacing each observation by its fractional rank, i.e. the rank normalized by the total number of genes). For differential expression analysis, we calculated log2foldchange from rank normalized expression values. For statistical significance, we performed Wilcox test with Benjamini and Hochberg adjustment to p-values. R (v3.3.2) and ggplot2 (2.2.1) were used for the statistical analysis.

Western blotting

Cells were lysed in RIPA buffer with protease and phosphatase inhibitors (Thermo Fisher Scientific) and total protein concentration was measured using the DC Protein Assay (Bio-Rad). Protein samples were resolved in SDS-PAGE, transferred onto a nitrocellulose membrane using the iBlot 2 dry blotting system (Thermo Fisher Scientific) and incubated overnight at 4°C with primary antibodies dissolved in 5% Blotting-Grade Blocker (Bio-Rad). The following primary antibodies (provider, clone, dilution) were used: BRG1 (Abcam, EPR3912, 1/1000), BRM (Cell Signaling Technology, D9E8B, 1/1000), BAF47 (Abcam, EPR12014, 1/5000), BAF155 (Abcam, EPR12395, 1/5000), BAF170 (Cell Signaling Technology, D8O9V, 1/10000), BAF45B (Atlas Antibodies, polyclonal, HPA049148, 1/1000), BAF53A (Abcam, EPR7443, 1/2000), BAF250A (Abcam, EPR13501, 1/1000). After 3 washes, the membrane was incubated with secondary antibody conjugated to horseradish peroxidase for 1h at room temperature. After 3 washes, signal was visualized by chemiluminescence using the Luminata Forte substrate (Thermo Fisher Scientific) and images were acquired with the ChemiDoc™ Touch Imaging System (Bio-Rad, Hercules, CA).

Cell culture and cell viability assay

BIN-67 cell line (SCCOHT cell line) was kindly gift from Prof. B. Vanderhyden (Ottawa Hospital Research Institute, Ottawa, CANADA). Others cell lines were obtained from the American Type Culture Collection (ATCC): NCI:H1299 (ATCC® CRL5803™) (human non-small lung cancer cells), NIH:OVCAR3 (ATCC® HTB161™), SKOV3 (ATCC® HTB77™) (human high-grade endometrioid adenocarcinoma of the ovary) and G401 (ATCC® CRL1441™) (malignant rhabdoid tumor). Cells lines were tested and certified as mycoplasma-free. Cells were maintained in Dulbecco's modified Eagle/F12- Dulbecco's modified Eagle's medium for BIN-67 and RPMI-1640 for others (GIBCO/ Life Technologies, Saint Aubin, France). Mediums were supplemented with 10% fetal calf serum (PAN Biotech, Aidenbach, Germany) and 1% penicillin/streptomycin (GIBCO/ Life Technologies, Saint Aubin, France; 100µg/mL) in a 37°C incubator with 5% CO₂ saturation.

Cell viability was determined using MTT assay. In brief, cells were seeded into 96-well plates (1.2×10^4 cells/well, 100 µL working volume per well), treated with drugs at increasing concentrations, incubated for 72 hours at 37°C and collected at 0h, 24h and 72h timepoints. MTT was added to the wells and incubated for 3 hours. Cell media was replaced by DMSO and pipetted to dissolve the formed formazan crystals. Then, optical density of the colored solution was quantified at 560 nm wavelength using a microplate reader (Victor X4 Multilabel Reader: Perkin Elmer, Courtaboeuf, France). Following compounds were tested at different concentrations (1-10-100nM): Trichostatin-A (TSA) (T8552, Sigma); 5'-Azacytidine (5'dAZAC) (Sigma, A2385). Negative control for all the assays was represented by the untreated medium containing vehicle DMSO (0.1%). Viability ratios (using the 0h timepoint as reference) between treated and untreated cells at 72h were compared using a two-way ANOVA test.

Part III. The SWI/SNF complex has tumor-promoting functions in prostate cancer.

1. Background

1.1 Prostate cancer: introduction

Prostate cancer (PCa) is the second most commonly diagnosed malignancy and the fifth cause of cancer-related death in men worldwide, with an estimated 1,28 million new cancer cases (13.5% of all cancers in men) and about 359,000 deaths (6.7% of all cancer-related deaths) in 2018 (Bray et al., 2018; Rawla, 2019). PCa is the third predicted cause of cancer-related death in EU men in 2020, with 78,800 predicted deaths (Carioli et al., 2020); this would represent a -7.1% decrease in the death rate since 2015, attributed to therapeutic advancements. In the United States, it was predicted to represent the most commonly diagnosed cancer in men and the second cause of cancer-related death in 2019, with 174,650 new cases (20% of all cancer cases in men) and 31,620 deaths in (10% of all cancer-related deaths in men) in that year (Siegel, Miller, & Jemal, 2019).

Despite the very high incidence of PCa, the overall mortality rate is relatively low compared to other cancer types, as most PCa patients are effectively treated by local therapies (surgery and/or radiotherapy) or can be followed by active surveillance. Only an estimated 20% of patients diagnosed with localized PCa develop metastatic disease during long-term (>10 year) follow-up, which emphasizes the importance of accurately predicting prognosis in patients with localized PCa to avoid overtreatment (Kamoun et al., 2018). For men who develop metastatic recurrence or present with metastases at initial diagnosis, the mainstay of treatment for metastatic PCa is androgen deprivation therapy (ADT), but resistance to these therapies ultimately develops, with progression to castration-resistant prostate cancer (CRPC). CRPC harbors a “luminal” (usual adenocarcinoma) differentiation (CRPC-Adeno), and still depends on

androgen receptor (AR) signaling, but is able to evade ADT treatment through various molecular alterations – including, but not limited to, AR gene amplifications, activating AR mutations, AR splice variants, mutations in AR coactivators/corepressors, or alternative androgen production (Abida et al., 2019b; Beltran et al., 2016; Robinson et al., 2015). Docetaxel chemotherapy was classically the first line treatment for CRPC, but more recently, more potent AR signaling inhibitors (ARSi) have been developed to target this form of disease. These are exemplified by abiraterone acetate (a CYP17 inhibitor) and enzalutamide (an AR signaling inhibitor), which have led to marked improvement of overall survival in CRPC patients (de Bono et al., 2011; Fizazi et al., 2012; Scher et al., 2012). According to most recent clinical trial data, these novel treatments show ever more benefit the earlier they are initiated; the STAMPEDE and LATITUDE clinical trials demonstrated practice-changing improvement in overall survival for men with newly diagnosed metastatic PCa by adding abiraterone to standard ADT (Fizazi et al., 2017; James et al., 2017). However, even though they are highly effective, resistance to these treatments ultimately develops as well.

1.2 Neuroendocrine prostate cancer.

Neuroendocrine PCa (NEPC) is classically described as having a distinct histomorphology, positive expression of neuroendocrine markers (such as synaptophysin or chromogranin A), and loss of AR expression (Abida et al., 2019b; Epstein et al., 2014; Rickman, Beltran, Demichelis, & Rubin, 2017) (**Fig.24**). NEPC can be present at diagnosis (*de novo* NEPC), but this is a rare event (<1% of cases) (E.J. Small, 2014); in most cases, the neuroendocrine phenotype is seen in the CRPC setting, where transdifferentiation to NEPC represents a form of treatment resistance (Davies et al., 2018). In our recent review of 440 CRPC patients (**Appendix I**), NEPC was encountered in 11% of the biopsies (Abida et al., 2019b).

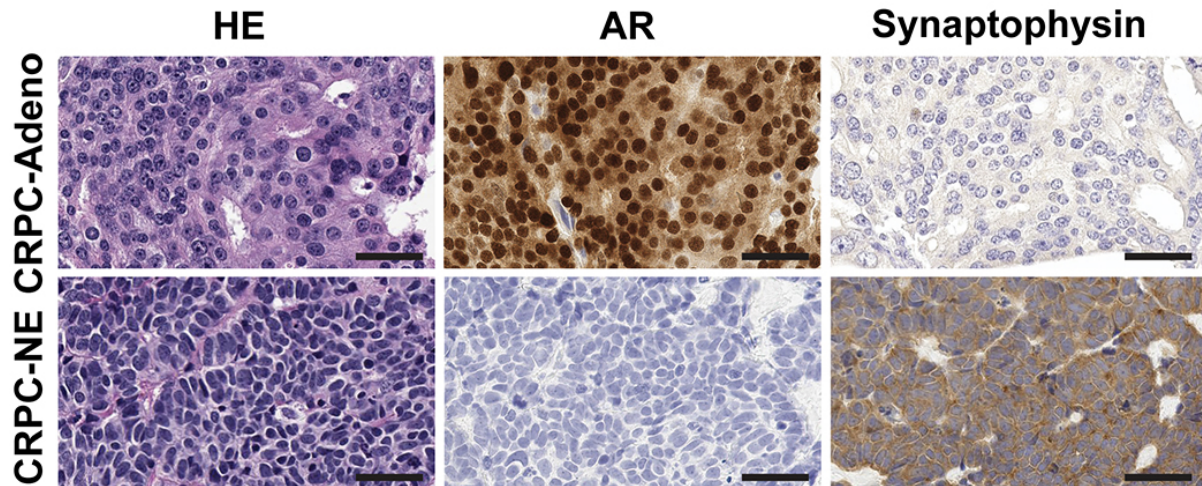


Figure 24. Pathology of castration-resistant prostate cancer: adenocarcinoma (CRPC-Adeno) versus neuroendocrine carcinoma (CRPC-NE). CRPC-Adeno shows typical morphology characterized by abundant cytoplasm, prominent nucleoli and possible gland formation. Conversely, CRPC-NE displays scant cytoplasm, a high nucleus-to-cytoplasm ratio, nuclear molding, fine or “salt and pepper” chromatin and inconspicuous nucleoli. On immunohistochemistry, CRPC-Adeno shows strong AR expression, while CRPC-NE shows loss of AR expression, consistent with AR-signaling indifference, and positivity for neuroendocrine markers such as synaptophysin. Further pathology classification of CRPC-NE into distinct subtypes has been proposed in: (Epstein et al., 2014). HE: hematoxylin eosin, AR: androgen receptor, Scale bar, 50 um.

It is important to define what NEPC actually means. The bona fide example of NEPC, and the term that is still sometimes used interchangeably with NEPC, is “small cell carcinoma”. However, in recent years, the “NEPC” category of tumors has been further subdivided based on morphologic and/or molecular criteria, and new phenotypic entities have been proposed. Epstein et al. outlined six distinct pathology categories of prostatic tumors with neuroendocrine differentiation: usual adenocarcinoma with neuroendocrine differentiation; small cell neuroendocrine carcinoma; large cell neuroendocrine carcinoma (rare); mixed neuroendocrine carcinoma-acinar adenocarcinoma; adenocarcinoma with Paneth cell-like differentiation (rare); and carcinoid tumor (well-differentiated neuroendocrine tumor)(rare) (Epstein et al., 2014). The SU2C/PCF/AACR West Coast team subsequently proposed a new subtype of CRPC, termed Intermediate Atypical Carcinoma (IAC), which was reportedly encountered in up to 24% of

metastatic CRPC biopsies from their cohort and characterized by morphologic and transcriptomic features intermediate between usual adenocarcinoma (CRPC-Adeno) and small cell carcinoma, and genomic features similar to usual adenocarcinoma (Small et al., 2020). However, IAC is not widely accepted as a distinct diagnostic entity for the moment. Another special variant of CRPC is the “double-negative” carcinoma, characterized by simultaneous negativity for AR and for neuroendocrine markers, and driven by activation of FGF and MAPK pathways (Bluemn et al., 2017). Yet another recent study distinguished five subtypes of CRPC based on their transcriptomic profiles: AR-high adenocarcinoma; AR-low adenocarcinoma; small cell neuroendocrine carcinoma; amphicrine carcinoma (which co-expresses AR and neuroendocrine markers); and double-negative carcinoma, which can evolve towards squamous differentiation in some cases (Labrecque et al., 2019). In summary, CRPC-Adeno and CRPC-NE are likely to represent a spectrum of phenotypes, and the criteria used to define a “neuroendocrine phenotype” should be explicitly stated in scientific studies, to optimize the interpretation of results.

Notwithstanding these considerations regarding the definition of NEPC, acquisition of neuroendocrine features in PCa is thought to represent one of the possible mechanisms of treatment resistance to ARSi (Abida et al., 2019b; Aggarwal et al., 2018; Beltran et al., 2016; Davies et al., 2018). A similar mechanism has been described in lung adenocarcinoma resistance to EGFR-TKI (Sequist et al., 2011). This phenomenon is in line with a broader framework proposed by Garraway et al. to appreciate the complexity of cancer cell drug resistance, describing three foundational resistance routes: pathway reactivation, pathway bypass and pathway indifference (Konieczkowski, Johannessen, & Garraway, 2018). CRPC-NE no longer responds to ARSi and carries a dismal prognosis, with a mean overall survival of 12 months (Metzger et al., 2019), and no specific standard of care treatment options are available to date (Beltran et al., 2016; H. T. Wang et al., 2014).

There is mounting evidence that CRPC-Adeno can progress to an AR-indifferent state through a mechanism of transdifferentiation under specific genomic conditions, including but not limited to *TP53*, *RB1*, and *PTEN* loss (Beltran et al., 2016; Davies et al., 2018; Ku et al., 2017; Mu et al., 2017). However, work by our group and others suggests that epigenetic mechanisms are also critical in this process (**Appendix II**) (Beltran et al., 2016; Davies et al., 2018; Mu et al., 2017). For example, DNA methylation profiles differ between CRPC-Adeno and CRPC-NE, and several DNA methyltransferases are significantly upregulated in CRPC-NE (Beltran et al., 2016). Similarly, the Polycomb (PRC2) protein EZH2 is highly expressed in CRPC-NE (Puca et al., 2018), and EZH2 inhibition can restore enzalutamide sensitivity and reverse lineage plasticity in cells deficient for *PTEN*, *RB1* and *TP53* (Ku et al., 2017).

Although the mammalian Switch Sucrose Non-Fermenting (mSWI/SNF) complex is another major epigenetic regulator well known for its role in physiological processes and frequently mutated in cancer (Kadoch & Crabtree, 2015; Kadoch et al., 2013; Shain & Pollack, 2013), SWI/SNF alterations in CRPC-NE have not been studied to date. Given the importance of SWI/SNF in cell fate determination, and the antagonistic relationship between SWI/SNF and PRC2 described in some cancers, we hypothesized that SWI/SNF may be mutated or deregulated through PCa progression, and in particular in CRPC-NE.

2. Results

2.1 SWI/SNF genes are rarely altered in prostate cancer.

While PCa is usually not cited among tumor types with recurrent SWI/SNF alterations, the exact frequency of such alterations in PCa, and especially in aggressive subtypes such as CRPC-Adeno and CEPC-NE, has not been reported. Thus, we started by specifically querying the status of SWI/SNF genes across prostate cancer disease states using three large published WES datasets: the TCGA dataset (Cancer Genome Atlas Research, 2015), the Weill Cornell

Medicine (WCM) dataset (Beltran et al., 2016) and the SU2C-PCF dataset (Abida, Cyrta et al., 2019a). The TCGA cohort consists of hormone treatment-naïve primary prostatic adenocarcinomas, while the WCM and the SU2C-PCF cohorts comprise CRPC patients. Samples from the WCM and SU2C-PCF cohorts were classified as CRPC-Adeno or CRPC-NE via a centralized consensus review by pathologists with experience in GU pathology, based on morphological criteria described in (Epstein et al., 2014). I was involved in coordinating the review process for the SU2C-PCF cohort (**Appendix I**).

Overall, using these three cohorts, our analysis of SWI/SNF genes was performed on 600 unique patient samples: 299 cases of localized treatment-naïve prostatic adenocarcinoma, 245 cases of CRPC-Adeno and 56 cases of CRPC-NE. The results are summarized in **Fig.25** and **Table 7**.

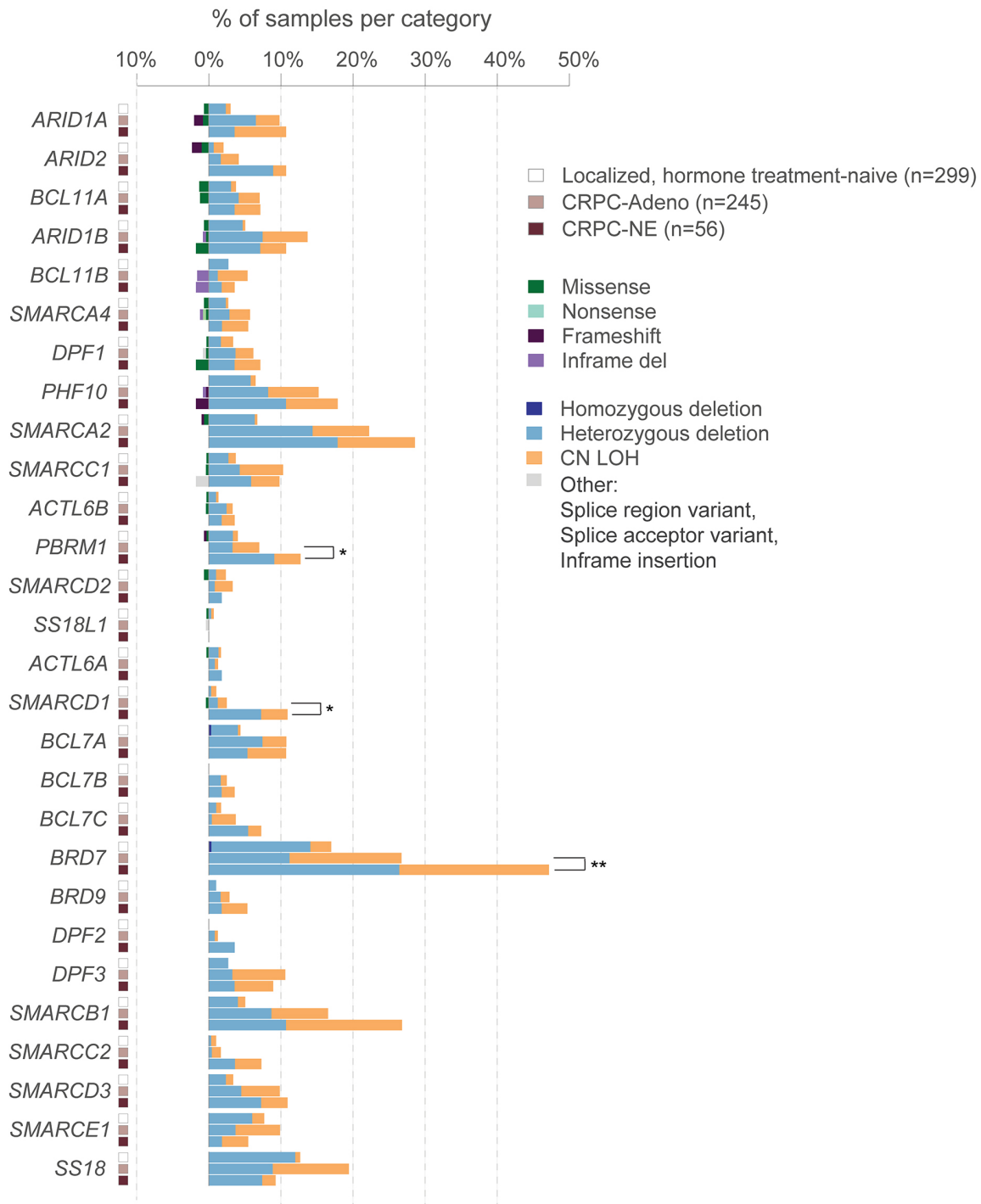


Figure 25. Genomic alterations (WES) in SWI/SNF genes across PCa disease states.
 CNL: copy number neutral LOH.

Gene name	Tumor category	Variant type	Protein Change	Ref reads normal	Alt reads normal	VAF normal	Ref reads tumor	Alt reads tumor	VAF tumor	Tumor purity
ACTL6A	CRPC-Adeno	Missense	p.G65S	134	1	0.007407	155	99	0.39	89%
ACTL6B	PCa	Missense	p.R386C	50	0	0	40	7	0.15	52%
ACTL6B	CRPC-Adeno	Splice_Site		15	0	0	42	15	0.26	83%
ACTL6B	PCa	Missense	p.V184D	78	0	0	104	7	0.06	24%
ACTL6B	CRPC-NE	Missense	p.S105N	56	0	0	20	21	0.51	NA
ARID1A	CRPC-Adeno	Missense	p.D322E	15	0	0	3	6	0.67	95%
ARID1A	CRPC-Adeno	Missense	p.M1300I	351	0	0	264	71	0.21	50%
ARID1A	PCa	Missense	p.P1305S	85	0	0	64	10	0.14	33%
ARID1A	PCa	Missense	p.R1551H	131	0	0	68	21	0.24	81%
ARID1A	PCa	NONSENSE	p.Q1650*	68	0	0	86	7	0.08	22%
ARID1A	PCa	Missense	p.P1756R	181	0	0	142	11	0.07	32%
ARID1A	CRPC-Adeno	Frame_Shift_Del	p.S2068fs	542	1	0.001845	206	51	0.20	NA
ARID1A	CRPC-NE	Frame_Shift_Del	p.P1478fs	96	0	0	28	33	0.53	NA
ARID1A	CRPC-Adeno	Frame_Shift_Del	p.V1753fs	61	0	0	76	3	0.04	NA
ARID1B	PCa	Missense	p.A781T	73	0	0	61	22	0.27	NA
ARID1B	CRPC-Adeno	Missense	p.R1062Q	146	0	0	178	144	0.45	89%
ARID1B	CRPC-Adeno	Splice_Site	p.R1217R	87	0	0	142	42	0.23	96%
ARID1B	CRPC-NE	Missense	p.T1639M	419	2	0.004751	131	87	0.40	90%
ARID1B	CRPC-Adeno	Frame_Shift_Del	p.L883fs	510	0	0	40	18	0.32	NA
ARID2	CRPC-Adeno	Missense	p.D243G	140	0	0	135	52	0.28	80%
ARID2	PCa	Missense	p.L327S	77	0	0	57	30	0.34	NA
ARID2	PCa	Missense	p.Q679H	118	0	0	130	17	0.12	NA
ARID2	CRPC-Adeno	Missense	p.I848M	347	0	0	454	56	0.11	96%
ARID2	CRPC-NE	NONSENSE	p.Q1207*	58	1	0.016949	41	21	0.34	NA
ARID2	PCa	Missense	p.C1271G	36	0	0	22	15	0.41	82%
ARID2	CRPC-Adeno	Missense	p.D1352G	69	0	0	81	18	0.18	NA
ARID2	CRPC-Adeno	Missense	p.A1773V	47	0	0	66	23	0.26	79%
ARID2	PCa	Frame_Shift_Del	p.I1764fs	92	0	0	61	47	0.42	NA
ARID2	PCa	Frame_Shift_Del	p.S1443fs	486	0	0	313	73	0.20	NA
ARID2	PCa	Frame_Shift_Ins	p.K35fs	61	0	0	7	7	0.50	NA
ARID2	PCa	Frame_Shift_Del	p.T1398fs	55	0	0	58	10	0.16	NA
ARID2	PCa	Frame_Shift_Del	p.P1087fs	148	0	0	108	28	0.20	NA
BCL11A	PCa	Missense	p.V551M	67	1	0.014706	48	4	0.08	33%
BCL11A	PCa	Missense	p.A459V	30	0	0	18	14	0.44	NA
BCL11A	CRPC-Adeno	Missense	p.P270L	22	0	0	9	4	0.31	NA
BCL11A	CRPC-Adeno	Missense	p.P263S	32	0	0	8	4	0.33	NA
BCL11A	CRPC-Adeno	NONSENSE	p.R196*	238	0	0	219	49	0.18	89%
BCL11A	CRPC-Adeno	Missense	p.R118K	320	1	0.003115	548	129	0.19	94%

BCL11A	CRPC-Adeno	Frame_Shift_Del	p.P265fs	205	1	0.004854	75	20	0.21	NA
BCL11B	CRPC-NE	Missense	p.A700V	193	0	0	109	9	0.08	NA
BCL11B	PCa	Missense	p.A240E	25	0	0	32	6	0.16	49%
BCL7A	CRPC-Adeno	Missense	p.V6F	9	0	0	3	5	0.63	67%
BCL7B	CRPC-Adeno	Frame_Shift_Del	p.Q80fs	107	1	0.009346	145	166	0.53	NA
BCL7C	CRPC-NE	Missense	p.P168H	29	0	0	69	20	0.22	92%
BCL7C	CRPC-Adeno	Missense	p.R123H	184	0	0	250	35	0.12	NA
BRD7	CRPC-NE	Splice_Site		54	1	0.021	20	3	0.17	NA
BRD7	PCa	Splice_Site		137	1	0.007353	149	7	0.05	NA
BRD7	PCa	Splice_Site		30	0	0	35	3	0.08	NA
BRD9	CRPC-Adeno	Missense	p.P171A	105	0	0	114	29	0.20	94%
BRD9	PCa	Missense	p.D167G	56	0	0	36	11	0.23	NA
BRD9	PCa	Missense	p.P157S	32	0	0	16	16	0.50	97%
BRD9	PCa	Missense	p.P140L	171	0	0	94	6	0.06	30%
BRD9	PCa	Intron		90	0	0	54	18	0.27	NA
DPF1	CRPC-NE	Missense	p.T316I	117	0	0	57	10	0.15	37%
DPF2	CRPC-NE	Missense	p.R300C	543	0	0	148	207	0.58	88%
PBRM1	CRPC-Adeno	Missense	p.Y994H	65	0	0	45	40	0.47	93%
PBRM1	PCa	Missense	p.K930N	144	0	0	108	10	0.08	NA
PBRM1	CRPC-Adeno	Missense	p.R710G	305	0	0	346	49	0.12	63%
PBRM1	CRPC-Adeno	Splice_Site		158	0	0	60	5	0.08	NA
PBRM1	CRPC-NE	Frame_Shift_Ins	p.N333fs	61	0	0	91	5	0.06	NA
PHF10	PCa	Missense	p.L257V	215	0	0	200	16	0.07	23%
PHF10	PCa	Missense	p.P244L	315	0	0	252	98	0.28	NA
SMARCA2	PCa	Missense	p.A1137V	145	0	0	116	17	0.13	35%
SMARCA2	CRPC-NE	Missense	p.D1399E	39	0	0	24	20	0.45	80%
SMARCA2	PCa	Frame_Shift_Del	p.I802fs	153	0	0	168	3	0.02	NA
SMARCA2	CRPC-NE	In_Frame_Ins	p.1362E>DIYI* LQQ	46	0	0	86	3	0.04	NA
SMARCA2	CRPC-NE	Frame_Shift_Ins	p.-804fs	86	0	0	103	3	0.04	NA
SMARCA2	PCa	Frame_Shift_Del	p.L1252fs	47	0	0	31	15	0.33	NA
SMARCA4	CRPC-Adeno	NONSENSE	p.Y64*	78	0	0	123	13	0.10	38%
SMARCA4	CRPC-NE	Missense	p.V1529A	80	0	0	63	5	0.07	25%
SMARCB1	CRPC-Adeno	Missense	p.T232R	474	0	0	110	297	0.73	96%
SMARCC1	CRPC-NE	Frame_Shift_Ins	p.A489fs	106	0	0	20	3	0.14	NA
SMARCC1	CRPC-Adeno	Splice_Site		10	0	0	14	8	0.38	NA
SMARCC2	CRPC-Adeno	Missense	p.D434E	261	0	0	327	45	0.12	46%
SMARCC2	PCa	Missense	p.W215R	45	0	0	32	7	0.18	NA

SMARCC2	PCa	Frame_Shift_Ins	p.-328fs	198	0	0	148	6	0.04	NA
SMARCD2	PCa	Missense	p.R357H	26	0	0	14	9	0.39	65%
SMARCD2	CRPC-Adeno	Missense	p.R195Q	22	0	0	10	7	0.41	98%
SMARCD2	PCa	Missense	p.R133W	114	0	0	77	67	0.47	92%
SMARCD2	CRPC-Adeno	Frame_Shift_Del	p.P92fs	192	0	0	24	4	0.14	NA
SMARCE1	PCa	Missense	p.N88D	339	0	0	190	71	0.27	77%
SMARCE1	CRPC-NE	Missense	p.N44H	72	0	0	162	19	0.10	42%
SS18	CRPC-Adeno	Missense	p.P292L	95	0	0	96	14	0.13	63%
SS18L1	CRPC-Adeno	Missense	p.S308F	132	0	0	172	72	0.30	75%

Table 7. SNV and indel results (WES) from 600 PCa patients with various disease states.

Overall, of the 28 SWI/SNF genes, no recurrent SWI/SNF somatic mutations were observed and there was a low overall rate of point mutations and insertions/deletions in those genes (59 samples, 9.8% of all cases).

In contrast to low point mutation rates in SWI/SNF genes, we did observe frequent loss-of-heterozygosity (LOH) by heterozygous deletion or by copy number neutral loss in those genes (**Fig.25**). However, this is most likely in keeping with the overall genomic instability of PCa, a well described characteristic of this cancer type (Beltran et al., 2016; Cancer Genome Atlas Research, 2015). We did observe an increased frequency of LOH in SWI/SNF genes in CRPC-Adeno as compared to localized hormone treatment-naïve PCa for 27 of 28 genes (significant for 15 genes, proportion test, alpha=0.05), and in CRPC-NE compared to localized hormone treatment-naïve PCa for 26 of 28 genes, which again could be in line with increased overall genomic instability through PCa progression.

Conversely, there were less differences when comparing CRPC-Adeno and CRPC-NE, and a significant increase in the fraction of LOH in CRPC-NE was only noted for three genes: *BRD7* (50% vs 28% respectively, p=0.005), *SMARCD1* (11% vs 3%, p=0.04), and *PBRM1* (18% vs 8%, p=0.049). To evaluate whether these alterations may potentially display functional significance, we verified the expression levels of these genes in CRPC-NE (**Fig.26**). For *PBRM1* and *SMARCD1*, gene expression was not significantly different in the CRPC-NE group. For

BRD7, expression was significantly lower in CRPC-NE compared to CRPC-Adeno, but not in CRPC-NE compared to localized PCa. It was previously shown that heterozygous loss of the region encompassing the *CYLD* gene is a frequent event in CRPC-NE (Beltran et al., 2016). Thus, given the genomic proximity between *BRD7* and *CYLD*, *BRD7* loss is likely part of a larger heterozygous deletion event centered around the *CYLD* gene. Overall, there is insufficient evidence to suggest that the above genomic alterations in *BRD7*, *PBRM1* and *SMARCD1* carry any functional significance.

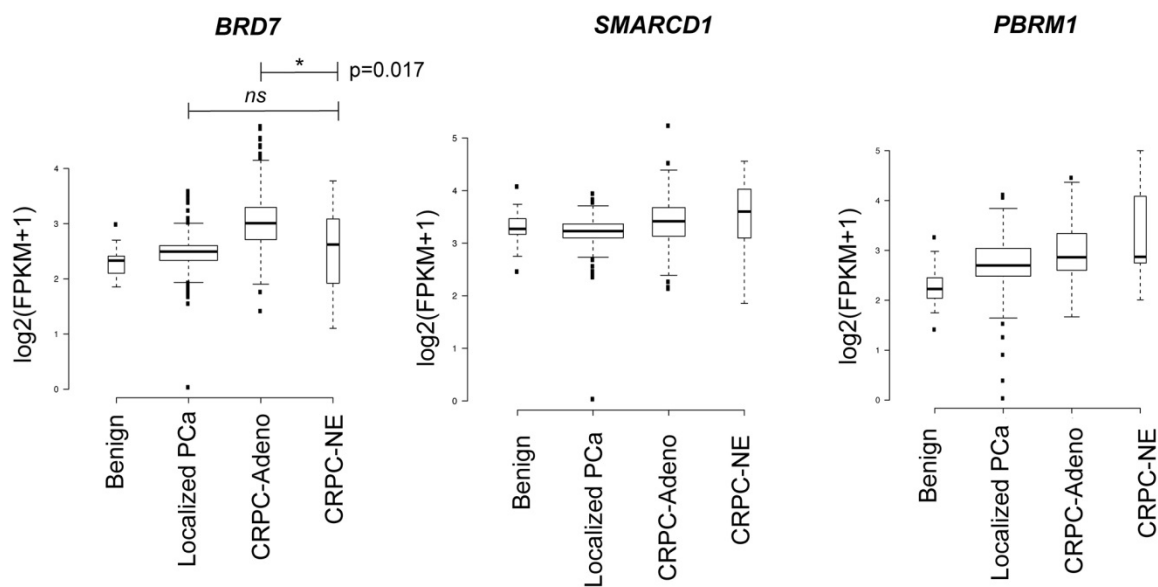


Figure 26. Gene expression levels (RNA-seq) of *BRD7*, *SMARCD1* and *PBRM1* across PCa disease states.

2.2 The expression of several SWI/SNF subunits varies across PCa disease states

Given the modest differential abundance of genomic lesions, we next queried the expression levels of SWI/SNF subunits by examining RNA-seq data from 572 unique patient samples: benign prostatic tissue (n=32), localized hormone treatment-naïve PCa (n=400), CRPC-Adeno (n=120) and CRPC-NE (n=20). The analyzed data came from previously published cohorts: benign (Beltran et al., 2011; Chakravarty et al., 2014), PCa (Beltran et al., 2011; Cancer Genome Atlas Research, 2015; Chakravarty et al., 2014), CRPC-Adeno and

CRPC-NE (Abida et al., 2019a) (the two latter studies authored or co-authored by our team, please see **Appendix I and II**).

The first notable finding was a significant upregulation of the *SMARCA4* (BRG1) ATP-ase subunit in CRPC-NE as compared to CRPC-Adeno: mean difference of averaged $\log_2(\text{FPKM}+1)=0.55$ ($p=0.015$, Mann-Whitney-Wilcoxon test). In addition, we observed a downregulation of the mutually exclusive paralogue, *SMARCA2* (BRM) in CRPC-NE as compared to CRPC-Adeno: mean difference of averaged $\log_2(\text{FPKM}+1)=-0.60$ ($p=0.02$) (**Fig.27**). A concordant result was observed when comparing *SMARCA4/SMARCA2* gene expression ratios in individual patients, which was significantly higher in the CRPC-NE group (median ratio=3.06) than in the CRPC-Adeno group (median ratio=1.07, $p=0.007$) (**Fig.27**).

The second notable finding was a strong upregulation of neural-specific mSWI/SNF subunit transcripts in CRPC-NE: *ACTL6B* (BAF53B), *DPF1* (BAF45B) and *SS18L1* (CREST) (means: 2.79, 1.19 and 3.58, respectively) compared to CRPC-Adeno (mean 0.24, $p=4.86e-06$; mean 0.35, $p=0.0016$; and mean 2.76, $p=6.85e-05$, respectively). These subunits have to date mainly been found to be expressed in post-mitotic neurons, as they serve instructive functions in neuronal differentiation (Lessard et al., 2007).

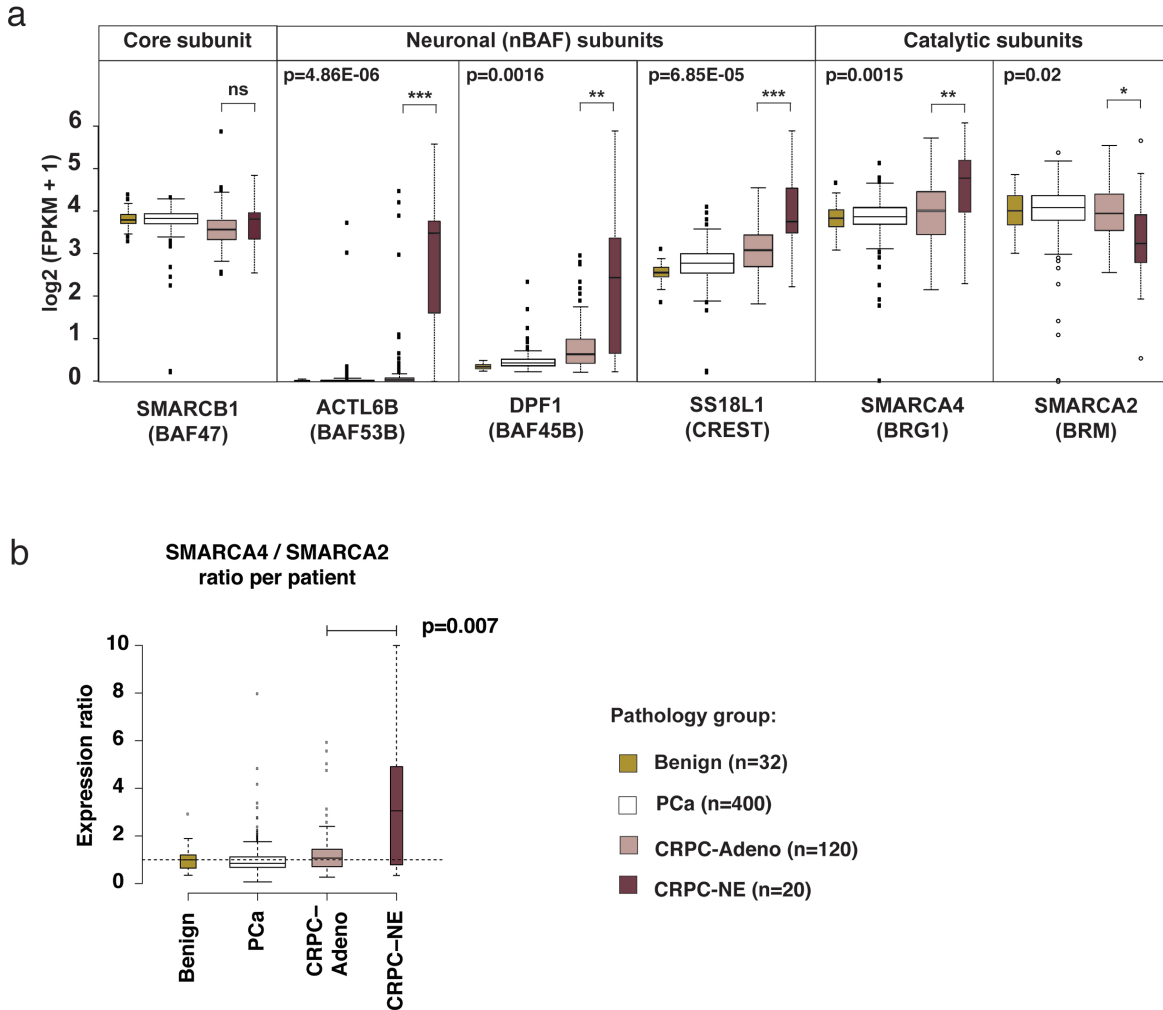


Figure 27. Gene expression levels of selected SWI/SNF genes across prostate cancer disease states. PCa: localized hormone treatment-naïve prostatic adenocarcinoma.

To validate these findings at the protein level, we performed immunohistochemistry (IHC) on clinical patient samples from a spectrum of PCa disease states using tissue microarrays (TMA) and whole tissue sections (cases were classified according to criteria mentioned above by pathologists from Weill Cornell) (**Fig.28**). We confirmed a trend towards higher SMARCA4 and lower SMARCA2 expression with increasing disease states. In particular, the highest SMARCA4 expression was observed in CRPC-NE. A subset of CRPC-NE samples showed a decrease or even a complete loss of expression of the SMARCA2 paralogue. We also performed a validation analysis of BAF53B (ACTL6B) and BAF45B (DPF1) expression by IHC

in patient samples and confirmed that these proteins were highly expressed at the nuclear level in CRPC-NE, but absent from benign prostate, localized PCa or CRPC-Adeno samples, suggesting that these two subunits may be specific markers of neuroendocrine differentiation. We chose to focus on BAF53B and BAF45B, and did not perform IHC for CREST (SS18L1), because mRNA expression levels suggested that CREST expression was less specific of CRPC-NE.

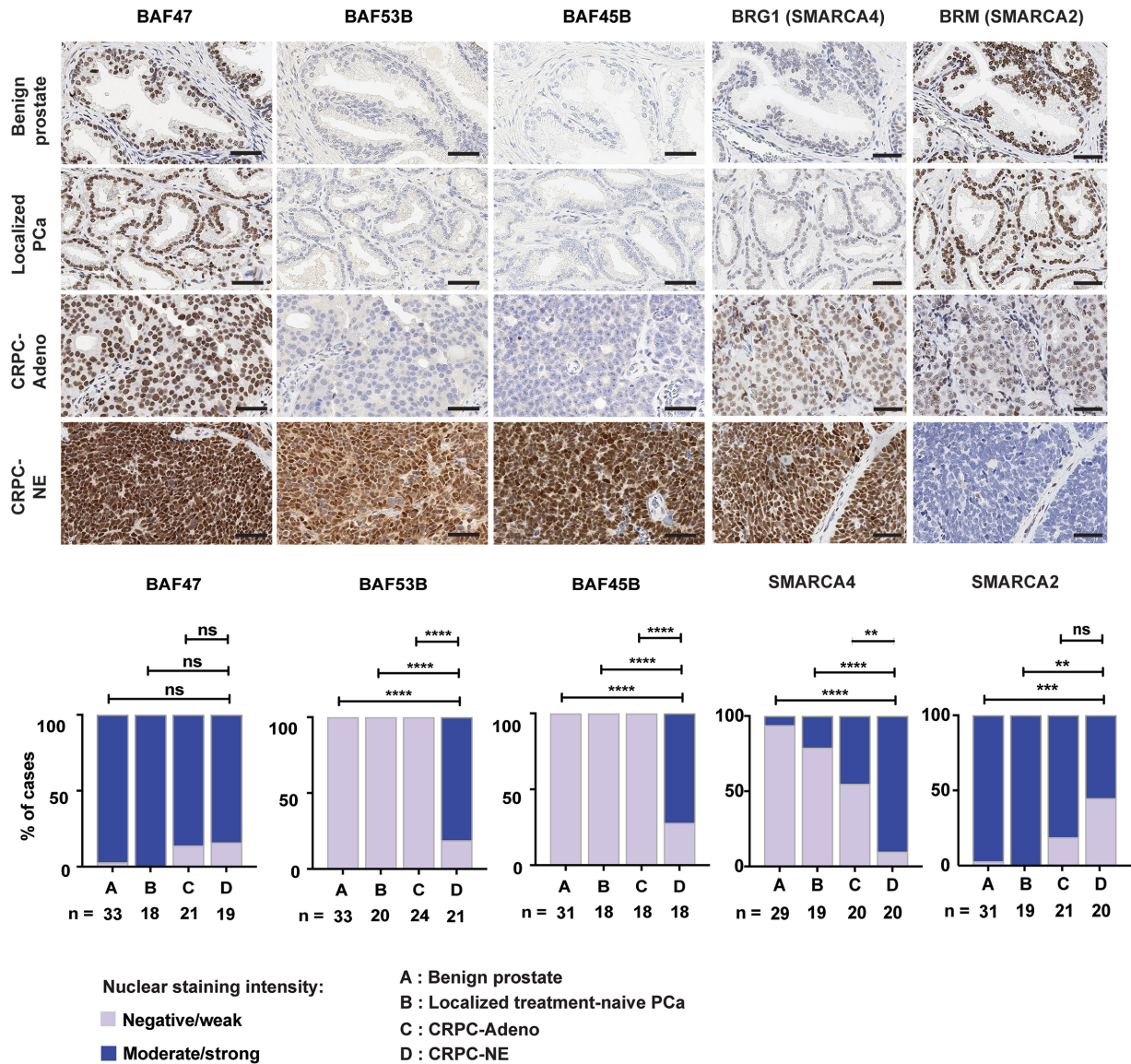


Figure 28. Immunohistochemistry for selected SWI/SNF subunits across PCa disease states. BAF47 (SMARCB1) is shown as control. **p < 0.01, ***p < 0.001, ****p < 0.0001, ns: non-significant (two-sided Fisher's exact test). Scale bars, 50 μ m.

Of note, we observed concordant results (i.e. higher SMARCA4 expression in less differentiated and/or more proliferative cell populations, and an inverse staining pattern of SMARCA2) in pilot examples of non-PCa patient samples (**Fig.29**).

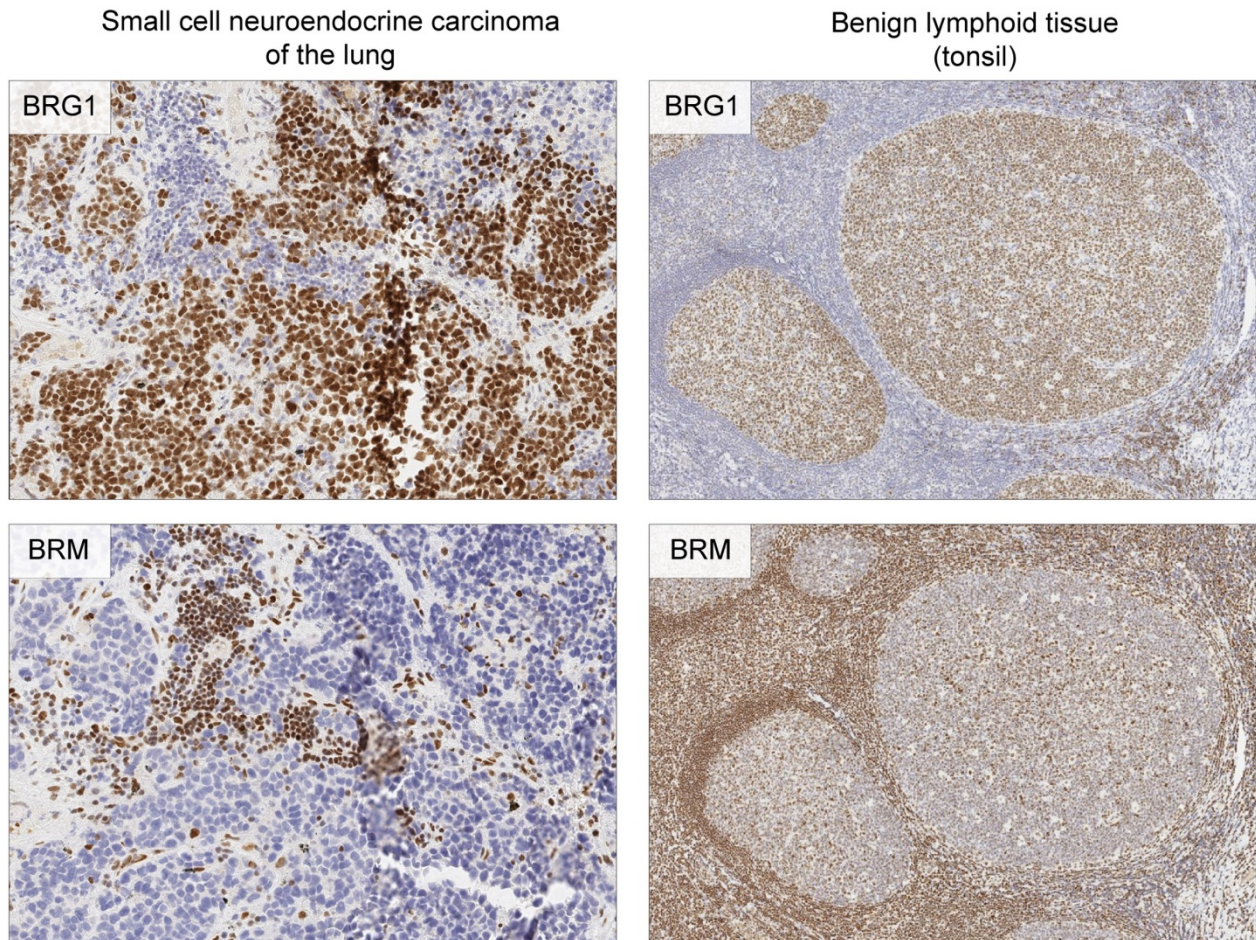


Fig.29. Examples of anti-SMARCA4 (BRG1) and anti-SMARCA2 (BRM) IHC in non-prostate tissues. *Left:* small cell neuroendocrine carcinoma of the lung, showing high SMARCA4 expression and negativity for SMARCA2 in tumor cells, consistent with our findings in prostate cancer (CRPC-NE). Conversely, the benign lymphocytic infiltrate within the tumor shows the opposite staining pattern (i.e. SMARCA4 low, SMARCA2 high). *Right:* benign lymphoid tissue (tonsil) showing high SMARCA4 expression in the germinal centres, and low expression in the interfollicular areas; while SMARCA2 displays the inverse staining pattern.

We next queried expression of SWI/SNF subunits in PCa cells lines by Western blotting, including in three CRPC-NE cell lines (NCI-H660, WCM154 and WCM155 (Puca et al., 2018)) (**Fig.30**). The “neuronal” SWI/SNF subunit BAF53B was expressed in CRPC-NE cell lines and

was also present, albeit at lower levels, in two synaptophysin-positive cell lines VCaP and 22Rv1, which bear some degree of transcriptomic similarities to neuroendocrine PCa cell lines although they are considered as adenocarcinoma cell lines (Bluemn et al., 2017). The second “neuronal” SWI/SNF subunit, BAF45B, was also expressed in CRPC-NE cell lines, but in addition, it was detected in several CRPC-Adeno cell lines (DU145, PC3 and MSKCC-PCA3). Although in neurons, BAF53B has been characterized as a mutually exclusive paralog to BAF53A, our data revealed that in CRPC-NE, BAF53A expression was maintained in cell lines expressing BAF53B. Unfortunately, BAF53A expression could not be queried in human samples by IHC, due to technical difficulties with antibody workup. As opposed to what we observed in human samples, Western blotting did not reveal a clear trend towards high SMARCA4/low SMARCA2 expression in CRPC-NE cell lines. However, this result should be interpreted with care, as several factors, including media composition and differences in cell proliferation rates between the cell lines, could influence the expression of these SWI/SNF catalytic subunits.

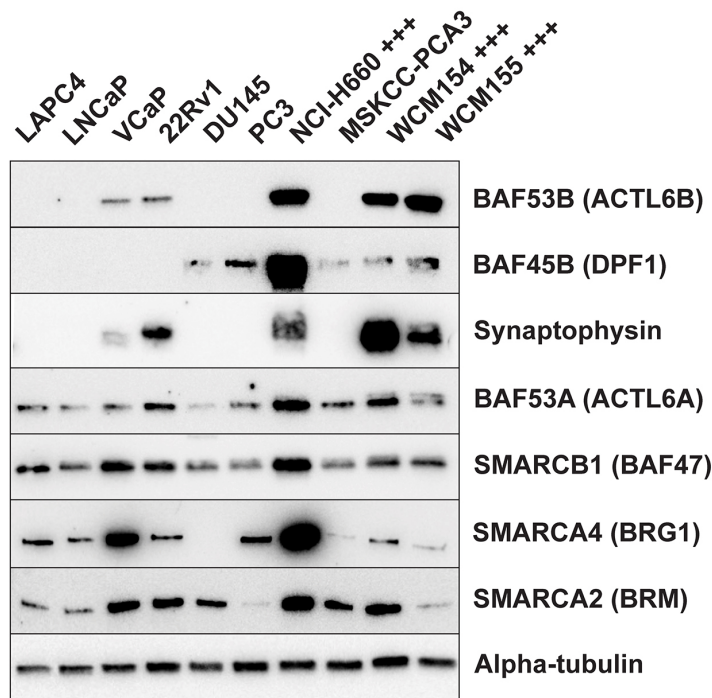
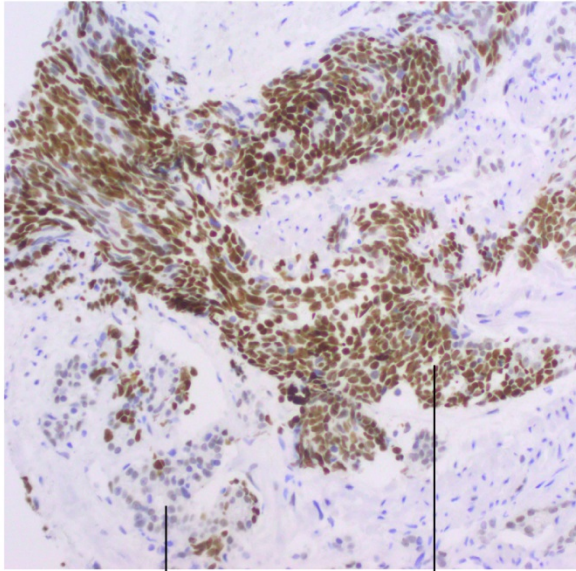


Figure 30. Immunoblotting for SWI/SNF subunits in PCa cell lines, including three CRPC-NE cell lines (denoted with +++).

2.3 Evidence of intra-tumor heterogeneity of SWI/SNF subunit expression

In addition, we noted intra-tumor heterogeneity in the expression of these subunits. In a pilot case of mixed prostatic carcinoma (combining areas with various degrees of adenocarcinoma or neuroendocrine differentiation), SMARCA4 expression was higher in the neuroendocrine component, while SMARCA2 expression was negative (**Fig.32**). These results are consistent with our RNA-seq findings, supporting a tendency towards a SMARCA4 high/SMARCA2 low phenotype in CRPC-NE as compared to adenocarcinomas. In a second case of mixed carcinoma (**Fig.33**), a similar tendency was observed, with higher SMARCA4 expression and lower SMARCA2 expression in the poorly differentiated component than in the glandular component. Of note, the poorly differentiated component in this case also showed loss of AR expression, high SOX2 expression, and positivity for BAF45B, but was not positive for neuroendocrine markers and for BAF53B (not shown).

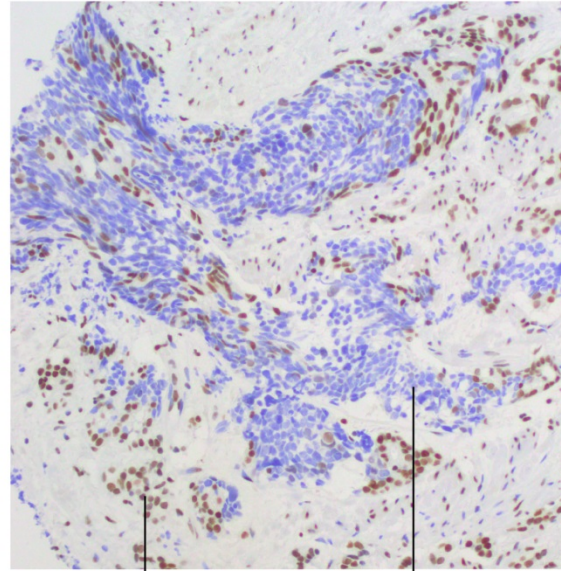
SMARCA4 (BRG1)



adenocarcinoma
component

neuroendocrine
carcinoma
component

SMARCA2 (BRM)



adenocarcinoma
component

neuroendocrine
carcinoma
component

Figure 32. Immunohistochemistry for BRG1 and BRM in a case of mixed prostatic carcinoma with neuroendocrine and acinar components.

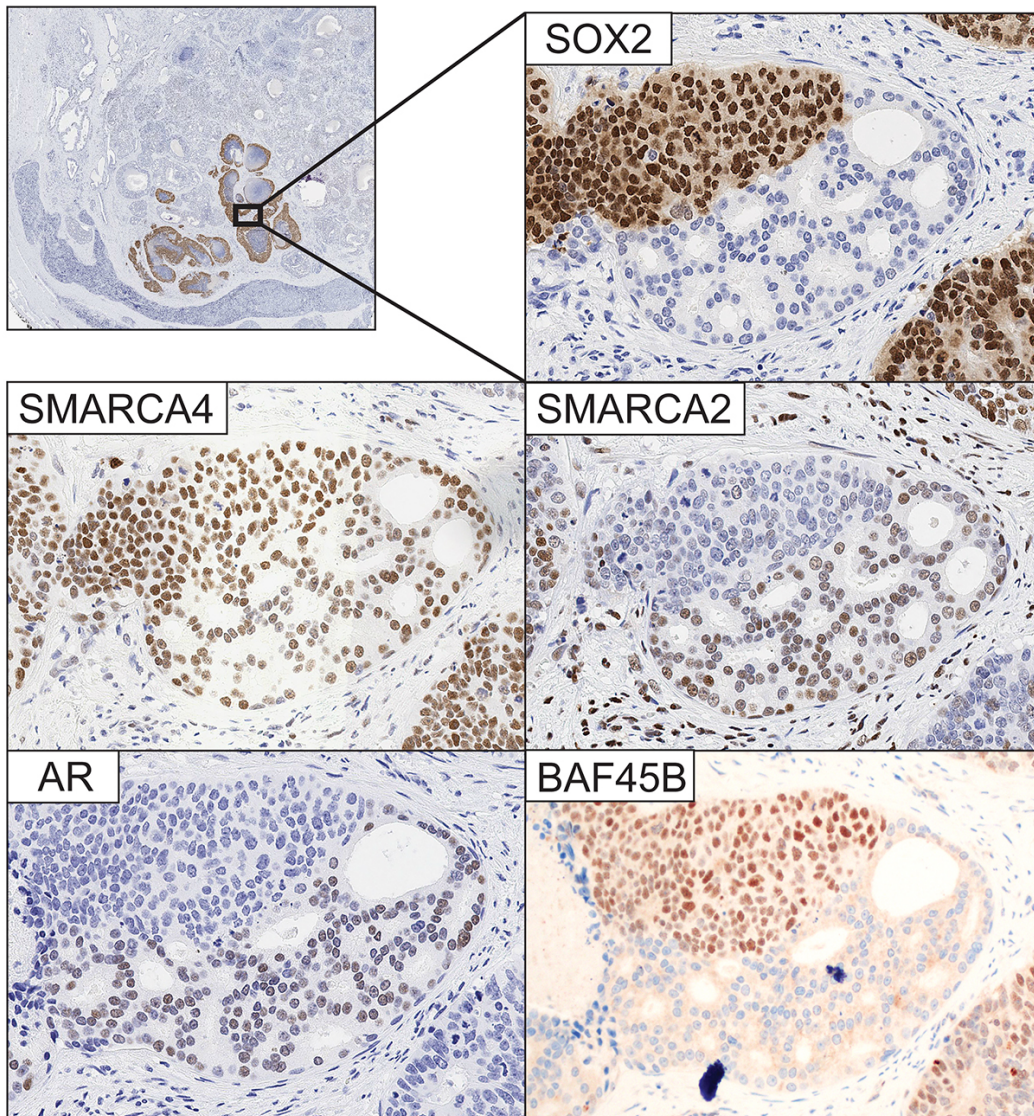


Figure 33. Immunohistochemistry for SWI/SNF subunits in a case of mixed prostatic carcinoma with poorly differentiated, AR-negative tumor areas with high SOX2 expression and areas of acinar adenocarcinoma with retained AR expression; BRG1=SMARCA4, BRM=SMARCA2, BAF45B=DPF1. The poorly differentiated areas were negative for neuroendocrine markers.

Lastly, we queried expression levels of SWI/SNF subunits by IHC in CRPC-NE organoids (WCM155 and WCM154) grown in 3D culture (formalin-fixed paraffin-embedded samples). We also observed intra-tumor heterogeneity in this setting, with the presence of distinct cell clusters that showed higher SOX2 expression, higher SMARCA4 expression and

lower SMARCA2 expression than other clusters (**Fig.34**). These clusters could represent a sub-population of less differentiated tumor cells, consistent with increased expression of the “stemness” regulator SOX2 and lacking the expression of terminal neural markers (synaptophysin or BAF53B), although putative tumor-perpetuating properties of this subpopulation remain to be verified functionally.

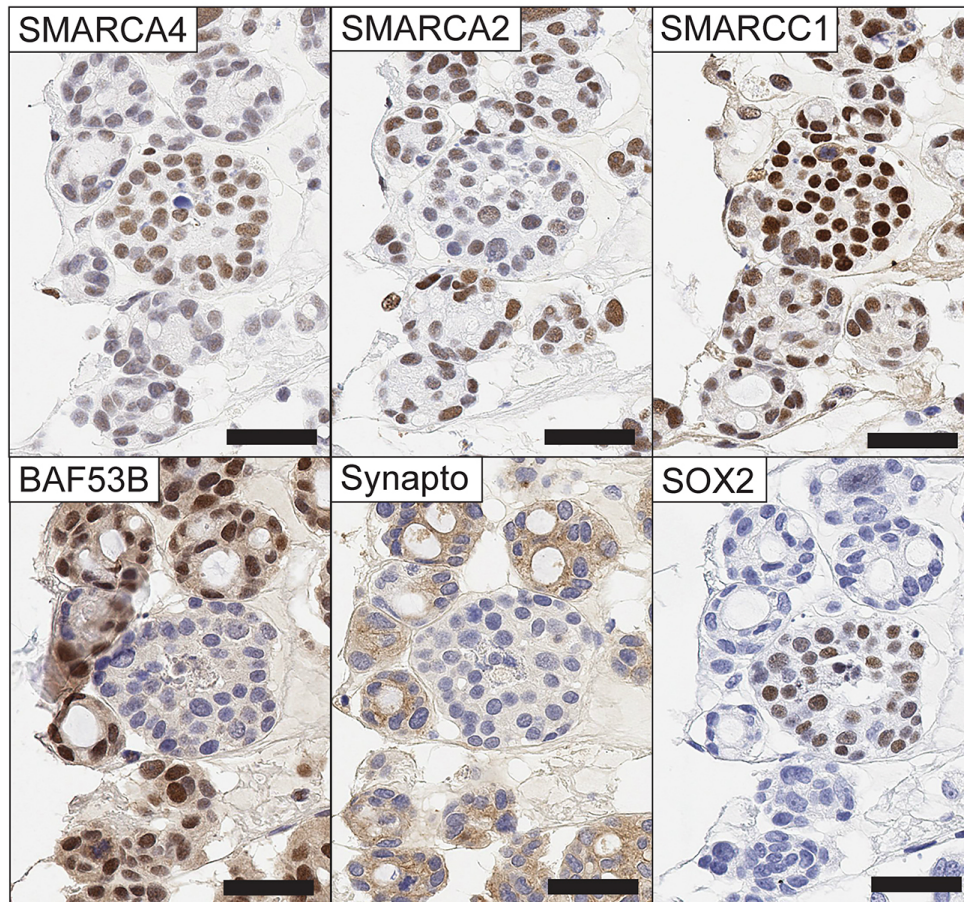


Figure 34. Immunohistochemistry for various SWI/SNF subunits and differentiation markers in a patient tumor-derived CRPC-NE organoid (3D culture) after FFPE processing. The case illustrates intra-tumor heterogeneity of the expression levels of SWI/SNF subunits in organoid cultures. Synapto: synaptophysin.

Overall, these observations suggest that SWI/SNF composition varies across PCa disease states and could be particularly relevant to CRPC-NE. This prompted a deeper exploration of the deregulated subunits.

2.4 SMARCA4 (BRG1) is overexpressed in aggressive PCa.

Given that SMARCA4 was more highly expressed in aggressive subtypes of PCa both at the transcriptomic and the protein level, we postulated that high SMARCA4 expression and/or low SMARCA2 expression could also be associated with worse prognosis in localized, treatment-naïve PCa. We used TMAs representing a cohort of radical prostatectomy specimens, described in (Spahn et al., 2010), with available clinical data (median follow-up of 79 months). Nuclear staining intensity was evaluated according to a semi-quantitative scale (0=absent, 1=weak, 2=moderate, 3=strong). Overall, 203 patients were included in the analysis. High SMARCA4 protein expression in primary PCa (score 3) was associated with a significantly shorter overall survival (HR=2.17 [97.5% CI: 1.07-4.42], p=0.028, log-rank test) (**Fig.35**). This relationship remained significant in multivariable models (performed with two predictors at a time), i.e. after adjustment for single predictors that have known association with PCa outcome (**Table 8**). Multivariable analysis with more predictors could not be performed, because of the risk of overfitting due to a relatively low number of events (a total of 31 deaths).

In patients with high SMARCA2 protein expression in primary PCa (score 3), there was a trend towards a better overall survival, but this relationship did not reach statistical significance.

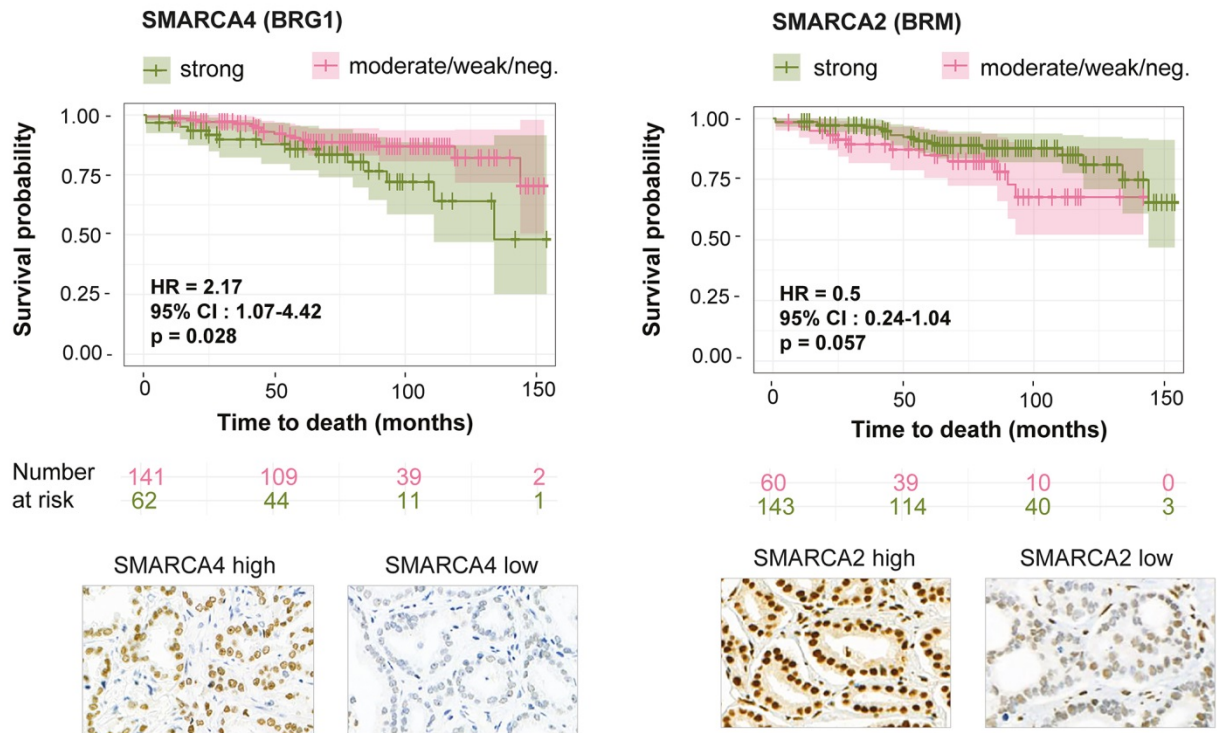


Figure 35. Overall survival in PCa patients stratified by SMARCA4 and SMARCA2 IHC staining intensity. Kaplan-Meier curves showing the association between SMARCA4 (BRG1) and SMARCA2 (BRM) IHC expression and overall survival in 203 patients with localized PCa (Log-rank test).

SMARCA4 (BRG1) IHC scoring			
	HR	95% CI	p.value
SMARCA4_score 3	2.17	1.07, 4.42	0.03
SMARCA4_score 3	2.3	1.13, 4.69	0.02
Positive lymph nodes	0.58	0.26, 1.26	0.17
SMARCA4_score 3	2.2	1.07, 4.53	0.03
Gleason grade group 1	2.21	0.29, 16.89	0.45
Gleason grade group 2/3	4.11	0.53, 32.08	0.18
Gleason grade group 4/5	2.05	0.22, 18.73	0.52
SMARCA4_score 3	2.12	1.03, 4.36	0.04
Extraprostatic extension	1.24	0.42, 3.61	0.7
SMARCA4_score 3	2.21	1.08, 4.51	0.03
Margins_positive	0.79	0.36, 1.73	0.55
Margins_not available	0.88	0.33, 2.35	0.79

SMARCA4_score 3	2.14	1.05, 4.35	0.04
Adjuvant Abiraterone	1.87	0.65, 5.35	0.25
BRG1_score 3	2.13	1.04, 4.34	0.04
Adjuvant radiation	1.57	0.47, 5.21	0.46
SMARCA2 (BRM) IHC scoring			
	HR	95% CI	p.value
SMARCA2_score 3	0.5	0.24, 1.04	0.06
SMARCA2_score 3	0.46	0.22, 0.97	0.04
Positive lymph nodes	0.57	0.26, 1.25	0.16
SMARCA2_score 3	0.51	0.24, 1.07	0.08
Gleason grade group 1	2.48	0.33, 18.88	0.38
Gleason grade group 2/3	4.24	0.54, 33.07	0.17
Gleason grade group 4/5	2.25	0.25, 20.39	0.47
SMARCA2_score 3	0.5	0.24, 1.04	0.06
Extraprostatic extension	1.44	0.5, 4.16	0.5
SMARCA2_score 3	0.49	0.23, 1.04	0.06
Margins_positive	0.86	0.39, 1.9	0.71
Margins_ not available	1.05	0.39, 2.81	0.93
SMARCA2_score 3	0.48	0.23, 1.01	0.05
Adjuvant Abiraterone	2.02	0.7, 5.8	0.19
SMARCA2_score 3	0.47	0.22, 0.99	0.05
Adjuvant radiation	2.08	0.62, 7	0.23

Table 8. Analysis of association between SMARCA4 (BRG1) and SMARCA2 (BRM) protein expression (strong vs. moderate/weak/negative) and patient's overall survival adjusted for single covariates known to have an impact on PCa prognosis.

2.5 Effects of SMARCA4 (BRG1) and SMARCA2 (BRM) knock-down on gene expression levels in LNCaP cells.

To explore the effects of SMARCA4 and SMARCA2 depletion in PCa cell lines, we performed siRNA-mediated knock-down of SMARCA4 or SMARCA2 in an androgen-sensitive prostatic adenocarcinoma cell line LNCaP and compared global transcriptomic alterations using RNA-seq. While SMARCA2 knock-down led to modest transcriptomic alterations, SMARCA4 knock-down demonstrated a stronger effect on the transcriptome (**Fig. 36a**). Among the genes

most significantly deregulated upon SMARCA4 knock-down there were several of known significance in PCa progression, including: downregulation of *EZH2* and downregulation of the oncogenic long non-coding RNA *PCAT-1*, as well as an up-regulation in tumor suppressors such as cell cycle regulators *CDKN1A* (p21) and *BTG2* (Beltran et al., 2016; Coppola et al., 2013; Prensner et al., 2014) (**Fig. 36b**). Changes in p21 and EZH2 expression were validated at the protein level by Western blotting (**Fig. 36c**).

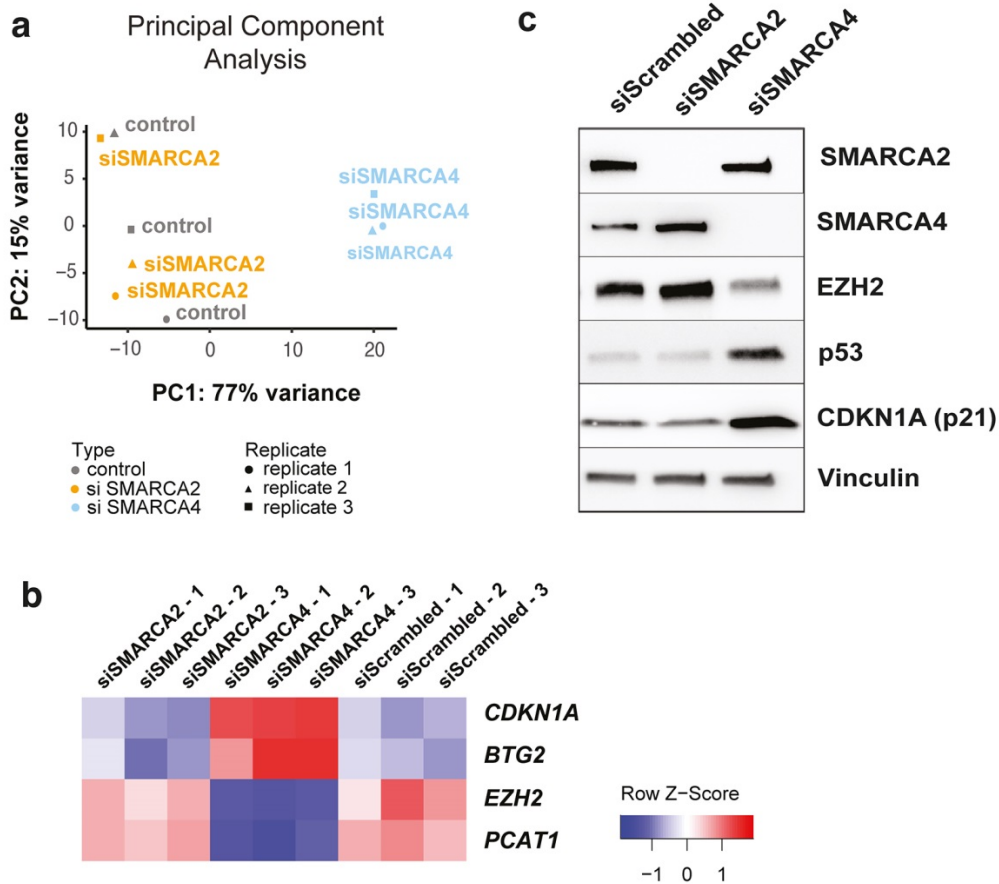


Figure 36. Transcriptomic changes upon siRNA-mediated SMARCA4 (BRG1) or SMARCA2 (BRM) knock-down in LNCaP cells. (a) Principal Component Analysis (RNA-seq data). **(b)** Heatmap representation of the expression levels (RNA-seq data) of selected genes with significantly ($FDR < 0.05$) differentially expressed upon SMARCA4 (BRG1) knock-down. **(c)** Western blot validating a decrease in EZH2 expression and an increase in p21 expression upon SMARCA4 (BRG1) knock-down.

Gene set enrichment analysis (GSEA) showed that upon SMARCA4 knock-down in

LNCaP cells, there was, among others, a negative enrichment for E2F targets (HALLMARK_E2F_TARGETS: genes encoding cell cycle related targets of E2F transcription factors) and a positive enrichment of EZH2 targets (NUYTEN_EZH2_TARGETS_UP: genes upregulated upon knock-down of EZH2 in prostate cancer PC3 cells)(**Fig.37**). This further suggests that the transcriptomic effects of SMARCA4 knock-down in LNCaP cells are opposed to cell cycle progression, and are somewhat similar to the effects of EZH2 knock-down. These enrichments were not observed for the SMARCA2 knock-down condition (not shown).

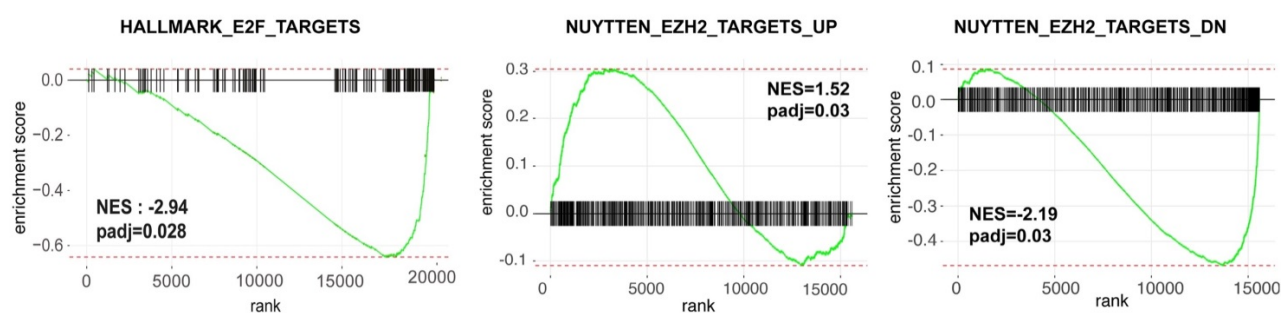


Figure 37. Gene Set Enrichment Analysis (GSEA) for SMARCA4 knock-down in LNCaP cells. results for selected gene sets are shown.

In addition, GSEA analysis comparing transcriptomic changes in LNCaP cells upon SMARCA4 and SMARCA2 depletion to the SHEN_SMARCA2_TARGETS_UP gene set (genes whose expression positively correlates with SMARCA2 in prostate samples) showed negative enrichment in cells with SMARCA2 knock-down (as expected), and positive enrichment in cells with SMARCA4 knock-down, further suggesting differences between the SMARCA4 and SMARCA2 knock-down conditions (**Fig.38**).

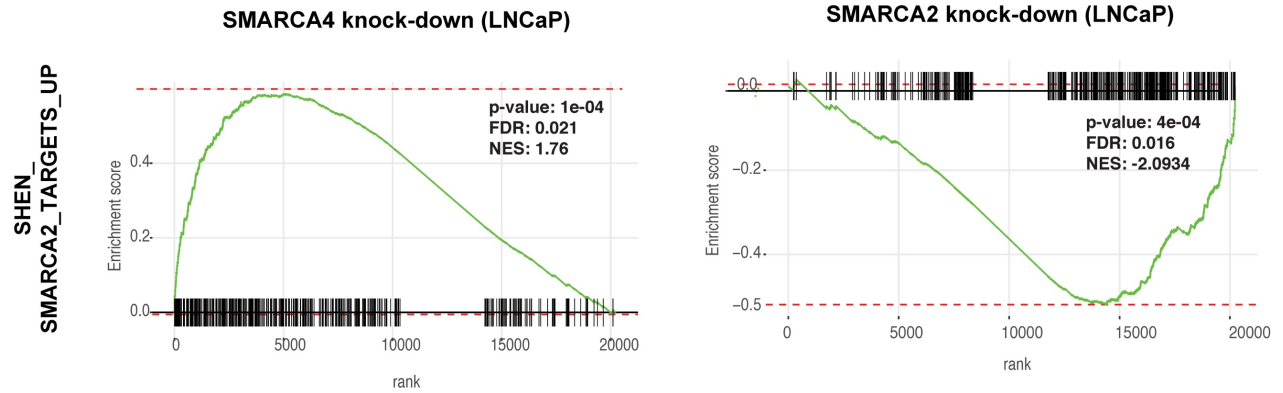


Fig.38. GSEA analysis comparing SMARCA4 and SMARCA2 knock-down (our study) in LNCaP cells to the SHEN_SMARCA2_TARGETS_UP gene set.

A study by Prensner et al. previously demonstrated that high expression of the long non-coding RNA SChLAP1 promoted prostate cancer aggressiveness, and suggested an antagonistic relationship between SChLAP1 and SWI/SNF (Prensner et al., 2013). We performed GSEA analysis comparing gene sets from their study (genes upregulated or downregulated upon SChLAP1 depletion in PCa cells) with our data from SMARCA4 and SMARCA2 knock-down in LNCaP cells (**Fig.39**). While the results suggested an inverse relationship between SMARCA2 knock-down and SChLAP1 knock-down, this was less evident for the SMARCA4 knock-down condition, and there was even a tendency towards positive enrichment of “genes upregulated upon SChLAP1 knock-down” in the SMARCA4 knock-down condition.

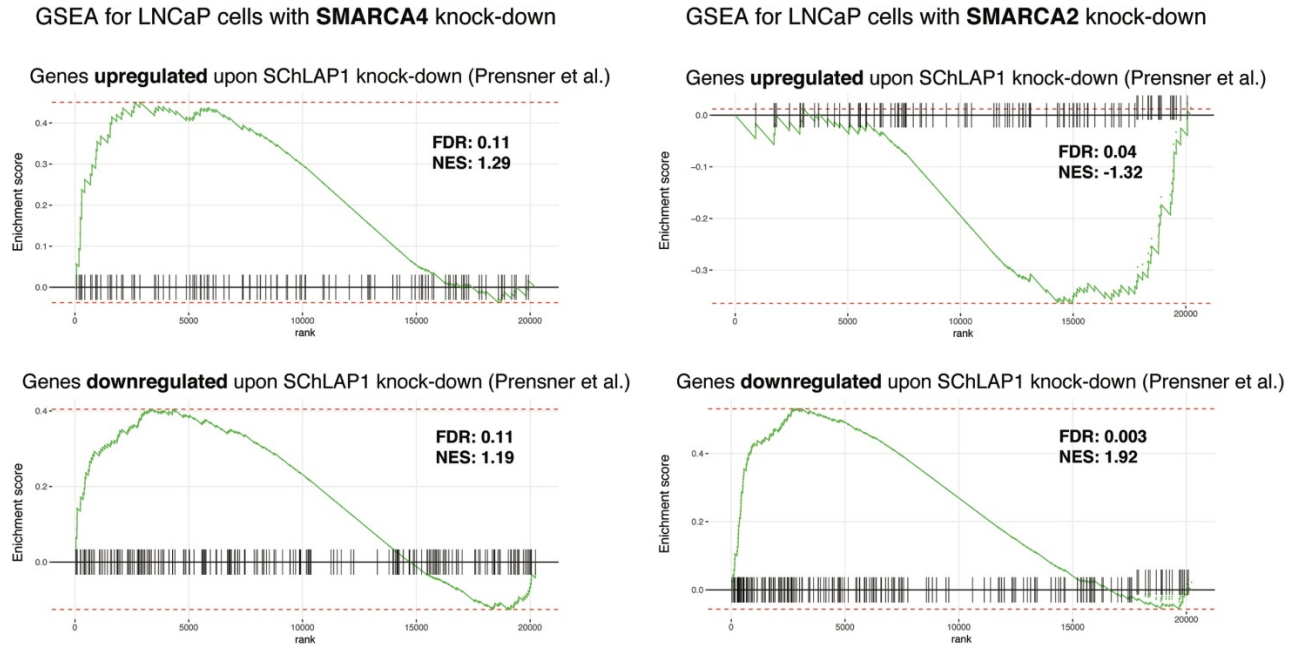


Fig.39. GSEA analysis comparing SMARCA4 and SMARCA2 knock-down (our study) and SchLAP1 knock-down (Prensner et al.) in LNCaP cells.

2.6 Effects of SMARCA4 (BRG1) and SMARCA2 (BRM) knock-down on cell growth in PCa cell lines.

Based on this exploratory analysis, which showed deregulation of cell cycle-related genes and pathways in PCa cells upon SMARCA4 knock-down, we sought to explore the requirement for SMARCA4 and SMARCA2 for PCa cell growth. Depletion of SMARCA4, but not of SMARCA2, significantly reduced proliferation of the AR-dependent adenocarcinoma cell line LNCaP and the CRPC-Adeno cell line C4-2 (which is derived from LNCaP), but not of the 22Rv1 cell line (not shown) or of the CRPC-NE cell line WCM154 (this discrepancy is discussed later in the text) (**Fig.40**).

Of note, combined knock-down of SMARCA4 and SMARCA2 in LNCaP and C4-2 cells did not have a stronger effect on cell growth than did the knock-down of SMARCA4 alone.

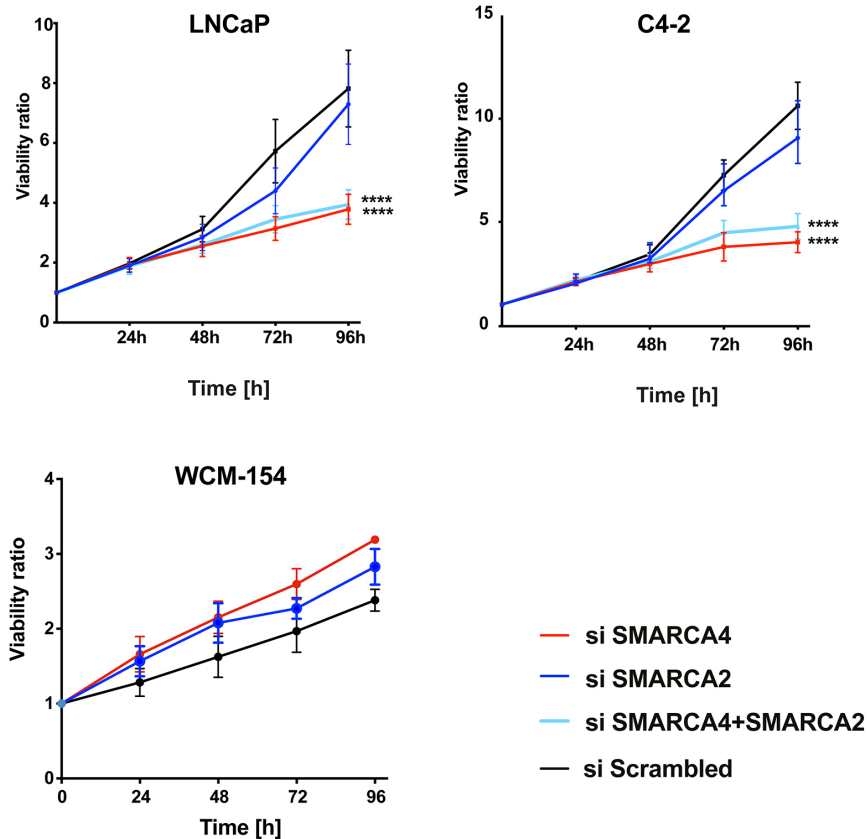


Figure 40. Effects of SMARCA4 and SMARCA2 knock-down on cell growth of prostatic adenocarcinoma cells (LNCaP), CRPC-Adeno cells (C4-2) and CRPC-NE cells (WCM-154). Cell viability assay (Cell Titer Glo), three pooled independent replicates, two-way ANOVA test. **** $p < 0.0001$.

CRPC-NE are known to show an enrichment of loss-of-function alterations in *TP53* and *RB1* (Beltran et al., 2016). To verify whether the observed effect was dependent on the presence of functional p53 and/or Rb, we used LNCaP cells with CRISPR-Cas9 mediated knock-out of *TP53*, *RB1* or both genes (**Fig.41**). This experiment showed that the effect of SMARCA4 knock-down on cell proliferation was not entirely abrogated by the absence of functional p53 and/or Rb.

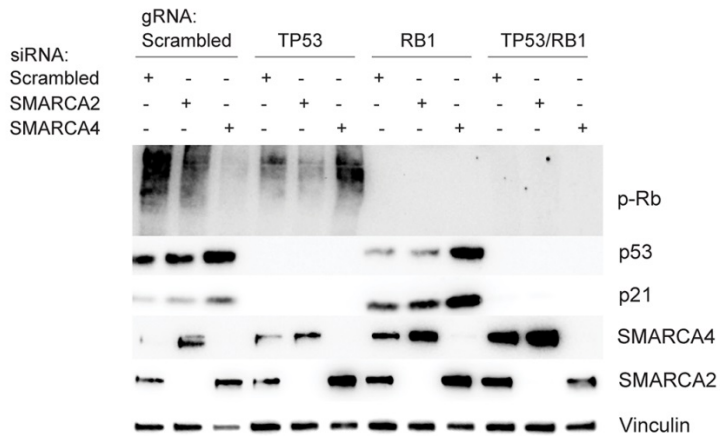
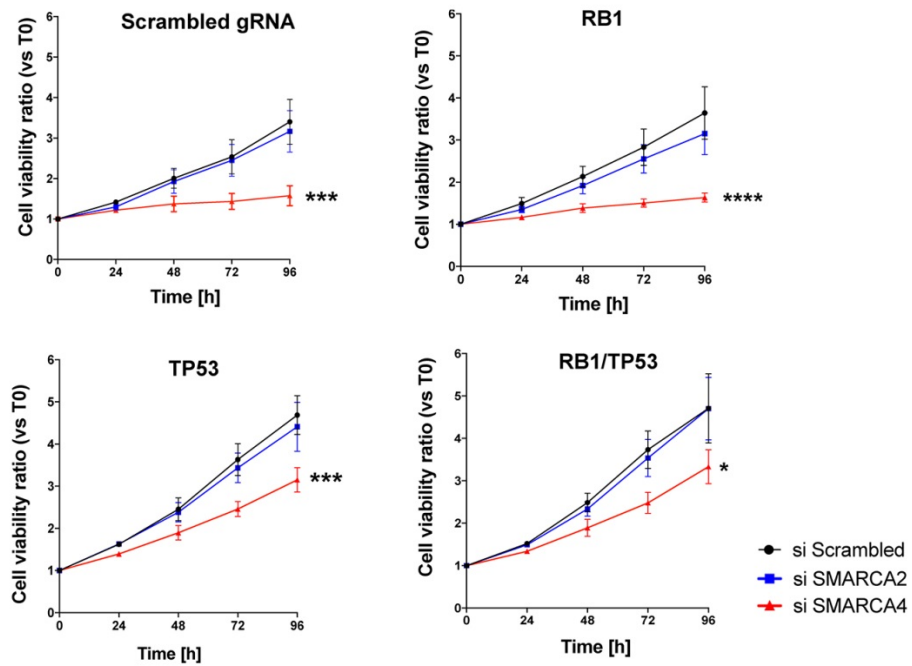


Figure 41. The effects of SMARCA4 (BRG1) and SMARCA2 (BRM) knock-down on cell growth in LNCaP cells with p53 and/or Rb loss. Cell viability assay (Cell Titer Glo), three pooled independent replicates, two-way ANOVA test (* $p < 0.05$, ** $p < 0.001$, **** $p < 0.0001$). The Western blot shows representative results from one of the three replicates.

2.7 Effects of SMARCC1 and SMARCC2 knock-down on PCa cell growth

We also observed that knock-down of SMARCA4, but not of SMARCA2, in PCa cells induced a decrease of some SWI/SNF subunits, including SMARCC1 and BAF53A, at the protein level, but not at the mRNA level (**Fig.42**), which suggests that SMARCA4 loss may destabilize the complex and lead to the degradation of the released subunit. This suggests that the effects observed upon SMARCA4 knock-down may, at least in part, be due to a global destabilization of the complex.

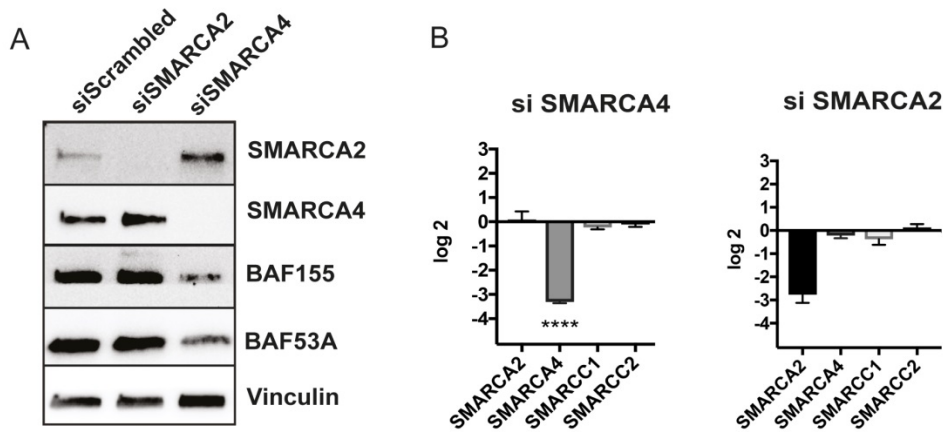


Figure 42. Effects of SMARCA4 or SMARCA2 knock-down on expression levels of other SWI/SNF subunits in LNCaP cells. **A.** Western blot showing that the protein levels of the SWI/SNF subunits SMARCC1 (BAF155) and BAF53A decrease upon SMARCA4 depletion, but not upon SMARCA2 depletion. **B.** RT-PCR showing that changes of SMARCC1 (BAF155) expression at the protein level are not explained by changes at the mRNA level.

Although transcriptomic data did not reveal a significant upregulation of *SMARCC1* in PCa disease progression, we sought to determine whether SMARCC1 knock-down would also have an effect on PCa cell growth (**Fig.43**). siRNA-mediated knock-down of SMARCC1 decreased the growth of prostatic adenocarcinoma cells, but also of the CRPC-NE cell line WCM154. A similar effect was observed upon knock-down of the SMARCC2 paralog in prostatic adenocarcinoma cells, although not in CRPC-NE cells.

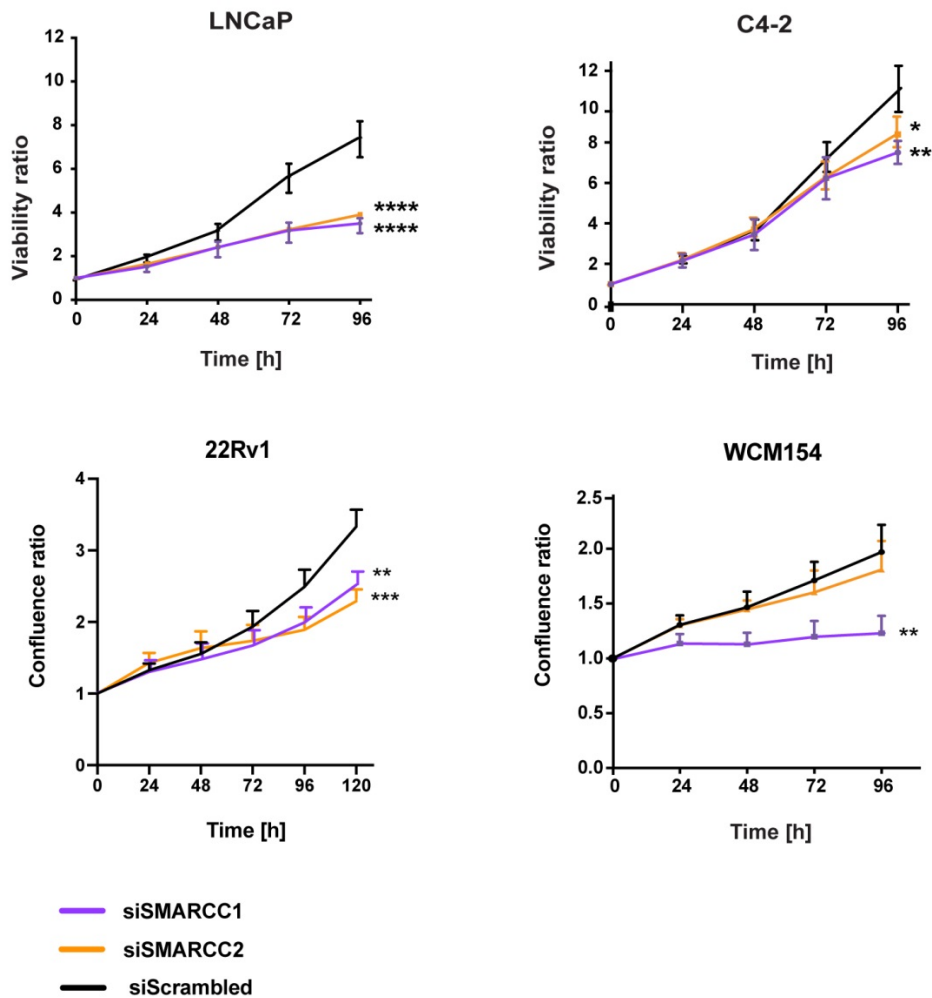


Figure 43. Effects of SMARCC1 or SMARCC2 siRNA-mediated knock-down on PCa cell growth. LNCaP: prostatic adenocarcinoma cells, C4-2: CRPC-Adeno cells (derived from LNCaP), 22Rv1: prostatic adenocarcinoma cells (considered CRPC-Adeno in some studies), WCM154: CRPC-NE cells. Viability ratio: results of the cell viability assay (Cell Titer Glo) relative to measurement at T0. Confluence ratio: cell confluence relative to measurement at T0, as measured using the Incucyte® system. Each graph represents pooled data from 3 independent experiments. *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001, two-way ANOVA, error bars: standard error.

2.8 Overexpressing SMARCA4 and SMARCA2 in PCa cells

To strengthen our hypothesis that SMARCA4 may show a tumor-promoting role in prostate cancer cells, we sought to investigate the effects of SMARCA4 overexpression in prostate cancer cells. 22Rv1 prostate adenocarcinoma cells (which show intermediate levels of

SMARCA4 protein by immunoblotting as compared to other PCa cells, **Fig.30**), were stably transduced with a lentiviral vector designed for overexpression of either SMARCA4 or SMARCA2 (pEZ-Lv203 or pEZLv216 vectors, respectively; Genecopoeia, MD, USA), or with a matched empty control vector, and sorted based on the expression of the fluorescent reporter (eGFP or mCherry, respectively). Despite strong expression of the fluorescent reporters, no increase in SMARCA4 or SMARCA2 protein levels was seen (**Fig.44**). However, after an additional 24h treatment with the proteasome inhibitor MG132, the expected overexpression of SMARCA4 and SMARCA2 could readily be detected. These observations suggest that the levels of catalytic SWI/SNF subunits may be subject to tight regulation and may depend on the cellular context at a given time. Forced, isolated overexpression of one subunit seems to result in rapid degradation of this excess subunit. Thus, SMARCA4 overexpression may be necessary, but not sufficient as a sole event, to promote an aggressive phenotype in prostate cancer cells.

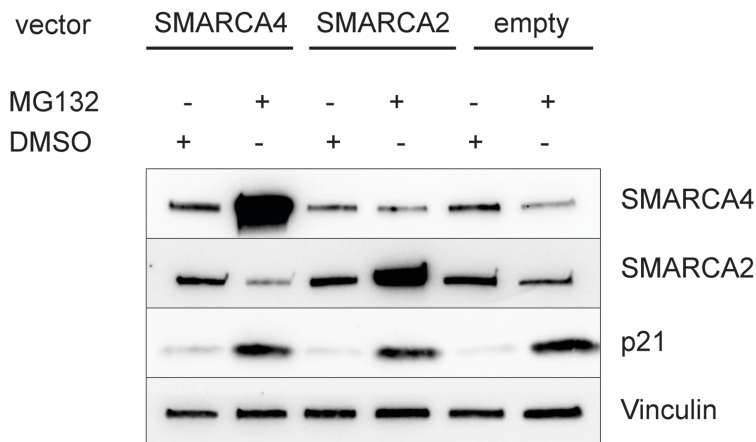


Figure 44. Immunoblot showing SMARCA4 and SMARCA2 expression levels in 22Rv1 cells transduced with lentiviral overexpression vectors and treated for 24h with MG132 or DMSO (control). The expression levels of the p21 protein are shown as a positive control for the MG132 treatment effect (Kwon et al., 2002).

2.9 Correlations between SMARCA4 knock-down scores and prognostic parameters in patient PCa samples.

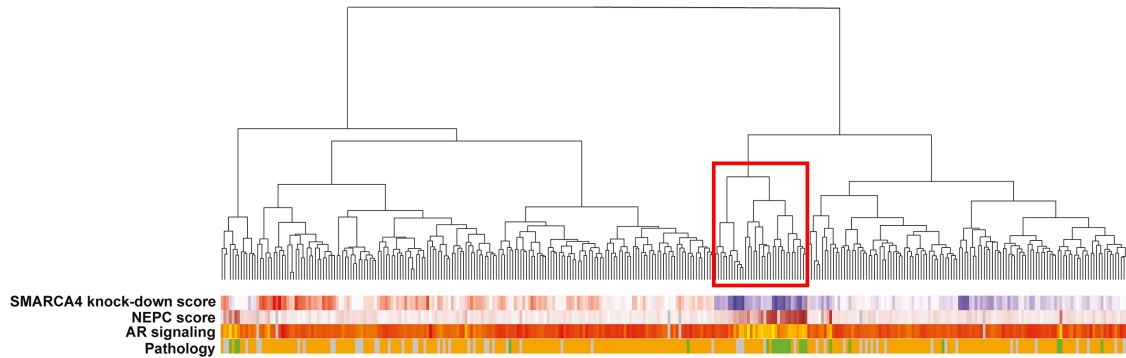
Using the RNA-seq data generated in LNCaP cells upon *SMARCA4* knock-down, we derived a *SMARCA4* knock-down signature using the 419 most deregulated genes. We postulated that since *SMARCA4* expression is higher in more aggressive PCa cases, the *SMARCA4* knock-down signature score would prove to be lower in such cases. To explore this hypothesis, we applied the signature to several highly annotated PCa cohorts with available RNA-seq data was available.

First, we examined two CRPC cohorts: the recently published Stand Up To Cancer-Prostate Cancer Foundation (SU2C-PCF) trial (n=332 patients) (Abida, Cyrta et al. 2019b) and the Weil Cornell Medicine (WCM) cohort (n=47 patients) (Beltran et al., 2016). Both cohorts include CRPC-Adeno and CRPC-NE cases. In the SU2C-PCF cohort, when considering patients from the highest (top 25%) and lowest (bottom 25%) quartiles of the *SMARCA4* knock-down signature scores (n=138), a low *SMARCA4* knock-down signature score (n=73) was significantly more frequent in CRPC-NE cases (n=16, 100%) than in CRPC-Adeno cases ($p=1.77e-05$; Fisher's test) (**Fig. 45A**). A similar result was obtained in the WCM cohort when considering the highest and lowest quartiles (n=25): a low *SMARCA4* knock-down signature score was seen in 89% (n=8) of CRPC-NE cases *versus* 31% (n=5) of CRPC-Adeno cases ($p=0.011$) (**Fig. 45B**). A low *SMARCA4* knock-down signature was also associated with a higher transcriptomic NEPC score (Beltran et al., 2016) and a lower AR signaling score (Hieronymus et al., 2006) in both cohorts (**Fig.45A and B, Table 9**).

A

SU2C-PCF 2019

SMARCA4 knock-down signature	Pathology		Fisher test p. value
	CRPC-NE	CRPC-Adeno	
LOW	16	57	1.768e-05
HIGH	0	65	



B

WCM Nat Med 2016

SMARCA4 knock-down signature	Pathology		Fisher test p. value
	CRPC-NE	CRPC-Adeno	
LOW	8	5	0.01117
HIGH	1	11	

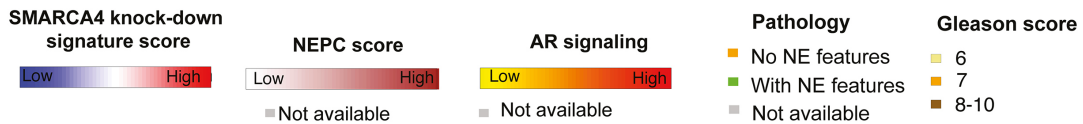
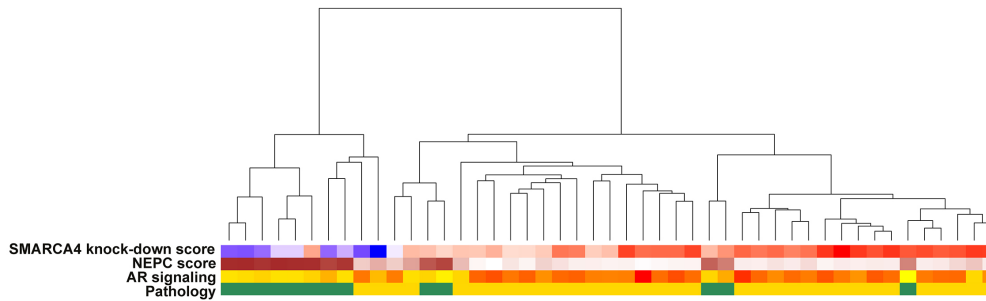


Figure 45. The transcriptomic SMARCA4 knock-down score, NEPC score, AR signaling score, and pathology data analyzed in: **A.** 332 cases of CRPC from the SU2C-PCF cohort (table, n=138); **B.** 47 cases of CRPC from the WCM cohort (table, n=25).

RNA-seq based unsupervised clustering of samples from each cohort evidenced specific

clusters with low *SMARCA4* knock-down signature scores, high CRPC-NE scores, and low AR signaling scores (red boxes) (**Fig.45A and B**).

Cohort			SMARCA4 knock-down signature		Fisher test p value
			Low	High	
SU2C-PCF	NEPC score	≥ 0.4	19	1	1.40E-05
		< 0.4	63	79	
	AR signaling	≤ 0.25	27	7	0.0001
		> 0.25	55	73	
WCM	NEPC score	≥ 0.4	7	1	0.009
		< 0.4	5	11	
	AR signaling	≤ 0.25	8	2	0.03
		> 0.25	4	10	

Table 9. Correlations between the transcriptomic NEPC and AR signaling scores, and the *SMARCA4* knock-down signature score.

In hormone treatment-naïve PCa, the *SMARCA4* knock-down signature was also correlated with the Gleason grade group categories (ISUP) (Epstein et al., 2015). From 495 patients from The Cancer Genome Atlas (TCGA) cohort (Cancer Genome Atlas Research, 2015), among the 248 patients from the highest and lowest quartiles with respect to the *SMARCA4* knock-down signature, tumors in the highest Grade groups (IV and V) more often displayed low *SMARCA4* knock-down signature scores than tumors in Grade groups I-III ($p < 2.2e-16$, Fisher's test) (**Fig.46A**). Decipher® is a prognostic test that uses the RNA expression of 22 markers for risk stratification after prostatectomy, to predict the risk of metastasis and cancer-specific mortality. In a cohort of 5,239 localized PCa from the Decipher Genomic Resource Information Database (GRID) study (Spratt et al., 2017), samples with a low *SMARCA4* knock-down signature (lowest 10%) were significantly enriched (62%) with high Decipher scores as compared to 14% in samples with high *SMARCA4* knock-down signature (highest 10%) (**Fig.46B**). In this same cohort, *SMARCA4* knock-down signature scores in the

Gleason 9-10 group (mean=-0.13) were significantly lower than in the Gleason 6 group (mean=0.29, $p < 2e-16$) (Fig.46C).

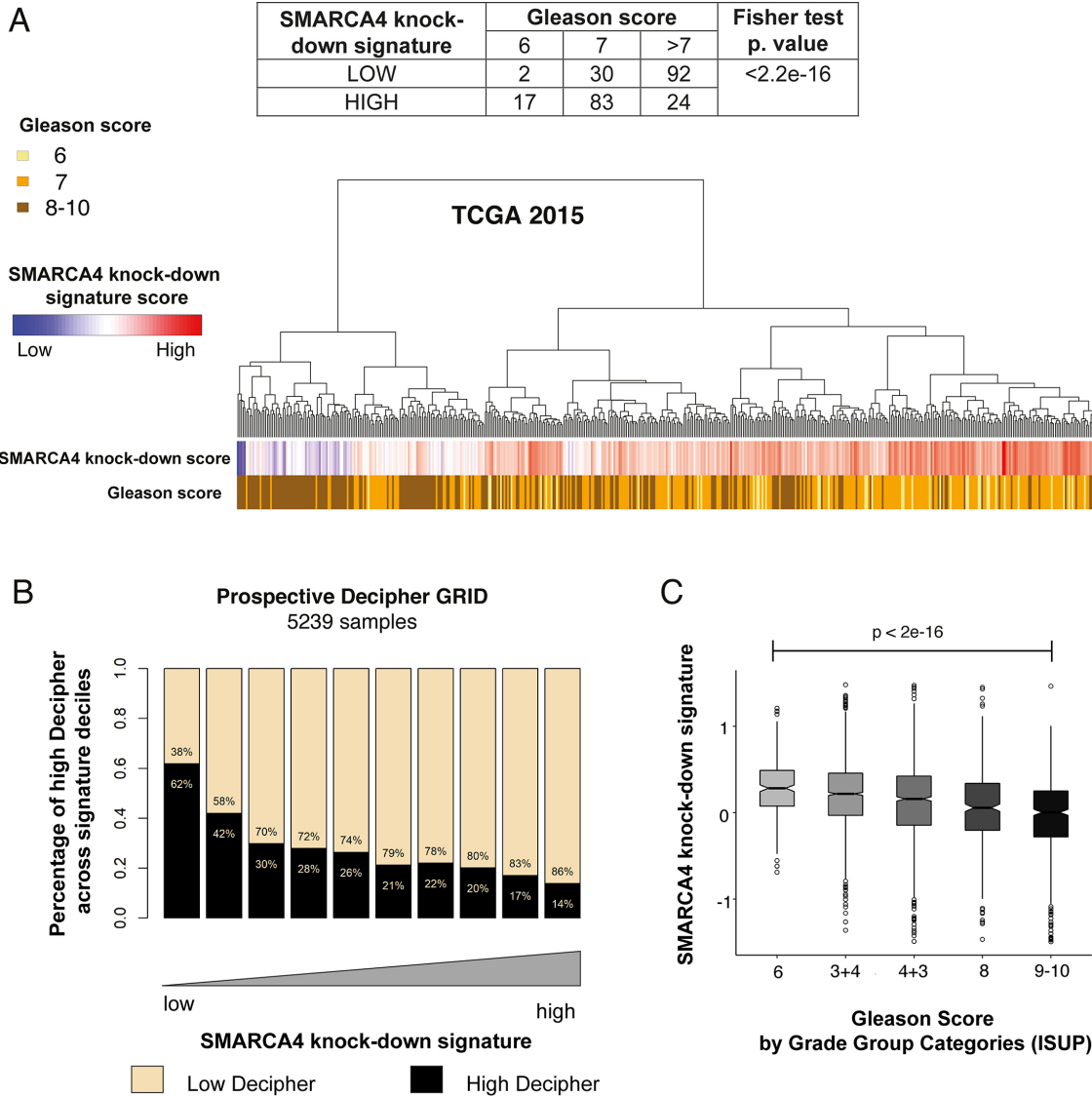


Figure 46. Correlations between prognostic parameters and the *SMARCA4* knock-down signature score in localized hormone treatment-naïve PCa. A. Correlations between the Gleason score *SMARCA4* knock-down signature score in the TCGA cohort (495 cases total, table: 248 cases). **B.** Low *SMARCA4* knock-down signature scores are associated with high Decipher scores (surrogate for risk of metastasis) in 5,239 primary PCa samples from the Prospective Decipher GRID (Mann Whitney U test). **C.** Low *SMARCA4* knock-down signature scores are associated with higher Gleason scores in the same Decipher GRID cohort (Mann-Kendall trend test).

Lastly, we analyzed an independent retrospective cohort of localized PCa cases with clinical outcome data from Johns Hopkins Medical Institution (JHMI), which has been previously described (Ross et al., 2016). A cluster of samples with low *SMARCA4* knock-down scores was enriched in cases with a metastatic outcome, and this cluster showed particularly high expression of proliferation-related genes (**Fig.47**).

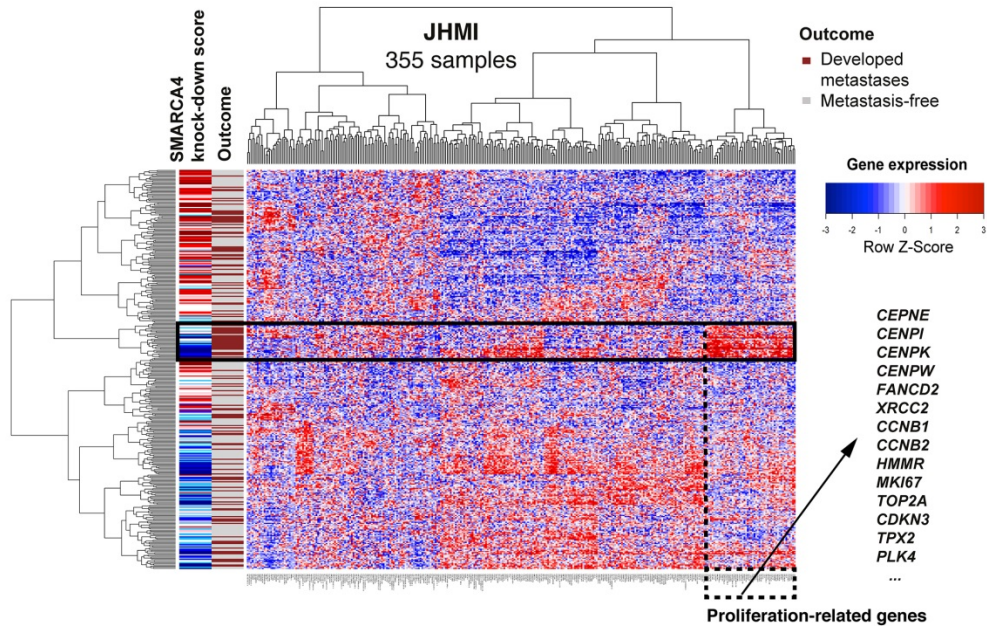


Fig.47. Unsupervised clustering of patients from in the JHMI natural history PCa cohort (Johns Hopkins Medical Institute, n=355) based on genes from the *SMARCA4* knock-down signature (using downregulated genes only), and compared to metastatic outcome (brown: metastatic recurrence, grey: metastasis-free). In particular, overexpression of a subset of genes, many of which are related to proliferation, is seen in a cluster of patients who presented metastatic outcome (black box).

Patients with low *SMARCA4* knock-down signature scores from the JHMI cohort showed a tendency towards worse metastasis-free survival, although this relationship did not reach statistical significance (**Fig.48**).

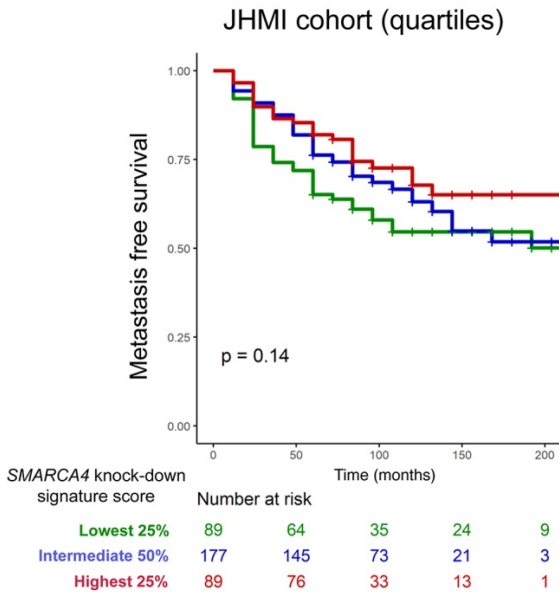


Fig.48 Metastasis-free survival in the JHMI cohort stratified by *SMARCA4* knock-down scores. p refers to the Kaplan-Meier analysis and Cox proportional hazard model comparing groups with lowest (green) and highest (red) *SMARCA4* knock-down signature scores.

2.10 Exploring the role of BAF53B and BAF45B in CRPC-NE

To understand whether BAF53B and BAF45B, the two other subunits overexpressed in CRPC-NE, potentially regulate similar gene expression programs as *SMARCA4*, we performed shRNA-mediated knock-down of BAF53B in the CRPC-NE cell line WCM155 followed by RNA-seq. No significant deregulation of transcriptional programs was observed upon BAF53B knock-down, despite good knock-down efficiency at the protein level. Neither BAF53B nor BAF45B knock-down had an effect on CRPC-NE cell proliferation in WCM155 cells (**Fig.49**).

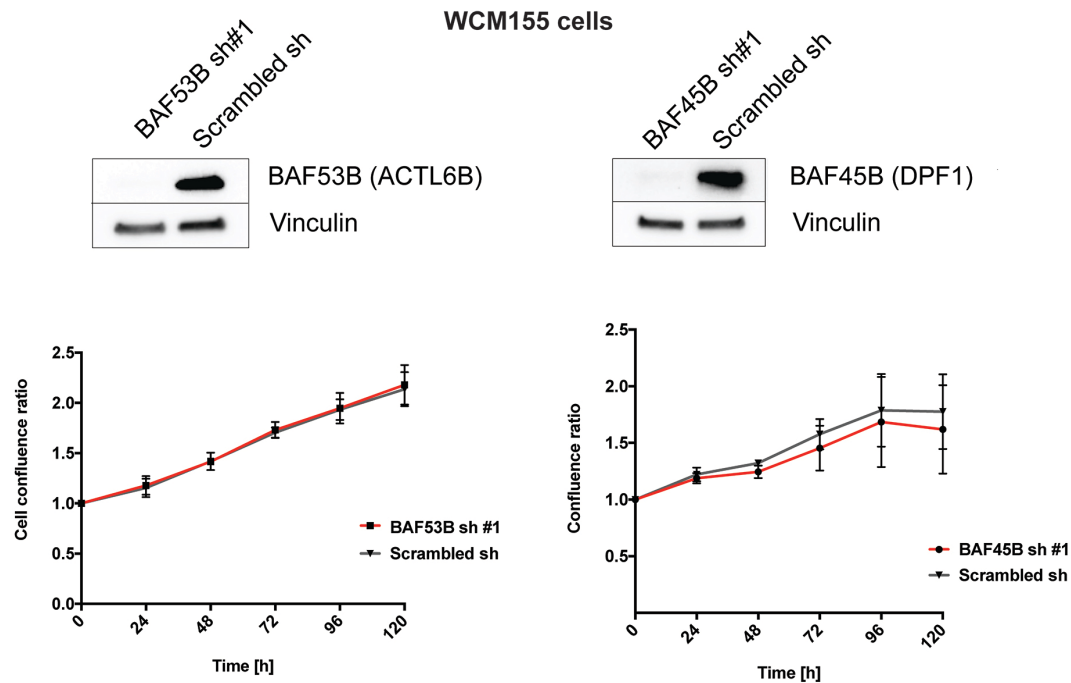


Figure 49. The effects of BAF53B or BAF45B shRNA-mediated knock-down on cell growth of a CRPC-NE cell line (WCM155). The Western blots show knock-down efficiency control (one representative experiment). Each growth curve shows pooled results from three independent experiments (bars: standard error).

These results suggest that BAF53B and/or BAF45B overexpression may be a terminal event in CRPC-NE transdifferentiation, rather than a mediator of this process. For BAF53B, this could be consistent with data from embryonic development, whereby the protein is expressed in post-mitotic, terminally differentiated neurons. In neuron development, BAF53B is de-repressed following downregulation of the RE1-Silencing Transcription factor (REST), a negative regulator of neuron-specific genes (Yoo et al., 2009). Of note, downregulation of REST has also previously been shown to promote a neuroendocrine phenotype in PCa cell lines (Chang et al., 2017; R. Chen, Li, Buttyan, & Dong, 2017; Zhang et al., 2015; Y. Zhu et al., 2014). Concordantly to what has been described in neurons, in the prostatic adenocarcinoma cell line LNCaP, we observed that short-term REST knock-down led to an increase of BAF53B at the protein and mRNA levels, but the effect was modest, whereas other terminal neuronal genes known to be

negatively controlled by REST (e.g. synaptophysin) were highly upregulated (**Fig.50**). No difference in BAF45B expression was observed.

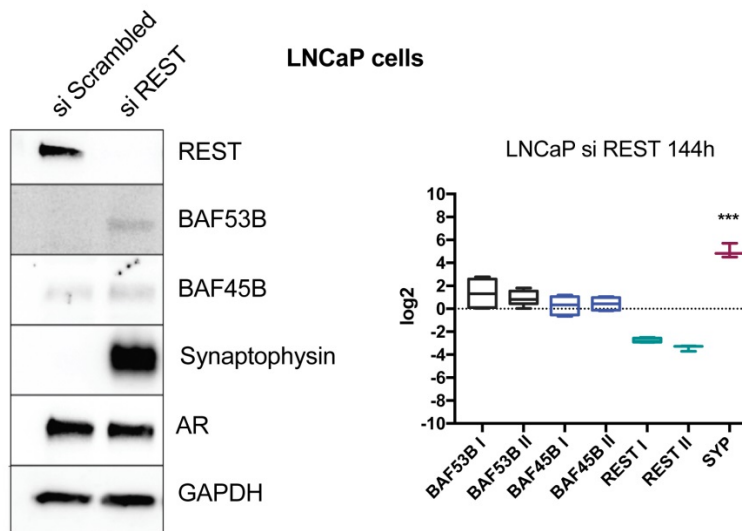


Figure 50. Effects of REST knock-down on the expression of BAF53B and BAF45B in LNCaP cells. Western blot and RT-PCR (4 independent replicates) show strong upregulation of synaptophysin (SYP) upon REST knock-down, a modest increase in BAF53B mRNA, and no significant change in BAF45B mRNA. For RT-PCR, I and II indicate different pairs of RT-PCR primers, data were analyzed using one-way Anova and SYP increase is significant (p value <0.0001).

2.11 SWI/SNF associates with different lineage-specific factors in CRPC-Adeno and in CRPC-NE

We also sought to identify interactors of SWI/SNF in the context of CRPC-NE and prostate adenocarcinoma cell lines. To this end, we performed co-immunoprecipitation (co-IP) with an antibody directed against SMARCA4 in the prostatic adenocarcinoma cell line LNCaP and in the CRPC-NE cell line NCI-H660) followed by mass spectrometry (MS) (**Fig.51**). Proteins that immunoprecipitated with SMARCA4 in CRPC-NE cells, but not in adenocarcinoma cells, included BAF53B, the transcription factor NKX2.1 (TTF-1) related to neural development and the microtubule-associated factor MAP2. Proteins that precipitated with SWI/SNF in adenocarcinoma cells, but not in CRPC-NE cells, included NKX3.1, HOXB13 and REST. BAF45B was detected in both cell lines. Another protein that co-immunoprecipitated with SWI/SNF in both cell lines was CHD4, a member of the NuRD chromatin remodeling complex.

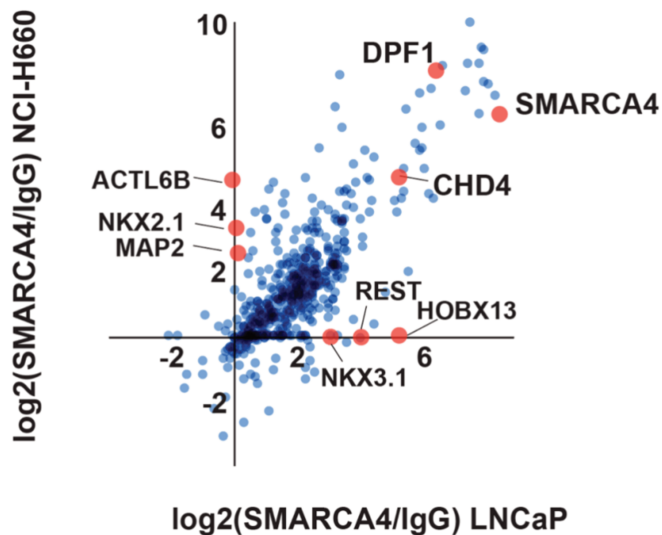


Fig. 51. SWI/SNF associates with different transcriptional regulators in CRPC-NE and in adenocarcinoma cells. A qualitative representation of MS data, comparing proteins associated with SWI/SNF in NCI-H660 (CRPC-NE) and in LNCaP (adenocarcinoma) cells. The plot represents averaged data from two co-IP experiments; \log_2 fold change values of SMARCA4/IgG in LNCaP cells (x-axis) and of SMARCA4/IgG in NCI-H660 (y-axis) are plotted for proteins present in both cell lines with sufficient evidence in each cell line (i.e. present in both replicates in one or both cell lines).

A second co-IP experiment (not shown) followed by MS was performed with an antibody directed against SMARCC1, a core SWI/SNF subunit, in LNCaP-AR cells (LNCaP cells engineered to overexpress the androgen receptor, (Chen et al., 2004)) and in NCI-H660 cells. Proteins that immunoprecipitated with SMARCC1 in CRPC-NE cells, but not in adenocarcinoma cells, included BAF53B, BAF45B and NKX2.1 (TTF-1). Proteins that immunoprecipitated with SMARCC1 in adenocarcinoma cells, but not in CRPC-NE cells, included HOXB13, while NKX3.1 and REST were not detected. CHD4 and another member of the NuRD chromatin remodeling complex, MTA1, were found to immunoprecipitate with SMARCC1 in both cell lines.

While interactions of SWI/SNF with lineage-specific factors in different forms of PCa are intriguing, these experiments do not demonstrate their functional significance. As a proof-of-principle that some of these interactions may indeed be functional, we compared genome occupancy of SMARCC1 with genome occupancy of HOXB13, a homeobox transcription factor involved in AR signaling (Yao et al., 2019), as well as with the location of the active chromatin histone mark H3K27ac and the inactive chromatin mark H3K27me3 in LNCaP cells, using published ChIP-seq datasets. ChIP-seq peaks for *SMARCC1* and *HOXB13* in LNCaP cells were

downloaded from GEO: GS110655 and GSE94682, respectively, and ChIP-seq peaks for H3K27ac and H3K27me3 came from the study: (Sandoval et al., 2018). This analysis showed that SMARCC1 and HOXB13 colocalized at active chromatin sites (11,824 sites), suggesting a potentially functional nature of this interaction, while there was almost no overlap (9 sites only) at inactive sites (**Fig.52**).

Collectively, these results suggest that SWI/SNF associates with different lineage-specific factors in prostatic adenocarcinoma cells and in CPRC-NE cells.

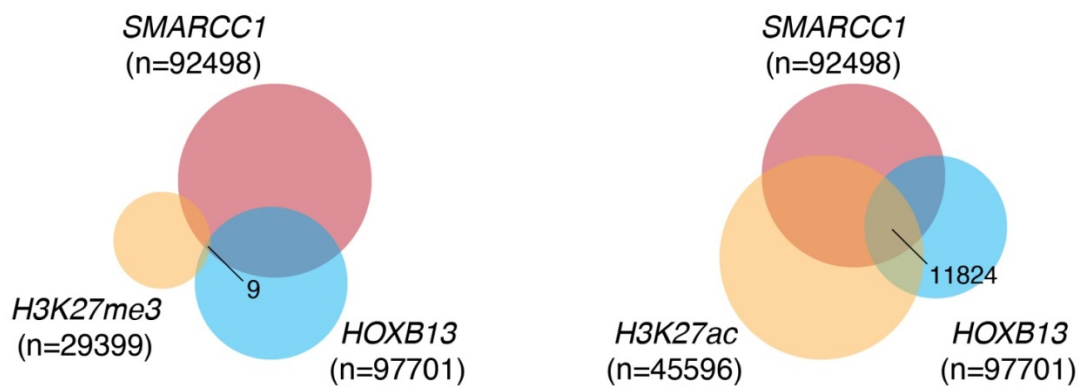


Fig.52. SWI/SNF colocalizes with HOXB13 at active chromatin sites in prostatic adenocarcinoma cells. Venn diagrams illustrating the overlap in genome occupancy sites for SMARCC1, HOXB13, H3K27me3 and H3K27ac in LNCaP cells, assessed by ChIP-seq (analysis performed using published datasets).

2.12 Expression of BAF53B and BAF45B in non-prostatic neuroendocrine tissues and tumors: a follow-up study

Based on our findings in PCa, as a next step, we decided to study BAF53B and BAF45B expression in a wider range of non-prostate neuroendocrine tissues (benign tissues, well-differentiated neuroendocrine tumors and poorly differentiated neuroendocrine carcinomas), in order to better characterize their expression patterns across tissue types and to assess whether they could be useful biomarkers of neuroendocrine differentiation.

Preliminary results showed strong nuclear expression of BAF53B in benign neuroendocrine cells, such as found in the pancreas and the gut (**Fig.53**). Conversely, there

was no nuclear expression of BAF45B in those cells; instead, a heterogeneous cytoplasmic staining was observed.

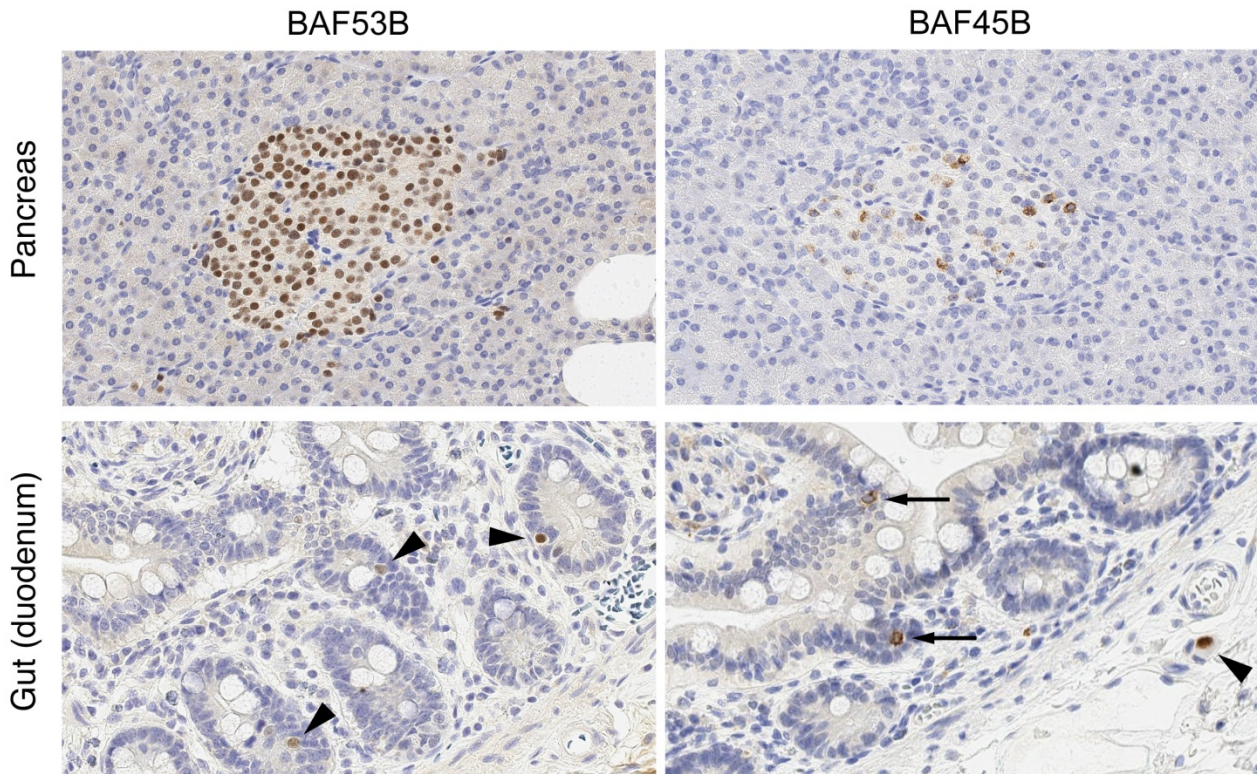


Fig.53. Examples of anti-BAF53B (ACTL6B) and anti-BAF45B (DPF1) IHC in benign tissues. BAF53B shows nuclear positivity in the neuroendocrine pancreas (islets of Langerhans), but not in the exocrine pancreas, and nuclear positivity in scattered cells within intestinal crypts (arrowheads), interpreted as neuroendocrine cells. BAF45B shows cytoplasmic staining in some cells of the neuroendocrine pancreas and in scattered cells within intestinal crypts (arrows), but no nuclear staining is seen. However, ganglion cells within the submucosa do show nuclear staining for BAF45B (arrowhead).

In a pilot cohort of over 150 well-differentiated neuroendocrine tumors (from the lung, gut and pancreas), BAF53B showed strong and diffuse nuclear staining in virtually all cases (**Fig.54**), while it was only focally positive in parathyroid adenomas and negative in cortical adrenal adenomas. BAF45B showed focal nuclear positivity in parathyroid adenomas, pituitary adenomas and a few cases of lung carcinoids, but was negative in other well-differentiated neuroendocrine tumors and in cortical adrenal adenomas.

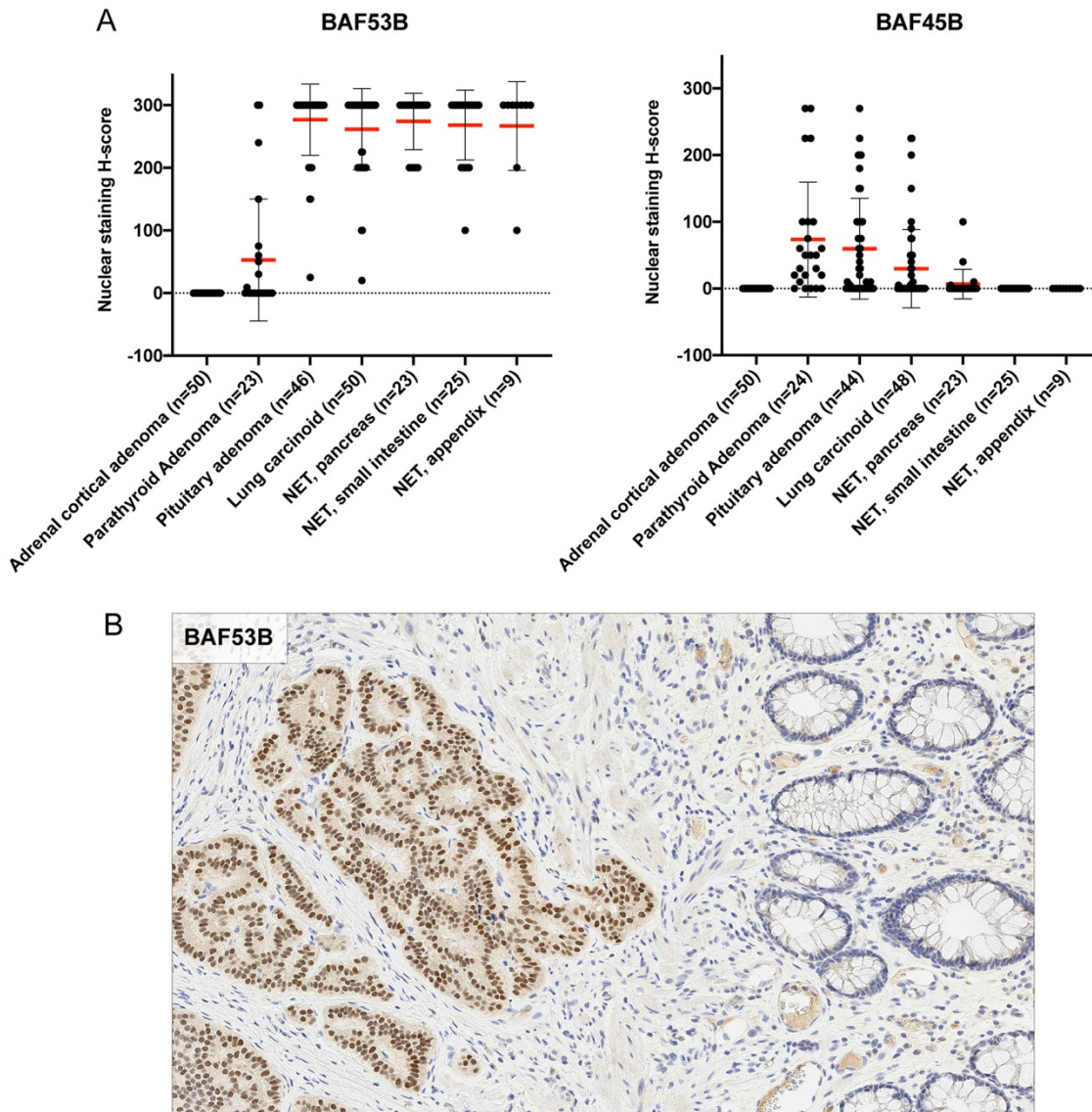


Fig.54. A. Anti-BAF53B (ACTL6B) and anti-BAF45B (DPF1) IHC scoring in various tumor types using TMAs (pilot cohort). The H-score represents the product of the staining intensity multiplied by the percentage of stained cells. Mean and standard deviation are shown. **B.** Example of anti-BAF53B staining in a well-differentiated neuroendocrine tumor of the colon showing diffuse nuclear positivity (left), while benign epithelium (right) remains negative.

Lastly, we tested these IHC in a pilot cohort of 29 poorly differentiated neuroendocrine carcinomas of the lung and 7 non-neuroendocrine carcinomas (2 adenocarcinomas and 5 squamous cell carcinomas) (**Fig.55**). In poorly differentiated neuroendocrine carcinomas,

staining for BAF45B appeared to be more sensitive than staining for BAF53B, contrarily to what we had observed in well-differentiated neuroendocrine tumors. Both stainings appeared to be more sensitive in small cell neuroendocrine carcinomas than large cell neuroendocrine carcinomas. Of note, BAF45B also showed focal nuclear staining in two squamous cell carcinomas, while BAF53B appeared to be more specific of the neuroendocrine phenotype.

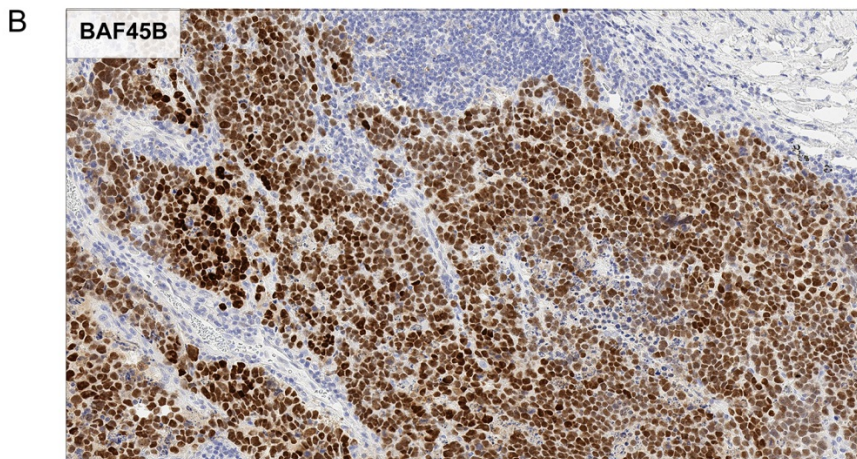
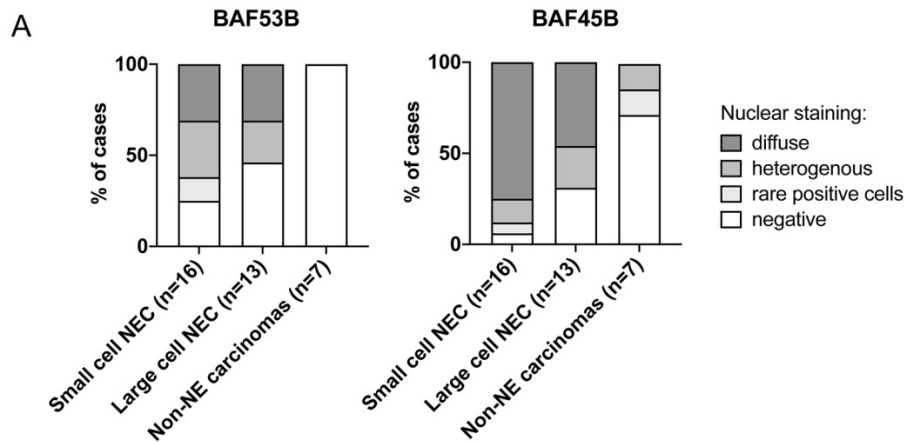


Fig.55. A. Anti-BAF53B (ACTL6B) and anti-BAF45B (DPF1) IHC scoring in poorly differentiated neuroendocrine carcinomas of the lung. **B.** Example of anti-BAF45B staining in a poorly differentiated small cell neuroendocrine carcinoma of the lung.

Taken together, these preliminary observations suggest that BAF53B and BAF45B display different expression patterns across neuroendocrine tissues and tumor types. In

particular, BAF53B seems to be consistently expressed in well-differentiated neuroendocrine tissues and tumors, which could be in line with its role as a “terminal neuronal differentiation” marker, while it is variably expressed in poorly differentiated neuroendocrine carcinomas. BAF45B, on the other hand, seems to show more consistent nuclear expression in poorly differentiated neuroendocrine carcinomas, but is rarely positive in well-differentiated neuroendocrine tumors and tissues. The significance of cytoplasmic staining for BAF45B in such tissues, which was observed in some cases, remains unclear. As a next step, I am planning to analyze BAF53B and BAF45B IHC in a series of ~100 well-differentiated neuroendocrine tumors and ~100 neuroendocrine carcinomas (in collaboration with Pr Jean-Yves Scoazec, Institut Gustave Roussy), to allow for a more systematic exploration of these staining patterns in a wider variety of tumors. Additional studies are warranted to compare the performance of these markers to those routinely used to detect neuroendocrine differentiation (synaptophysin, chromogranin A and INSM1), to assess the relationship between BAF53B and BAF45B expression and the molecular subtype of tumors (such as the molecular subtypes described for small cell neuroendocrine carcinoma of the lung (Rudin et al., 2019)), and to study their potential utility as prognostic markers.

3. Discussion

3.1 A subset of prostate cancers show dependency on SWI/SNF.

Our results show that SWI/SNF genes are rarely altered in prostate cancer (**Fig.25**), in contrast to other cancer types in which SWI/SNF may behave as a tumor suppressor and be inactivated through genomic events (Kadoch et al., 2013; Shain & Pollack, 2013). However, some exceptions should be noted. In an analysis of the “long tail” of genomic alterations in 1,013 cases of prostate cancer (including some overlap with the datasets used in our study), Armenia et al. pointed out that mutations in *ARID1A* and *ARID2* were present in 1.6% and 1.3% of cases, respectively, with roughly half of the variants being truncating. This is consistent with our analysis, as truncating (i.e., nonsense or frameshift) mutations in *ARID1A*, *ARID1B* and *ARID2* were indeed seen in a few cases, although they remained very rare (n=4, 2 and 6 patients, respectively, out of a total of 600 PCa patients). Another particular example is the CRPC-Adeno cell line DU-145, which harbors an inactivating *SMARCA4* splice site mutation (Wong et al., 2000) accompanied by an absence of *SMARCA4* protein expression (**Fig.30**). These examples suggest that some heterogeneity in regards to the role of SWI/SNF may exist among PCa. However, they need to be interpreted in the context of the unique genomic and epigenetic makeup of each tumor - for example, the DU145 cell line harbors a hypermutated profile with microsatellite instability (Wong et al., 2000). The role of these alterations in tumor progression (passenger vs. driver) and of the timing (occurring early in the process of oncogenesis vs. late in tumor progression) also need to be taken into account.

Our results provide further evidence for dependency of some PCa, including its most aggressive forms, on the SWI/SNF complex. Although the role of SWI/SNF in PCa has not been extensively explored, some previous studies support this hypothesis. Inhibition of the SWI/SNF subunits *SMARCE1* or *BAF53A* in PCa cells has been shown to abrogate androgen-dependent cell proliferation (Jin et al., 2018; Link et al., 2008). Similarly, Sandoval et al. reported that

SWI/SNF interacts with ERG in PCa cells harboring the *TMPRSS2:ERG* gene fusion and is required to activate specific gene programs and maintain cell growth (Sandoval et al., 2018).

Most recently, two studies demonstrated that SMARCA4 was required for the growth of prostatic adenocarcinoma cells (Ding et al., 2019; Muthuswami et al., 2019), consistent with our findings. Muthuswami et al. showed that treatment with ADAADi, a biologic inhibitor of SMARCA4, significantly reduced the growth of PC3 cells both *in vitro* and in xenografts (Muthuswami et al., 2019). ADAADi (active DNA-dependent ATPase A Domain inhibitor) is a byproduct of bacterial metabolism by the APH (3')-III enzyme that is responsible for aminoglycoside resistance. Treatment with ADAADi was previously shown to decrease the growth of triple-negative breast cancer cells MDA-MB-231 and, importantly, ADAADi, has been suggested to show preference for inhibition of SMARCA4 over SMARCA2 in *in vitro* experiments (Wu et al., 2016).

Interestingly, Ding et al. proposed a synthetic lethal relationship between PTEN and SMARCA4 in PCa, identified through a CRISPR-Cas9 screen (Ding et al., 2019). They showed that *in vitro*, SMARCA4 knock-down leads to decreased cell proliferation in PTEN-negative cell lines (LNCaP, C4-2 and PC3), in keeping with our results, but not in PTEN-competent cells (22Rv1, BPH-1, and LAPC4). They confirmed these findings in a mouse model of early PCa, by conditionally inactivating *Pten* and *Smarca4* in prostate tissue in mice and studying *in vivo* tumor growth and mouse-derived organoid growth in the context of *Smarca4* loss with and without *Pten* loss.

Findings by Ding et al. do not preclude SMARCA4 as a potential therapeutic target in PCa; indeed, up to 80% of CRPC demonstrate homozygous *PTEN* deletion (Abida et al., 2019b; Aggarwal et al., 2018). However, a synthetically lethal relationship between PTEN and SMARCA4 could explain why SMARCA4 knock-down did not affect cell growth of the CRPC-Adeno cell line 22Rv1 or of the WCM154 CRPC-NE cell lines in our experiments, as these cell lines are PTEN-competent (Puca et al., 2018).

Nevertheless, two points require further attention:

- In the study by Sandoval et al., SMARCA4 knock-down impaired the growth of VCaP cells (Sandoval et al., 2018). However, this PCa cell line has been described to show retained functional PTEN expression (Hodgson et al., 2011). Although this discrepancy is only based on the example of one cell line, it could suggest additional molecular complexity for identifying PCa that are susceptible to SMARCA4 inhibition.
- Even if SMARCA4 inhibition proves to only be effective in PTEN-negative PCa, our results showed that SMARCC1 knock-down effectively impaired the growth not only of PTEN-negative PCa cell lines, but also of two PTEN-competent cells: the CRPC-Adeno cell line 22Rv1 and the CRPC-NE cell line WCM154. This suggests that PTEN wild-type PCa may still be vulnerable to disruption of other SWI/SNF subunits, and warrants further studies of SMARCC1 as a putative therapeutic target.

In contrast to the above studies, and to our results, a paper by Prensner et al. advocates that SWI/SNF acts as a tumor suppressor in PCa (Prensner et al., 2013). In particular, the authors showed a physical interaction between SWI/SNF and the pro-oncogenic long non-coding RNA SChLAP1, and demonstrated an antagonistic relationship between SChLAP1 and the core SWI/SNF subunit SMARCB1 (BAF47). Several observations may help reconcile their result with other above cited studies and with our findings:

- A subsequent study failed to confirm that SChLAP1-SWI/SNF interaction leads to depletion of SWI/SNF from the genome (Raab et al., 2019). It suggested that the SChLAP1-SWI/SNF interaction is likely to represent a non-specific interaction between SWI/SNF and transcribing RNA, rather than a functionally meaningful relationship, and that SChLAP1 may function in a SWI/SNF-independent manner.
- Prensner et al. propose that SChLAP1 and SWI/SNF regulate gene expression in an opposing manner. This interpretation is based, in part, on Gene Set Enrichment Analyses

(GSEA) (Supplementary Fig.10 from their study), whereby they compare LNCaP and 22Rv1 cells treated with SChLAP1 siRNA to gene sets related to SMARCA2 targets, and they show inverse correlations between the SChLAP1 and SMARCA2 signatures. However, based on our results from the present study, SMARCA2 knock-down signatures may not be a reliable surrogate for the function of SWI/SNF as such in a given cell model. It is even theoretically possible that SMARCA2 could define a specific subset of complexes with tumor-suppressive activities, as opposed to tumor-promoting complexes incorporating SMARCA4. Thus, the fact that depletion of SChLAP1 and depletion of SMARCA2 show opposite effects, is not sufficient to demonstrate that all forms of SWI/SNF play tumor-suppressive roles in PCa.

- In our GSEA analysis comparing RNA-seq data in LNCaP cells upon knock-down of SMARCA4 or SMARCA2 to transcriptomic data from Prensner et al. upon knock-down of SChLAP1 (**Fig.39**), we observed an inverse relationship between the effects of SMARCA2 knock-down and the effects of SChLAP1 knock-down. Conversely, although the relationship between SMARCA4 knock-down and SChLAP1 knock-down did not reach statistical significance, there was a tendency towards a positive relationship between genes upregulated upon SMARCA4 knock-down and genes upregulated upon SChLAP1 knock-down. This further suggests that SChLAP1 may have a different functional relationship with different forms of SWI/SNF complexes, depending on their composition.
- Prensner et al. also observed an inverse correlation between SChLAP1 and SMARCB1 (BAF47) by GSEA. This result is more difficult to explain, as the effects of SMARCB1 knock-down have not been studied in PCa cells neither in studies published to date, nor in our study.

Taken together, our and the previously published findings support the hypothesis that PCa may expand the spectrum of cancer types that are dependent on SWI/SNF activity for their survival, growth and/or progression. This provides a rationale for further exploring selected SWI/SNF subunits as putative therapeutic targets in CRPC-NE, given that effective therapeutic approaches for ARSi-resistant forms of CRPC remain an unmet clinical need.

3.2 High SMARCA4 expression correlates with aggressive PCa.

We found that higher SMARCA4 expression by IHC correlated with worse overall survival in patients with localized PCa. We also observed a positive correlation between a low transcriptomic SMARCA4 knock-down signature score and features of poor prognosis in PCa, including neuroendocrine differentiation, a high Gleason score, and a high Decipher score.

Only a few studies have investigated the link between expression of SWI/SNF subunits and patient outcome in PCa. A study from 2008 found that SMARCC1 expression by IHC was higher in PCa than in benign prostatic tissue, and that high SMARCC1 expression was positively correlated with the Gleason score, clinical T stage and risk of recurrence in univariate analysis (Heeboll et al., 2008). Similar to our current study, Ding et al. found that high SMARCA4 expression by IHC was associated with a shorter time of biochemical recurrence, as determined by Prostate Specific Antigen (PSA) levels, in a population of Asian men with clinically localized PCa treated with surgery, and that this association was most relevant in cases with PTEN loss (Ding et al., 2019). Accordingly, Muthuswami et al. analyzed SMARCA4 mRNA levels in PCa patients from the TCGA cohort, and showed that the 25% of patients with highest SMARCA4 expression (highest quartile) had a significantly worse disease-free survival than the 25% with lowest SMARCA4 expression (lowest quartile) (HR=2.4, p=0.0028, log-rank test)(Muthuswami et al., 2019). The authors also observed a positive correlation between SMARCA4 mRNA levels and the Gleason scores in this cohort. No significant association between SMARCA2 mRNA expression and survival was found. Taken together, our results and these data align to suggest that high SMARCA4 expression is associated with aggressive PCa.

While the transcriptomic SMARCA4 knock-down signature, derived from *in vitro* experiments in LNCaP cells, could represent an additional argument in favor of this hypothesis, an important potential bias needs to be mentioned. Indeed, we saw that SMARCA4 knock-down significantly impacts cell proliferation and/or viability in LNCaP cells. Thus, the SMARCA4 knock-down signature may contain genes regulated by SMARCA4 (directly or indirectly), but

may also contain various genes related to cell proliferation and not regulated by SMARCA4. As such, given that high proliferation is more likely to be observed in aggressive forms of PCa, it is possible that the correlations we have observed between the SMARCA4 knock-down signature and parameters of aggressive PCa are mainly driven by differences in proliferation-related genes. Indeed, in the JHMI cohort, a cluster of samples showing low SMARCA4 knock-down scores and enriched in cases with a metastatic outcome (**Fig.47**) appeared to be mainly driven by high expression of proliferation-related genes. In conclusion, while results obtained using the SMARCA4 knock-down signature are consistent with the overexpression of SMARCA4 in aggressive PCa, they should be interpreted with care because of this potential bias.

3.3 SMARCA4 and SMARCA2 are not entirely redundant in PCa.

In the current study, we also demonstrate that the proportions of several SWI/SNF subunits in PCa are susceptible to change between disease states, and even between distinct sub-populations within the same tumor.

SMARCA4 and its paralog SMARCA2 show 86% similarity at the protein level (Muchardt & Yaniv, 2001). Several studies have suggested that synthetic lethality approaches could be effective in SMARCA4-null or SMARCA2-null tumors, by inhibiting the remaining catalytic subunit; this could be in line with the hypothesis functional redundancy between SMARCA4 and SMARCA2, although it does not prove it (Hoffman et al., 2014).

Among studies that demonstrated a synthetic lethal relationship between SMARCA4 and SMARCA2, Oike et al. showed that *SMARCA4*-mutated non-small cell lung carcinoma cell lines were sensitive to SMARCA2 depletion, and that this phenotype could be rescued by re-expression of functional SMARCA4, but not of its ATP-ase deficient form (Oike et al., 2013). Wilson et al. further demonstrated that *SMARCA4*-mutant cancer cells lines contain residual SWI/SNF complexes, and that they are sensitive to depletion of the core subunit SMARCB1 (Wilson et al., 2014). In a cancer vulnerability screen, *SMARCA2* was the top essential gene in

SMARCA4-mutant cancer cell lines as compared to *SMARCA4*-wild type cancer cell lines, and *SMARCA2* depletion in such cell lines led to a significant decrease of cell proliferation (Wilson et al., 2014). Similarly, knock-down of *SMARCA2* in *SMARCA4*-deficient melanoma cells inhibits their tumorigenicity *in vitro* (Keenen et al., 2010). Yet a different study using the small molecule assisted shut-off (SMASh) degron system, which allows to fine-tune the levels of depletion of the targeted protein, showed that a near-complete depletion of *SMARCA2* was necessary to impair cell growth of *SMARCA4*-mutated lung carcinoma cell lines (Rago et al., 2019). In squamous esophageal carcinoma cell lines, sensitivity to *SMARCA4* depletion was linked to low or absent expression of the *SMARCA2* paralog (Ehrenhofer-Wolfer et al., 2019).

However, there are also studies suggesting functional differences between the two subunits. *SMARCA4* and *SMARCA2* expression levels are known to vary throughout embryonic development: *SMARCA4* is the dedicated ATP-ase in the esBAF complex found in mouse embryonic stem cells, while *SMARCA2* is not encountered in this form of the complex (Ho, Jothi, et al., 2009; Ho et al., 2011; Ho, Ronan, et al., 2009; Yan et al., 2008). *SMARCA2* expression levels have been shown to increase with differentiation stage of the mouse embryo and of embryonic stem cells (LeGouy et al., 1998; Yan et al., 2008). In porcine embryos, overexpression of *SMARCA4* and *SMARCA2* at various stages of embryonic development had different effects on the expression levels of transcription factors and on embryo development (Magnani & Cabot, 2009). Results in knock-out mice are also often cited to point to differences between the two catalytic subunits: while homozygous *Smarca4* inactivation is lethal for mouse embryos (Bultman et al., 2000), mice with homozygous *Smarca2* inactivation develop normally and are fertile; however, the *Smarca2* knock-out mouse model has recently been questioned, as it may not represent a complete knockout (Thompson et al., 2015).

Functional differences between the two SWI/SNF ATP-ases have also been noted in cancer. *SMARCA2* mutations are much less frequent than *SMARCA4* mutations. An inverse expression pattern between *SMARCA4* and *SMARCA2*, whereby high *SMARCA4* expression is

associated with worse prognosis and high SMARCA2 expression tends to be associated with better prognosis, has been highlighted in several studies. In a recent computational meta-analysis, SMARCA4 was overexpressed, and SMARCA2 was underexpressed, in multiple tumor types as compared to corresponding benign tissues (Guerrero-Martinez & Reyes, 2018). Moreover, high SMARCA4 expression and low SMARCA2 expression was associated with worse prognosis in various tumor types (univariate analysis); high SMARCA4 expression was also positively correlated with higher histologic grade, while high SMARCA2 expression was associated with lower histologic grade, in renal cell and liver carcinomas. In the prostate, SMARCA4 and SMARCA2 also showed a reciprocal expression pattern by IHC as compared to benign tissue (Sun et al., 2017). Another study also showed that SMARCA2 expression by IHC was lower in PCa than in benign tissue, and inversely correlated with the Ki67 index (Shen et al., 2008). In addition, the authors used the *Smarca2*^{-/-} mouse model to show that homozygous knock-out induced lobe-specific epithelial hyperplasia in the prostate.

Our results in PCa, including its most aggressive forms, and in pilot examples of non-prostate tissues (**Fig.29**), further support this inverse relationship between SMARCA4 and SMARCA2 expression levels. In addition, we showed that SMARCA4 and SMARCA2 depletion have different effects on the transcriptome and on cell growth in PCa cells, whereby SMARCA4 knock-down has a potent effect on cell proliferation and on gene expression programs, while SMARCA2 knock-down does not affect cell growth and has less effect on the transcriptome (**Fig.36**).

Overall, these observations suggest that SMARCA4 and SMARCA2 are not redundant in a variety of cellular contexts. Two main hypotheses can be considered at this stage to explain the observed preference for a SMARCA4-high and SMARCA2-low expression profiles in cancer progression:

- Hypothesis 1: SMARCA4 and SMARCA2 show largely overlapping functions, but SMARCA4

is the dominant subunit. In this scenario, in tumors with high SMARCA4 expression, SMARCA2 downregulation could represent a secondary event, possibly related to lower incorporation of SMARCA2 in the complex (via competition with SMARCA4) and increased degradation of the unbound subunit. Conversely, in knock-down experiments, residual SMARCA2 could be insufficient to replace SMARCA4 in its functions in the event of SMARCA4 depletion (hence, the SMARCA4 knock-down phenotype), while in case of SMARCA2 depletion, SMARCA4 could effectively replace SMARCA2 (hence, the absence of effect of SMARCA2 knock-down on cell proliferation and the modest effects on the transcriptome).

- Hypothesis 2: functional differences exist between SMARCA4 and SMARCA2, and are responsible for their differential involvement in oncogenesis. As such, SMARCA4 could have tumor-promoting functions, while SMARCA2 could behave in a tumor-suppressive manner. Both SMARCA4 overexpression and SMARCA2 downregulation (whether it's a primary or a secondary event) could contribute to disease progression in this model.

In support of the first hypothesis, several studies suggest that SMARCA4 may play a dominant role in cell functions. In ChIP-seq experiments in the mouse erythroleukemia model, the majority of promoters were shown to interact with SMARCA4 rather than SMARCA2, and most transcription factors were associated with SMARCA4-containing complexes (Kadam & Emerson, 2003). In addition, SMARCA4 depletion leads to increased incorporation of the SMARCA2 subunit into the complex, and vice versa (Reyes et al., 1998; Wilson et al., 2014), and this is also consistent with our Western blot results (**Fig.42**).

In support of the second hypothesis, several examples suggest the presence of functional differences between the two subunits. *In vitro* protein binding experiments and experiments in a mouse erythroleukemia model showed that different protein families displayed a preferential specificity for SMARCA4 or SMARCA2, but not both; for example, zinc finger proteins were shown to function specifically with SMARCA4, while two components of the Notch

pathway (CBF-1 and ICD22) interacted exclusively with SMARCA2 (Kadam & Emerson, 2003). SMARCA4- and SMARCA2-containing complexes also showed distinct chromatin occupancy patterns throughout cell proliferation and differentiation in this model. Another study showed that in SMARCA4-null and SMARCA2-null cells (the SW13 cell line), AR-dependent activation of the PSA (*KLK3*) promoter was only induced by restoration of SMARCA2 expression, but not by SMARCA4 re-expression (Marshall et al., 2003), which could turn out to be relevant for PCa biology. In liver cancer stem cells, the long non-coding RNA *IncBRM* was shown to specifically sequester SMARCA2 and to prevent it from incorporating into the complex, thus leading to an increased incorporation of SMARCA4 (P. Zhu et al., 2016) and allowing for activation of specific signaling by SMARCA4-containing complexes, such as the YAP1 pathway implicated in cancer stem cell self-renewal.

While our study was not designed to formally distinguish between those two hypotheses, two observations support the existence of functional differences between SMARCA4 and SMARCA2. First, transcriptomic changes observed upon SMARCA4 and SMARCA2 knock-down did not overlap, but instead, demonstrated opposite trends in LNCaP cells for some gene sets in the GSEA analysis (**Fig.38-39**). Second, we observed that the combined knock-down of SMARCA4 and SMARCA2 did not have a synergistic effect on cell growth (**Fig.40**).

3.4 BAF53B and BAF45B expression in CRPC-NE: interpretation of results and follow-up studies

In neuronal development, cells committed to the neural lineage initially express a neural progenitor form of the SWI/SNF complex (termed npBAF), which incorporates among others the BAF53A, BAF45a/d and SS18 subunits (Lessard et al., 2007; J. I. Wu et al., 2007; Yoo et al., 2009). However, upon differentiation to post-mitotic neurons, the complex undergoes a switch to the neural variant of SWI/SNF (nBAF) and incorporates the respective paralogs of those three

subunits (i.e., BAF53B, BAF45B/C and SS18L1). This switch is mediated by repression of BAF53A by micro-RNAs in response to a downregulation of REST (Yoo et al., 2009).

Here, we report for the first time the expression of “neuronal” SWI/SNF subunits, BAF53B and BAF45B, outside of the central nervous system. Although their expression appeared to be specific of CRPC-NE, it remains unclear whether they play any role in activating neural-like gene programs or are simply expressed as a consequence of this process.

For BAF53B, we hypothesize that expression in CRPC-NE is rather a result of de-repression of neural-like gene expression programs, than a causal mechanism of CRPC-NE differentiation. Indeed, our transcriptomic analysis revealed only 20 significantly deregulated genes between the knock-down condition and the control samples (shRNA-mediated knock-down of BAF53B in WCM155 CRPC-NE cells), despite a good knock-down efficiency. In addition, BAF53B knock-down did not have an effect on CRPC-NE cell growth. Lastly, this scenario would be consistent with current knowledge about the BAF53B protein, which is expressed in post-mitotic neurons as part of the terminal differentiation process (Lessard et al., 2007; J. I. Wu et al., 2007; Yoo et al., 2009). This is also consistent with our results in PCa cell lines, where BAF53B expression always accompanied synaptophysin expression (**Fig.30**), and with the role of REST depletion on BAF53B de-repression (**Fig.50**).

Conversely, the role of BAF45B in CRPC-NE, and possibly in other tumor types, may warrant further investigation. Although BAF45B knock-down did not inhibit the growth of CRPC-NE cells, only one *in vitro* model (WCM155) has been tested so far, and the RNA-seq experiment was not conclusive because of technical problems. In addition, we saw that the expression pattern of BAF45B expression across PCa cell lines was different than the pattern of BAF53B: BAF45B expression was not limited to synaptophysin-positive cell lines, but instead, some level of expression was observed in CRPC-Adeno cells, such as PC3, DU145 and MSKCC-PCa3. Interestingly, in glioblastoma, it was shown that BAF45B (DPF1) was necessary to maintain “stemness” and tumor-initiating properties of glioma-initiating cells, and that

BAF45B knock-down improved survival in an orthotopic xenograft model in the mouse (Hiramatsu et al., 2017). Overall, our data in CRPC-NE are likely insufficient to affirm that BAF45B does not have potential tumor-promoting functions in neuroendocrine carcinomas.

3.5 The role of SWI/SNF in PCa phenotype plasticity: future directions

There is increasing evidence that CRPC-NE arises from CRPC-Adeno through a mechanism of lineage plasticity, rather than through selection of a sub-population of neuroendocrine cells. Lineage tracing used in a genetically engineered mouse model (GEMM) of PCa with combined *Trp53* and *Pten* loss provided evidence that neuroendocrine tumor cells can directly arise from pre-existing luminal adenocarcinoma cells and do not emerge from a second, independent population of neuroendocrine or intermediate cells (Zou et al., 2017). Patient-derived PCa xenografts can acquire neuroendocrine features following castration, but they retain genomic relatedness to the pre-existing adenocarcinoma (D. Lin et al., 2014). Patient CRPC-NE tumors are characterized by an overexpression of several epigenetic regulators, such as EZH2, and by a specific DNA methylation profile, as shown by our team and others (**Appendix II**) (Beltran et al., 2016; Davies et al., 2018; Puca et al., 2018). Of note, lineage plasticity induced in mouse PCa models upon *Pten*, *Trp53* and *Rb1* loss is reversible, as epigenetic therapy with EZH2 inhibitors can re-sensitize those tumors to ARSi (Ku et al., 2017). Overall, these data support the idea that PCa progression through lineage plasticity is regulated by both genomic and epigenetic changes (Mu et al., 2017; Zehir et al., 2017).

The “neuroendocrine phenotype” is not limited to the expression of terminal neuronal markers (e.g. synaptophysin, chromogranin...), but needs to include other critical characteristics of CRPC-NE, such as dedifferentiation, acquiring “stem cell”-like features (with the expression of stem cell factors such as SOX2), AR signaling indifference, and/or high proliferation (Davies et al., 2018). We hypothesize that high SMARCA4 expression observed in CRPC-NE is more related to stem cell-like features and/or to proliferation, and not to the expression of terminal

neuronal markers. In addition, it was previously shown in non-neuronal benign cells that in fact, SMARCA4 actually collaborates with REST and enables the repression of neuronal gene programs by REST in non-neural cells (Battaglioli et al., 2002; Ooi et al., 2006). This is in keeping with **Fig.34**, where clusters of CRPC-NE organoid cells with high expression of SOX2, a neural stem cell marker, show higher expression of SMARCA4 and SMARCC1, and lower expression of terminal neural markers (e.g. SYP). An interesting future direction would be to further analyze the correlation between SWI/SNF subunits and gene expression programs at the single cell level, using cell sorting or single cell RNA-seq.

One important limitation of this study is that we do not report gene expression profiling in a CRPC-NE cell line upon SMARCA4 knock-down. This was due to the rarity of available CRPC-NE cell lines and to difficulties in manipulating their gene expression levels. Nevertheless, such experiments are currently ongoing in Prof Rubin's laboratory and will complement the data presented herein.

Given that SWI/SNF is a major epigenetic regulator in embryonic development and in tissue differentiation, and that selected SWI/SNF subunits are differentially expressed in CRPC-NE as shown in our study, it would also be important to better understand whether SWI/SNF actually plays a role in PCa phenotype switching. During the course of this study, we considered using two possible *in vitro* models to address this question. Unfortunately, neither of these models provided a satisfactory rendering of neuroendocrine PCa transdifferentiation in our pilot experiments:

1) Knock-down of the neuronal repressor REST in prostatic adenocarcinoma cells has previously been proposed as a model of neuroendocrine PCa transdifferentiation (Chang et al., 2017; R. Chen et al., 2017; Zhang et al., 2015; Y. Zhu et al., 2014). We performed siRNA-mediated knock-down of REST in LNCaP-AR cells and profiled them by RNA-seq, compared to cells transfected with non-targeting siRNA (**Fig.56**). Upon REST knock-down, there was a significant upregulation of gene expression programs related to neuronal functions and synaptic

signaling. When considering specifically only genes from the 70-gene transcriptomic signature known as the “NEPC score” (Beltran et al., 2016 (Beltran et al., 2016)), only a few were upregulated upon REST knock-down, in particular those related to terminal neuronal differentiation (e.g., Secretogranin 3, SCG3). Concordantly, recent work by Labrecque et al. confirmed that REST knock-down leads to the expression of neuronal genes in PCa cells, but does not decrease AR expression or the AR transcriptomic score, nor does it increase the expression of transcription factors associated with small cell neuroendocrine carcinoma, such as NKX2-1, POU3F2 (BRN2) or SOX2 (Labrecque et al., 2019). Collectively, these findings suggest that REST knock-down allows for de-repression of “terminal” neuronal genes, but does not provide a satisfactory model of other key elements of neuroendocrine PCa plasticity. Nonetheless, a recent study by Chang et al. suggests that prolonged REST depletion could confer “stemness” features to PCa cells in addition to neural-like differentiation (Chang et al., 2017).

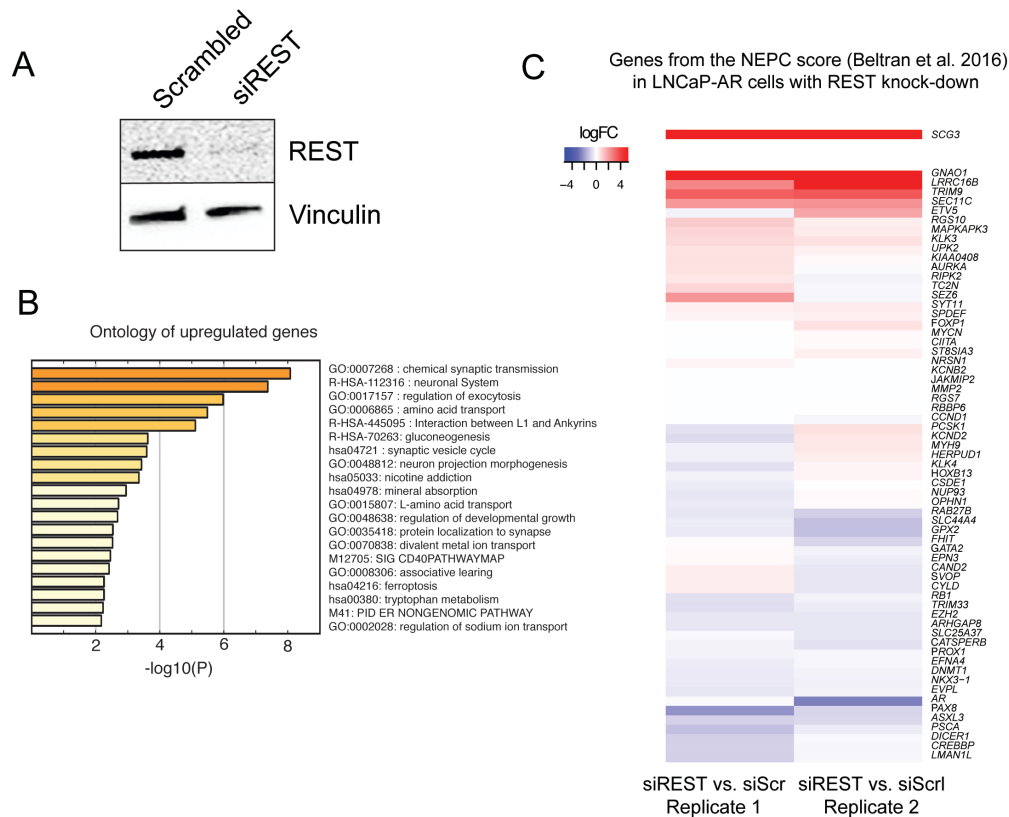


Figure 56. Transcriptomic changes in prostatic adenocarcinoma cells (LNCaP-AR) upon REST knock-down.

2) The inducible *TP53/RB1* knock-down model consists of LNCaP-AR cells stably transduced with Doxycyclin-inducible shRNA vectors against *TP53* and *RB1* (published by, and a kind gift from, Drs. Ping Mu and Charles Sawyers, MSKCC (Mu et al., 2017)). The authors showed that induction of concomitant *TP53/RB1* knock-down induced the expression of the neural stem cell transcription factor *SOX2*, of neural markers (*SYP*, *CHGA*) and of basal markers (*CK5*, *CK14*, *P63*), and decreased the expression of luminal markers (*AR*, *CK8*, *CK18*).

In our hands, short-term induction with Doxycyclin (96h) allowed to observe an increase in synaptophysin expression and a decrease of *AR* expression at the protein level, as expected (**Fig.57**). This was also accompanied by a slight increase of some SWI/SNF proteins, including *SMARCA4*, *SMARCC1* and *BAF45B*. *BAF53B* expression was not induced (not shown). However, due to several technical issues, we could not pursue experiments with this model:

- the results of Doxycyclin induction on the expression levels of neuroendocrine markers, AR and SWI/SNF subunits were highly variable between experiments, despite consistently good knock-down efficiency of *TP53* and *RB1*. This suggests that other factors present at the moment of induction (such as cell proliferation dynamics, epigenetic makeup...) may influence the process of phenotype plasticity.

- the observed phenotype switch appeared to be incomplete and transient. Long-term experiments (not shown) attenuated the results instead of enhancing them. This is also in line with our observations in LNCaP cells with a stable CRISPR-Cas9-induced knock-out of *TP53* and *RB1* (used for growth experiments in **Fig.41**), which do not display a neuroendocrine phenotype. Thus, while depletion of *TP53* and *RB1* may transiently “destabilize” the cell phenotype (either in a specific manner, or because of “cellular stress” linked to the sudden loss of these proteins), and represents an interesting proof of concept for the phenomenon of PCa phenotype plasticity, it may be insufficient to achieve complete neuroendocrine transdifferentiation.

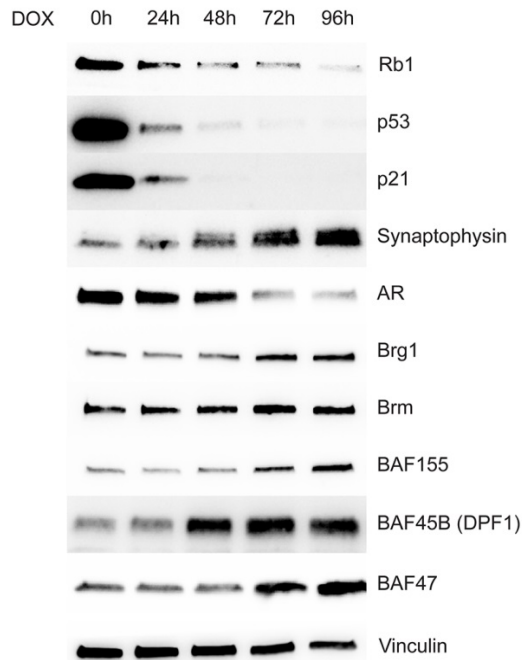


Figure 57. Western blot showing the effects of Doxycyclin-induced shRNA-mediated *TP53* and *RB1* knock-down in LNCaP-AR cells at different time points.

Taken together, reliable *in vitro* models that would allow to recapitulate all stages of neuroendocrine PCa transdifferentiation are still lacking. The potential role of SWI/SNF in PCa phenotype plasticity remains to be addressed *in vitro* models and potentially by using the above cited mouse models.

4. Methods

Genomic analysis

Matched tumor and normal WES data of localized and advanced prostate cancer from The Cancer Genome Atlas (Cancer Genome Atlas Research, 2015), SU2C-PCF (Abida et al., 2019b) and from the Weill Cornell Medicine cohort (Beltran et al., 2016) were uniformly analyzed for somatic copy number aberrations (SCNA) with CNVkit (Talevich, Shain, Botton, & Bastian, 2016), and for single nucleotide variations (SNVs) and indels with MuTect2 (Cibulskis et al., 2013). SNVs and Indels were annotated with variant effect predictor (VEP)(McLaren et al., 2016) and only mutations with HIGH or MODERATE predicted impact on a transcript or protein were retained

[https://www.ensembl.org/info/genome/variation/prediction/predicted_data.html]. All samples with tumor ploidy and purity estimated using CLONET(Prandi et al., 2014) were retained in the analyses and processed for allele specific characterization. The integrated dataset includes 299 unique hormone treatment-naïve prostatic adenocarcinoma (Adeno), 245 castration resistant prostate adenocarcinoma (CRPC-Adeno), and 56 castration resistant neuroendocrine prostate carcinoma (CRPC-NE) patients. Two-tailed proportion test has been used to check enrichment of heterozygous deletion and copy number neutral loss.

RNA-seq data analysis of human samples

RNA-seq data from 32 normal prostate samples (Beltran et al., 2011; Chakravarty et al., 2014), 400 localized PCa (Beltran et al., 2011; Cancer Genome Atlas Research, 2015;

Chakravarty et al., 2014) and 120 CRPC-Adenos and 20 CRPC-NE patients (Abida et al., 2019b; Beltran et al., 2016) were utilized for the initial investigation of the SWI-SNF complex units levels and were processed as follows. Reads (FASTQ files) were mapped to the human genome reference sequence (hg19/GRC37) using STAR v2.3.0e(Alexander Dobin et al., 2012), and the resulting alignment files were converted into Mapped Read Format (MRF) for gene expression quantification using RSEQtools(Habegger et al., 2010) and GENCODE v19 (<http://www.gencodegenes.org/releases/19.html>) as reference gene annotation set. A composite model of genes based on the union of all exonic regions from all gene transcripts was used, resulting in a set of 20,345 protein-coding genes. Normalized expression levels were estimated as FPKM. After converting the FPKM via $\log_2(\text{FPKM} + 1)$, differential expression analysis was performed using Mann-Whitney Wilcoxon test. RNA-seq data of the SU2C-PCF cohort were downloaded from original study (Abida et al., 2019b). NEPC score and AR signaling score were inferred as previously described (Abida et al., 2019b). Gleason scores of the TCGA PCas were retrieved from the original study(Cancer Genome Atlas Research, 2015). RNA-seq data and Gleason score from the TCGA PCa dataset were retrieved from the TCGA data portal using TCGAbiolinks R package(Colaprico et al., 2016).

Immunohistochemistry

Immunohistochemistry (IHC) was performed on sections of formalin-fixed paraffin-embedded patient tissue (FFPE) using a Bond III automated immunostainer and the Bond Polymer Refine detection system (Leica Microsystems, IL, USA). Slides were de-paraffinized and heat-mediated antigen retrieval using the Bond Epitope Retrieval 1 solution at pH6 (H1) or Bond Epitope Retrieval 2 solution at pH9 (H2) or enzyme-mediated antigen retrieval (E1) was performed. All antibodies, dilutions and conditions used are shown in **Table 10**.

Protein name	Antibody information			Immunoblotting	Immunohistochemistry		
	Company	Clone name	Catalogue number	Dilution	Dilution	Retrieval solution (pH)	Retrieval time
AR	Abcam	ER179(2)	ab108341	1/10000	-	-	-

BAF155	Abcam	EPR12395	ab172638	1/5000	1/300	H1 (pH6)	30 min
BAF170	Cell Signaling Technology	D8O9V	12760	1/10000	1/300	H1 (pH6)	30 min
BAF45B	Atlas Antibodies	polyclonal	HPA049148	1/1000	1/100 with casein	H2 (pH9)	40 min
BAF47 (INI-1)	BD Biosciences		bd612110	-	1/100	H2 (pH9)	30 min
BAF47 (INI-1)	Abcam	EPR12014	ab181976	1/5000	-	-	-
BAF53A	Abcam	EPR7443	ab131272	1/2000	-	-	-
BAF53B	Abcam	EP10101	ab180927	1/1000	1/50	H2 (pH9)	20 min
Brg1	Abcam	EPR3912	ab108318	1/1000	1/50	H2 (pH9)	60 min
Brm	Cell Signaling Technology	D9E8B	11966	1/1000	1/200	H1 (pH6)	30 min
CHD4	Cell Signaling Technology	D8B12	11912	1/1000			
EZH2	Active Motif	polyclonal	39933	1/5000	-	-	-
GAPDH	Millipore Sigma	polyclonal	AB2302	1/10000	-	-	-
Histone H3	Cell Signaling Technology	D1H2	4499	1/1000			
HOXB13	Novus Biologicals	polyclonal	NBP2-48778	1/500			
Ki-67	Dako	MIB-1	M7240	-	1/50	H2 (pH9)	20 min
MAP2	Abcam	polyclonal	ab32454	1/500			
Mouse Anti-rabbit IgG	Cell Signaling Technology	L27A9	5127	1/2000			
MTA1	Cell Signaling Technology	D40DY	5647	1/1000	-	-	-
NKX3.1	Cell Signaling Technology	D2Y1A	83700	1/1000	-	-	-
p-Rb1	Cell Signaling Technology	D20B12	8516	1/1000	-	-	-
p21	Cell Signaling Technology	12D1	2947	1/1000	-	-	-
p53	Santa Cruz	DO-1	sc-126	1/1000	-	-	-
Rb1	Abcam				-	-	-
REST	Millipore Sigma	polyclonal	07-579	1/1000	-	-	-
SOX2	Cell Signaling Technology	D6D9	3579	1/1000	1/100	H2 (pH9)	20 min
synaptophysin	Thermo Scientific	SP11	RM9111-S	-	1/100	H2 (pH9)	20 min
synaptophysin	Abcam	YE269	ab32127	1/1000	-	-	-
TTF1 / NKX2.1	Abcam	EP1584Y	ab76013	1/2000	-	-	-
VEGF	Abcam	polyclonal	ab69989	1/500			
Vinculin	Abcam	EPR8185	ab129002	1/5000	-	-	-

Table 10: antibodies and conditions used in this study.

The intensity of nuclear immunostaining for SWI/SNF subunits was evaluated on tissue micro-arrays (TMAs) and whole slide sections by a pathologist (J.C.) blinded to additional pathological and clinical data, and was scored as negative (score 0), weak (score 1), moderate (score 2) or strong (score 3). Association between disease state and staining intensity

(negative/weak vs. moderate/strong) was examined using the two-tailed Fisher's exact test.

Analysis of *SMARCA4* and *SMARCA2* expression in localized PCa versus clinical outcome

The patient cohort with localized PCa and available clinical and follow-up information has been previously described (Spahn et al., 2010). IHC for *SMARCA4* and *SMARCA2* was performed on TMAs constructed from these patients' prostatectomy specimens. Staining intensity was scored by a pathologist (J.C.) blinded to the clinical data, using the digital online TMA scoring tool Scorenado (University of Bern, Switzerland). The Kaplan-Meier method was used to estimate patients' overall survival. The association between *SMARCA4* and *SMARCA2* expression (strong vs. moderate/weak/negative) and overall survival was examined using the log-rank test and multivariable Cox proportional hazards regression models. Ninety-five percent confidence intervals were calculated to assess the precision of the obtained hazard ratios. All p-values were two-sided, and statistical significance was evaluated at the 0.05 alpha level. All analyses were performed in R (3.5.1) for Windows.

Development of a *SMARCA4* knock-down signature

We defined the *SMARCA4* knock-down signature by selecting a list of differentially expressed genes between *SMARCA4* siRNA-mediated knock-down and Scrambled control in the LNCaP cell line with a log fold change of 1.5 and an FDR < 0.01. For each sample, gene expression data were first normalized by z-score transformation. Then signature score was calculated as a weighted sum of normalized expression of the genes in the signature and was finally re-scaled with the 2.5% and 97.5% quantiles equaled -1 and +1, respectively. We defined samples with low *SMARCA4* knock-down signature score as the 25% of cases with the lowest scores, and samples with high signature score as the 25% of cases with the highest scores.

Validation of *SMARCA4* knock-down signature in multiple clinical cohorts

SMARCA4 knock-down generated signature was applied to two CRPC cohorts consisting of 332 patients from the Stand Up To Cancer-Prostate Cancer Foundation (SU2C-PCF) trial treated with ARSi (recently published by Abida et al (Abida et al., 2019b)) and 47 patients from the Weil Cornell Medicine (WCM) cohort (published by Beltran et al (Beltran et al., 2016)) and on one cohort of localized, hormone treatment-naïve PCa consisting of 495 patients from The Cancer Genome Atlas (TCGA).

Results from the signature was then correlated with NEPC score and AR signaling scores for the SU2C-PCF and the WCM dataset and with Gleason score for the TCGA dataset.

Decipher GRID analysis

For prospective Decipher GRID and JHMI cohort, tumor RNA was extracted from FFPE blocks or slides after macrodissection guided by a histologic review of the tumor lesion by a GU pathologist. RNA extraction and microarray hybridization were all done in a Clinical Laboratory Improvement Amendments (CLIA)-certified laboratory facility (GenomeDx Biosciences, San Diego, CA, USA). Total RNA was amplified and hybridized to Human Exon 1.0 ST GeneChips (Thermo-Fisher, Carlsbad, CA). All data was normalized using the Single Channel Array Normalization (SCAN) algorithm (Piccolo et al., 2012). Decipher scores were calculated based on the predefined 22-markers (Spratt et al., 2017). Patients with high Decipher (>0.7) were categorized as genomically high risk patients. Mann–Whitney U test was used to assess score differences across Gleason score groups and Mann-Kendall trend test was used to test the association between the percentage of high Decipher scores across deciles of the *SMARCA4* knock-down signature. Kaplan-Meier analysis and Cox proportional hazard model was used to associate *SMARCA4* knock-down signature with time to metastasis in the JHMI cohort.

Cell culture

Commercially available PCa cell lines (RWPE-1, LNCaP, 22Rv1, VCaP, LAPC4, PC3, DU145, NCI-H660, C4-2) were purchased from ATCC and maintained according to ATCC protocols. WCM154 and WCM155 CRPC-NE cell lines have been previously established and

were maintained in two-dimensional monolayer culture according to the previously described protocol (Puca et al., 2018). LNCaP-AR cells were a kind gift from Dr. Sawyers and Dr. Mu (Memorial Sloan Kettering Cancer Center) and were cultured as previously described (Mu et al., 2017). MSKCC-PCa3 CRPC-Adeno cells were a kind gift from Dr. Chen (Memorial Sloan Kettering Cancer Center) and were maintained identically to WCM154 and WCM155 cells. Cell cultures were regularly tested for *Mycoplasma* contamination and confirmed to be negative.

Cell transfection and siRNA-mediated knock-down

ON-TARGET plus siRNA SMARTpool siRNAs against *SMARCA4*, *SMARCA2*, *SMARCC1*, *SMARCC2* and REST were purchased from Dharmacon. Transfection was performed overnight on attached cells growing in 6-well plates using the Lipofectamine 3000 reagent (Thermo Fisher Scientific) to the proportions of 10 μ L of 20 μ M siRNA per well. Cells were harvested for protein and RNA extraction 72h after transfection.

Cell infection, shRNA-mediated knock-down and gene overexpression

The *ACTL6B* shRNA and the matching Scrambled shRNA control were a kind gift from Dr. Cigall Kadoch (Dana Farber Cancer Institute). The vector was pGIPZ and the target sequence was: sh#1 – TGGATCACACCTACAGCAA. The *DPF1* shRNA and the corresponding Scrambled shRNA control were purchased from Genecopoeia. The vector was psi-LVRU6GP and the target sequences were: sh#1 – GAATTAAGTTGTTCTGTGTAT, Scrambled control - GCTTCGCGCCGTAGTCTTA. For infection, WCM155 cells were collected, resuspended in media containing Polybrene (Millipore) and lentiviral particles, and centrifuged at 800xg at room temperature for 60 min. Both vectors included a GFP reporter and infection efficiency was confirmed by green fluorescence. Cells were harvested for protein and RNA extraction 72h after transfection. Given the short-term nature of the experiments, selection was not performed. For the *SMARCA4* or *SMARCA2* overexpression experiment, lentiviral particles were prepared as described above using the pEZ-Lv203 vector (*SMARCA4* gene, eGFP reporter), the pEZLv216 vector (*SMARCA2* gene, mCherry reporter) (all vectors Genecopoeia,

MD, USA; all sequence-verified). 22Rv1 cells were infected as described above, cultured and sorted based on the expression of the fluorescent reporter.

Immunoblotting

Cells were lysed in RIPA buffer with protease and phosphatase inhibitors (Thermo Fisher Scientific) and total protein concentration was measured using the DC Protein Assay (Bio-Rad). Protein samples were resolved in SDS-PAGE, transferred onto a nitrocellulose membrane using the iBlot 2 dry blotting system (Thermo Fisher Scientific) and incubated overnight at 4°C with primary antibodies dissolved in 5% Blotting-Grade Blocker (Bio-Rad). All primary antibodies and dilutions used are listed in Supplementary Table STm.1. After 3 washes, the membrane was incubated with secondary antibody conjugated to horseradish peroxidase for 1h at room temperature. After 3 washes, signal was visualized by chemiluminescence using the Luminata Forte substrate (Thermo Fisher Scientific) and images were acquired with the ChemiDoc™ Touch Imaging System (Bio-Rad, Hercules, CA).

CRISPR-Cas9 mediated *TP53* and *RB1* knock-out

To generate the stable p53 and RB1 knockout cells, all-in-one CRISPR plasmids with mCherry reporter were purchased from Genecopoeia (Cat # HCP218175-CG01, HCP216131-CG01). Cells were transfected with CRISPR plasmids, selected with puromycin and sorted for mCherry positivity. *TP53* gRNA sequences used: TCGACGCTAGGATCTGACTG, CGTCGAGCCCCCTCTGAGTC, CCATTGTTCAATATCGTCCG. *RB1* gRNA sequences used: CGGTGGCGGCCGTTTTTCGG, CGGTGCCGGGGGTTCCGCGG, CGGAGGACCTGCCTCTCGTC. Control gRNA sequence: GGCTTCGCGCCGTAGTCTTA.

RNA extraction from cells, RNA sequencing and analysis, qPCR

Total RNA was extracted from cells using the Maxwell 16 LEV simplyRNA Purification Kit and the Maxwell 16 Instrument. RNA integrity was verified using the Agilent Bioanalyzer 2100 (Agilent Technologies). cDNA was synthesized from total RNA using Superscript III (Invitrogen). Library preparation was performed using TruSeq RNA Library Preparation Kit v2.

RNA sequencing was performed on the HiSeq 2500 sequencer to generate 2x75bp paired-end reads. Sequence reads were aligned using STAR two-pass (A. Dobin et al., 2013) to the human reference genome GRCh37. Gene counts were quantified using the “GeneCounts” option. Per-gene counts-per-million (CPM) were computed and \log_2 -transformed adding a pseudo-count of 1 to avoid transforming 0. Genes with \log_2 -CPM <1 in more than three samples were removed. Unsupervised clustering was performed using the top 500 most variable genes, Euclidean distance as the distance metric and the Ward clustering algorithm. When required, the batch effect was removed using the function removeBatchEffect from the limma R package for data visualization. For differential expression the batch factor was included in the design matrix.

Differential expression analysis between knock-down cells and control samples was performed using the edgeR package (Nikolayeva & Robinson, 2014). Normalization was performed using the “TMM” (weighted trimmed mean) method and differential expression was assessed using the quasi-likelihood F-test.

Genes with FDR <0.05 and > 2-fold were considered significantly differentially expressed.

Gene Set Enrichment Analysis (GSEA) was performed using the Preranked tool (Subramanian et al., 2005) for C2 (canonical pathways) and H (hallmark gene sets) (Liberzon et al., 2011).

Genes were ranked based on the T-statistic from the differential expression analysis.

Primer sequences used for RT-PCR are available in **Table 11**.

Oligos (RT-PCR)	Sequence (5'-3')
MAP2 fw	CGAAGCGCCAATGGATTCC
MAP2 rv	TGAACTATCCTTGCAGACACCT
VGF fw	GGAACTGCGAGATTTTCAGTCC
VGF rv	GTGCGGGTTTCCGTCTCTG
MTA1 fw	CATCAGAGGCCAACCTTTTCG
MTA1 rv	GCACGTATCTGTCCGGTGGTC
SMARCC1 fw	TCTTGGGGCTGCTTACAAGTA
SMARCC1 rv	TCCATTCGAGATGGGTTCTGTAG
ACTB fw	TGACGTGGACATCCGCAAAG
ACTB rv	CTGGAAGGTGGACAGCGAGG

GAPDH fw	GACAGTCAGCCGCATCTTCT
GAPDH rv	TTAAAAGCAGCCCTGGTGAC
DPF1 fw	GTACAAGATCGACTGTGAAGCACC
DPF1 rv	CAACTGCTGTTTCTGACAGTCCATA
REST fw	GAACTCATACAGGAGAACGCCC
REST rv	GGCTTCTCACCTGAATGAGTACG
BAF53b fw	GAATGGCATGATCGAGGACTGGG
BAF53b rv	CGTGTGTTCCACGGAGCCTC

Table 11. Primer sequences for RT-PCR.

Cell growth experiments

Cells were treated with siRNA (3 pmol) against *SMARCA4*, *SMARCA2*, *SMARCC1*, *SMARCC2* or with a scrambled control for 24h. LNCaP and C4-2 cells were then seeded in Poly-L- Lysine coated 96-well plates (2000 cells / well) and WCM154 cells were seeded in a collagen-coated 96-well plates (5000 cells / well). Cell viability was determined after 24, 48, 72 and 96 hours with a Tecan Infinite M200PRO reader using the CellTiter-Glo® Luminescent Cell Viability Assay according to manufacturer's directions (Promega). Cell confluence was determined using the Incucyte S3 instrument and the IncuCyte S3 2018B software (Essen Bioscience, Germany). Values were calculated as x-fold of cells transfected with siRNA for 0 hours.

Co-immunoprecipitation and mass spectrometry analysis

For the second Co-IP using an anti-SMARCA4 antibody, SWI/SNF complexes were isolated from the nuclear fraction of LNCaP (adenocarcinoma) or NCI-H660 (CRPC-NE) cells, which was prepared using the Universal CoIP Kit (Active Motif). Briefly, anti-Brg-1 antibodies (H-10, Santa Cruz Biotechnology) were cross-linked using Dimethyl pimelimidate dihydrochloride (Sigma-Aldrich) to Protein G conjugated magnetic beads (Bio-Rad). 30µg of cross-linked antibodies were incubated with 0.8-1 mg of nuclear lysates overnight. Bead-bound BAF complexes were washed and eluted using 8M urea buffer. The obtained protein complexes were subjected to immunoblotting and MS analysis.

For MS analysis, the eluted proteins were precipitated with trichloroacetic acid (TCA, 20% w/v), rinsed three times with acetone, and dried at room temperature. The pellets were re-suspended in 50 μ L resuspension buffer (8M urea, 50mM ammonium bicarbonate, and 5mM DTT) and subjected to reduction and alkylation by adding 15mM iodoacetamide to each sample for 30 min in the dark at room temperature, followed by addition of 5mM DTT to quench the reaction. Samples were diluted to a final concentration of 2M urea and digested with LysC at room temperature overnight, and then diluted further at 1M urea and digested with Trypsin at 37°C overnight (for each enzyme a ratio of 1:125 enzyme:protein was used).

Samples were labeled using reductive dimethylation. Labeling was done while the peptides were bound to the solid phase C18 resin in self-packed STAGE Tip micro-columns. Stage tips were washed with methanol, acetonitrile (ACN) 70% v/v and formic acid (FA) 1% v/v. Samples were acidified by adding 100% FA to a final concentration of 2% FA before loading. After sample loading, stage tips were washed with 1% FA and phosphate/citrate buffer (0.23M sodium phosphate and 86.4mM citric acid [pH 5.5]). At this point, the “light” solution (0.4% CH₂O and 60mM NaBH₃CN), or “heavy” solution (0.4% CD₂O and 60mM NaBD₃CN) was added twice on each stage tip to label the peptides. A final wash with 1% FA was performed prior to elution with 70% ACN and 1% FA. Samples were dried under vacuum, resuspended in 5% FA, and mixed together in equal amounts for analysis using an Orbitrap Fusion Mass Spectrometer. Peptides were introduced into the mass spectrometer by nano-electrospray as they eluted off a self-packed 40cm, 75 μ m (ID) reverse-phase column packed with 1.8 μ m, 120Å pore size, SEPAX C18 resin. Peptides were separated with a gradient of 5–25% buffer B (99.9% ACN, 0.1% FA) with a flow rate of 350 nl/min for 65 min. For each scan cycle, one high mass resolution full MS scan was acquired in the Orbitrap mass analyzer at a resolution of 120K, AGC value of 500000, in a m/z scan range of 375-1400, max acquisition time of 100ms and up to 20 parent ions were chosen based on their intensity for collision induced dissociation (normalized collision energy=35%) and MS/MS fragment ion scans at low mass resolution in the

linear ion trap. Dynamic exclusion was enabled to exclude ions that had already been selected for MS/MS in the previous 40 sec. Ions with a charge of +1 and those whose charge state could not be assigned were also excluded. All scans were collected in centroid mode. Two biological replicates for each condition were processed and analyzed.

MS2 spectra were searched using SEQUEST (version 28 revision 13) against a composite database containing all Swiss-Prot reviewed human protein sequences (20,193 target sequences, downloaded from www.uniprot.org March 18, 2016) and their reversed complement, using the following parameters: a precursor mass tolerance of ± 25 ppm; 1.0 Da product ion mass tolerance; tryptic digestion; up to two missed cleavages; static modifications of carbamidomethylation on cysteine (+57.0214) and dimethylation on n-termini and lysines (+28.0313); dynamic modifications of methionine oxidation (+15.9949) and heavy dimethylation on N-termini and lysines (+6.03766). Peptide spectral matches (PSMs) were filtered to 1% FDR using the target-decoy strategy combined with linear discriminant analysis (LDA) using several different parameters including Xcorr, $\Delta Cn'$, precursor mass error, observed ion charge state, and predicted solution charge state. Linear discriminant models were calculated for each LC-MS/MS run using peptide matches to forward and reversed protein sequences as positive and negative training data. PSMs within each run were sorted in descending order by discriminant score and filtered to a 1% FDR as revealed by the number of decoy sequences remaining in the data set. The data were further filtered to control protein level FDRs. Peptides were combined and assembled into proteins. Protein scores were derived from the product of all LDA peptide probabilities, sorted by rank, and filtered to 1% FDR as described for peptides. The FDR of the remaining peptides fell dramatically after protein filtering. The data were further filtered to require a minimum of 8 PSMs per protein. All peptides were required to have a sum of heavy and light signal-to-noise (SN) ≥ 10 . Protein ratios were calculated as the \log_2 ratio of the total SN of all experimental sample peptide values over that for IgG control sample peptides. For a small number of the most highly enriched proteins, the control value was zero (this is the

theoretical ideal). In these cases, we imputed a value of one for ratio calculations. Subsequent visualization and statistical analysis was done with Perseus and R program.

ChIP-sequencing data analysis

ChIP-seq peaks for *SMARCC1* and *HOXB13* in LNCaP cells were downloaded from GEO: [GSE110655](#) and [GSE94682](#), respectively. ChIP-seq peaks for H3K27ac and H3K27me3 in LNCaP cells were from data published by Sandoval et al. (Sandoval et al., 2018). Peak comparison was performed using BEDTOOLS (<https://bedtools.readthedocs.io/en/latest/#>).

Part IV. Discussion and conclusions: a pleiotropic role of SWI/SNF in cancer.

In this dissertation, I presented two examples of malignancies in which *SMARCA4* plays diametrically different roles. In SCCOHT, *SMARCA4* loss is the bona fide oncogenic event; conversely, depletion of SWI/SNF subunits (such as *SMARCA4* and *SMARCC1*) impairs cell proliferation in at least a subset of PCa. These observations support a pleiotropic role for the SWI/SNF chromatin remodeling complex in cancer. Indeed, although the complex appears to play the role of a tumor suppressor in multiple cancer types (S. Jones et al., 2010; Kadoch et al., 2013; Versteeg et al., 1998; Witkowski et al., 2014), there is increasing evidence for dependency of other malignancies on SWI/SNF (Buscarlet et al., 2014; Clark et al., 1994; Hiramatsu et al., 2017; Jubierre et al., 2016; Kadoch & Crabtree, 2013; Laurette et al., 2019; Q. Wu, Madany, et al., 2016).

An important result of previously published studies and of our work in PCa is the fact that the expression of various SWI/SNF subunits is subject to inter-tumor and intra-tumor heterogeneity. As such, future studies should strive to define the specific composition of the complex, rather than propose conclusions about “the SWI/SNF complex” in general as if it were a single, homogenous entity. It is possible that specialized forms of the complex may display tumor-suppressive or tumor-promoting functions. For example, one attractive hypothesis is that esBAF-like complexes could be present in cancer cells and responsible for cancer progenitor cell maintenance and/or phenotype plasticity.

In tumors with genomic SWI/SNF alterations, a careful understanding of the functional consequences of each type of alteration is critical for proposing new treatment strategies. For instance, in synovial sarcoma, attenuated expression of the *SMARCB1*, a core SWI/SNF subunit, is seen by IHC (**Fig.6**). Based on this observation, synovial sarcoma was postulated to represent a “SWI/SNF-deficient tumor” and as such, to exhibit excessive activity of the PRC

(polycomb) complex and to be potentially sensitive to EZH2 inhibitors. However, functional studies from Cigall Kadoch's group demonstrated that attenuated SMARCB1 expression was in fact due to disruption of the complex composition by incorporation of the abnormal SS18-SSX fusion protein, leading to ejection of SMARCB1, which subsequently gets degraded (Kadoch & Crabtree, 2013). The SS18-SSX fusion product retargets the complex from enhancer sites to broad polycomb domains, where it reverses polycomb-mediated repression and leads to aberrant gene expression (Kadoch & Crabtree, 2013; McBride et al., 2018). Thus, EZH2 inhibition could be ineffective, or even potentially harmful, in these patients. This may explain why in the phase II study (NCT02601950) investigating the EZH2 inhibitor tazemetostat in adult patients with solid tumors, no objective responses were observed in synovial sarcoma patients (Schoffski P., 2017).

The role of SWI/SNF in tumors may be highly dependent on the genomic, epigenetic and “master transcription factor” context. This context-dependent paradigm has been gaining support both in regards to SWI/SNF and to other epigenetic regulators. For example, large B cell lymphomas with a centroblast-like phenotype are known to harbor activating *EZH2* mutations, while myeloid malignancies are characterized by loss-of-function *EZH2* alterations (Velcheti et al., 2019).

The fact that alterations in an epigenetic regulator seem to depend on the underlying cell type and differentiation stage could explain the relationships between the temporality of SWI/SNF alteration and oncogenesis. In mice with inducible *Smarcb1* inactivation, the phenotype is dependent on the temporality of *Smarcb1* inactivation; specifically, early inactivation (between days E6 and E10) results in the development of intracranial tumors reminiscent of human AT/RTs with highest penetrance (Han et al., 2016). In another mouse study, early *Smarcb1* inactivation in neural crests led to the development of MRT, while *Smarcb1* inactivation in Schwann cells at a later developmental stage and in conjunction with biallelic *Nf2* inactivation recapitulated schwannomatosis (Vitte et al., 2017).

Interesting observations were also made regarding other SWI/SNF subunits. Glaros et al. developed a conditional and lung-specific *Smarca4* knock-out mouse model to study the effects of *Smarca4* inactivation on lung cancer development (Glaros et al., 2008). In this model, heterozygous inactivation of *Smarca4* enhanced the number and size of lung tumors after carcinogen exposure, but homozygous *Smarca4* inactivation did not. The authors found that homozygous *Smarca4* loss led to increased apoptosis in normal lung tissue, and hypothesized that this loss of cell viability may prevent oncogenesis. However, when *Smarca4* inactivation was induced after carcinogen exposure, homozygous *Smarca4* loss potentiated tumor development. Similar results were obtained in a liver cancer mouse model upon liver-specific inactivation of *Arid1a* (Sun et al., 2017). Homozygous *Arid1a* inactivation had a protective role against initiation of hepatocellular carcinoma in two different toxin-induced models, while *Arid1a* overexpression increased the tumor burden. Conversely, *Arid1a* deletion after the tumors had developed accelerated tumor growth and metastasis. An ambiguous role of *Smarca4* in oncogenesis has also been shown in pancreatic cancer mouse models (Roy et al., 2015). Loss of *Smarca4* inhibited formation of pancreatic intraepithelial neoplasia (PanIn) from pancreatic acinar cells expressing oncogenic *Kras*, but promoted the development of intraductal papillary mucinous neoplasm (IPMN) from pancreatic ductal cells expressing oncogenic *Kras*. Unexpectedly, in *Smarca4*-deficient pancreatic ductal adenocarcinoma (PDA), re-expression of *Smarca4* enhanced tumor growth. The authors proposed a working model whereby SMARCA4 may inhibit dedifferentiation of mature cells, but may have tumor-promoting functions in established PDA.

While the results of these studies appear somewhat discrepant (e.g., SMARCA4 seems to act as a tumor suppressor in established tumors in the lung and to have tumor promoting functions in established tumors in the pancreas), and further studies are needed to better elucidate these relationships, they support the context-dependent role of SWI/SNF in oncogenesis.

In addition, in some situations, the effect SWI/SNF alterations may rely on the concomitant presence of other alterations. This is the case for concomitant *SMARCB1* and *NF2* alterations in schwannomatosis (Vitte et al., 2017). Likewise, in a mouse model of ovarian clear cell carcinoma, *ARID1A* mutations were shown to cooperate with *PIK3CA* mutations, which are also found in those tumors, in the tumorigenesis process (Chandler et al., 2015). And in medulloblastoma, *SMARCA4* mutations are more often observed in the Wnt molecular subtype than in other subtypes (Masliah-Planchon et al., 2015; Pugh et al., 2012).

In summary, there is increasing evidence that the role of SWI/SNF in cancer is highly dependent on the cell-type, the timing, as well as the composition of the complex achieved by combinatorial subunit assembly. Identifying the elusive “cell of origin” of some malignancies could be particularly helpful in trying to explain some paradoxical observations, such as the ambiguous role of *SMARCA4* in cancer as illustrated herein with the examples of SCCOHT and prostatic carcinoma.

References

- Abida, W., Cyrta, J., Heller, G., Prandi, D., Armenia, J., Coleman, I., . . . Sawyers, C. L. (2019a). Genomic correlates of clinical outcome in advanced prostate cancer. *Proc Natl Acad Sci U S A*, 116(23), 11428-11436. doi:10.1073/pnas.1902651116
- Abida, W., Cyrta, J., Heller, G., Prandi, D., Armenia, J., Coleman, I., . . . Sawyers, C. L. (2019b). Genomic correlates of clinical outcome in advanced prostate cancer. *Proc Natl Acad Sci U S A*. doi:10.1073/pnas.1902651116
- Abro, B., Kaushal, M., Chen, L., Wu, R., Dehner, L. P., Pfeifer, J. D., & He, M. (2019). Tumor mutation burden, DNA mismatch repair status and checkpoint immunotherapy markers in primary and relapsed malignant rhabdoid tumors. *Pathol Res Pract*, 215(6), 152395. doi:10.1016/j.prp.2019.03.023
- Agaimy, A., Bertz, S., Cheng, L., Hes, O., Junker, K., Keck, B., . . . Hartmann, A. (2016). Loss of expression of the SWI/SNF complex is a frequent event in undifferentiated/dedifferentiated urothelial carcinoma of the urinary tract. *Virchows Arch*, 469(3), 321-330. doi:10.1007/s00428-016-1977-y
- Agaimy, A., Daum, O., Markl, B., Lichtmanegger, I., Michal, M., & Hartmann, A. (2016). SWI/SNF Complex-deficient Undifferentiated/Rhabdoid Carcinomas of the Gastrointestinal Tract: A Series of 13 Cases Highlighting Mutually Exclusive Loss of SMARCA4 and SMARCA2 and Frequent Co-inactivation of SMARCB1 and SMARCA2. *Am J Surg Pathol*, 40(4), 544-553. doi:10.1097/PAS.0000000000000554
- Agaimy, A., Koch, M., Lell, M., Semrau, S., Dudek, W., Wachter, D. L., . . . Hartmann, A. (2014). SMARCB1(INI1)-deficient sinonasal basaloid carcinoma: a novel member of the expanding family of SMARCB1-deficient neoplasms. *Am J Surg Pathol*, 38(9), 1274-1281. doi:10.1097/PAS.0000000000000236
- Aggarwal, R., Huang, J., Alumkal, J. J., Zhang, L., Feng, F. Y., Thomas, G. V., . . . Small, E. J. (2018). Clinical and Genomic Characterization of Treatment-Emergent Small-Cell Neuroendocrine Prostate Cancer: A Multi-institutional Prospective Study. *J Clin Oncol*, 36(24), 2492-2503. doi:10.1200/JCO.2017.77.6880
- al., J. R. L. e. (2018). *Abstract 1612PD: A phase II, multicenter study of the EZH2 inhibitor tazemetostat in adults (rhabdoid tumor cohort) (NCT02601950)*.
- Andrysik, Z., Bernstein, W. Z., Deng, L., Myer, D. L., Li, Y. Q., Tischfield, J. A., . . . Bahassi el, M. (2010). The novel mouse Polo-like kinase 5 responds to DNA damage and localizes in the nucleolus. *Nucleic Acids Res*, 38(9), 2931-2943. doi:10.1093/nar/gkq011
- Bacci, C., Sestini, R., Provenzano, A., Paganini, I., Mancini, I., Porfirio, B., . . . Papi, L. (2010). Schwannomatosis associated with multiple meningiomas due to a familial SMARCB1 mutation. *Neurogenetics*, 11(1), 73-80. doi:10.1007/s10048-009-0204-2
- Bai, J., Mei, P., Zhang, C., Chen, F., Li, C., Pan, Z., . . . Zheng, J. (2013). BRG1 is a prognostic marker and potential therapeutic target in human breast cancer. *PLoS One*, 8(3), e59772. doi:10.1371/journal.pone.0059772

- Bai, J., Mei, P. J., Liu, H., Li, C., Li, W., Wu, Y. P., . . . Zheng, J. N. (2012). BRG1 expression is increased in human glioma and controls glioma cell proliferation, migration and invasion in vitro. *J Cancer Res Clin Oncol*, 138(6), 991-998. doi:10.1007/s00432-012-1172-8
- Baranov, E., McBride, M. J., Bellizzi, A. M., Ligon, A. H., Fletcher, C. D. M., Kadoch, C., & Hornick, J. L. (2020). A Novel SS18-SSX Fusion-specific Antibody for the Diagnosis of Synovial Sarcoma. *Am J Surg Pathol*, 44(7), 922-933. doi:10.1097/PAS.0000000000001447
- Battaglioli, E., Andres, M. E., Rose, D. W., Chenoweth, J. G., Rosenfeld, M. G., Anderson, M. E., & Mandel, G. (2002). REST repression of neuronal genes requires components of the hSWI.SNF complex. *J Biol Chem*, 277(43), 41038-41045. doi:10.1074/jbc.M205691200
- Beltran, H., Prandi, D., Mosquera, J. M., Benelli, M., Puca, L., Cyrta, J., . . . Demichelis, F. (2016). Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. *Nat Med*, 22(3), 298-305. doi:10.1038/nm.4045
- Beltran, H., Rickman, D. S., Park, K., Chae, S. S., Sboner, A., MacDonald, T. Y., . . . Rubin, M. A. (2011). Molecular characterization of neuroendocrine prostate cancer and identification of new drug targets. *Cancer Discov*, 1(6), 487-495. doi:10.1158/2159-8290.CD-11-0130
- Ben-Porath, I., Thomson, M. W., Carey, V. J., Ge, R., Bell, G. W., Regev, A., & Weinberg, R. A. (2008). An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. *Nat Genet*, 40(5), 499-507. doi:10.1038/ng.127
- Blanc-Durand, F., Lefeuvre-Plesse, C., Ray-Coquard, I., Chaltiel, D., Floquet, A., Meriaux, E., . . . Pautier, P. (2020). Dose-intensive regimen treatment for small-cell carcinoma of the ovary of hypercalcemic type (SCCOHT). *Gynecol Oncol*. doi:10.1016/j.ygyno.2020.07.019
- Bluemn, E. G., Coleman, I. M., Lucas, J. M., Coleman, R. T., Hernandez-Lopez, S., Tharakan, R., . . . Nelson, P. S. (2017). Androgen Receptor Pathway-Independent Prostate Cancer Is Sustained through FGF Signaling. *Cancer Cell*, 32(4), 474-489 e476. doi:10.1016/j.ccell.2017.09.003
- Bourdeaut, F., Lequin, D., Brugieres, L., Reynaud, S., Dufour, C., Doz, F., . . . Delattre, O. (2011). Frequent hSNF5/INI1 germline mutations in patients with rhabdoid tumor. *Clin Cancer Res*, 17(1), 31-38. doi:10.1158/1078-0432.CCR-10-1795
- Bourdeaut F., T. M. D., Bergthold G., Karski E. . (2017). ATRT-11. Marked response to atezolizumab in a patient with rhabdoid tumor: a case study from the imatrix-atezolizumab trial. *Neuro-Oncology*, Volume 19, Issue suppl_4, 1 June 2017, Page iv3.
- Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*, 68(6), 394-424. doi:10.3322/caac.21492

- Bultman, S., Gebuhr, T., Yee, D., La Mantia, C., Nicholson, J., Gilliam, A., . . . Magnuson, T. (2000). A Brg1 null mutation in the mouse reveals functional differences among mammalian SWI/SNF complexes. *Mol Cell*, 6(6), 1287-1295. doi:10.1016/s1097-2765(00)00127-1
- Buscarlet, M., Krasteva, V., Ho, L., Simon, C., Hebert, J., Wilhelm, B., . . . Lessard, J. A. (2014). Essential role of BRG, the ATPase subunit of BAF chromatin remodeling complexes, in leukemia maintenance. *Blood*, 123(11), 1720-1728. doi:10.1182/blood-2013-02-483495
- Calderaro, J., Masliah-Planchon, J., Richer, W., Maillot, L., Maille, P., Mansuy, L., . . . Bourdeaut, F. (2016). Balanced Translocations Disrupting SMARCB1 Are Hallmark Recurrent Genetic Alterations in Renal Medullary Carcinomas. *Eur Urol*, 69(6), 1055-1061. doi:10.1016/j.eururo.2015.09.027
- Cancer Genome Atlas Research, N. (2014a). Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature*, 507(7492), 315-322. doi:10.1038/nature12965
- Cancer Genome Atlas Research, N. (2014b). Comprehensive molecular profiling of lung adenocarcinoma. *Nature*, 511(7511), 543-550. doi:10.1038/nature13385
- Cancer Genome Atlas Research, N. (2015). The Molecular Taxonomy of Primary Prostate Cancer. *Cell*, 163(4), 1011-1025. doi:10.1016/j.cell.2015.10.025
- Carioli, G., Bertuccio, P., Boffetta, P., Levi, F., La Vecchia, C., Negri, E., & Malvezzi, M. (2020). European cancer mortality predictions for the year 2020 with a focus on prostate cancer. *Ann Oncol*, 31(5), 650-658. doi:10.1016/j.annonc.2020.02.009
- Carvajal, R. D., Tse, A., & Schwartz, G. K. (2006). Aurora kinases: new targets for cancer therapy. *Clin Cancer Res*, 12(23), 6869-6875. doi:10.1158/1078-0432.CCR-06-1405
- Chakravarty, D., Sboner, A., Nair, S. S., Giannopoulou, E., Li, R., Hennig, S., . . . Rubin, M. A. (2014). The oestrogen receptor alpha-regulated lncRNA NEAT1 is a critical modulator of prostate cancer. *Nat Commun*, 5, 5383. doi:10.1038/ncomms6383
- Chan-Penebre, E., Armstrong, K., Drew, A., Grassian, A. R., Feldman, I., Knutson, S. K., . . . Ribich, S. A. (2017). Selective Killing of SMARCA2- and SMARCA4-deficient Small Cell Carcinoma of the Ovary, Hypercalcemic Type Cells by Inhibition of EZH2: In Vitro and In Vivo Preclinical Models. *Mol Cancer Ther*, 16(5), 850-860. doi:10.1158/1535-7163.MCT-16-0678
- Chandler, R. L., Damrauer, J. S., Raab, J. R., Schisler, J. C., Wilkerson, M. D., Didion, J. P., . . . Magnuson, T. (2015). Coexistent ARID1A-PIK3CA mutations promote ovarian clear-cell tumorigenesis through pro-tumorigenic inflammatory cytokine signalling. *Nat Commun*, 6, 6118. doi:10.1038/ncomms7118
- Chang, Y. T., Lin, T. P., Campbell, M., Pan, C. C., Lee, S. H., Lee, H. C., . . . Chang, P. C. (2017). REST is a crucial regulator for acquiring EMT-like and stemness phenotypes in hormone-refractory prostate cancer. *Sci Rep*, 7, 42795. doi:10.1038/srep42795

- Chen, C. D., Welsbie, D. S., Tran, C., Baek, S. H., Chen, R., Vessella, R., . . . Sawyers, C. L. (2004). Molecular determinants of resistance to antiandrogen therapy. *Nat Med*, *10*(1), 33-39. doi:10.1038/nm972
- Chen, R., Li, Y., Buttyan, R., & Dong, X. (2017). Implications of PI3K/AKT inhibition on REST protein stability and neuroendocrine phenotype acquisition in prostate cancer cells. *Oncotarget*, *8*(49), 84863-84876. doi:10.18632/oncotarget.19386
- Chun, H. E., Johann, P. D., Milne, K., Zapatka, M., Buellesbach, A., Ishaque, N., . . . Kool, M. (2019). Identification and Analyses of Extra-Cranial and Cranial Rhabdoid Tumor Molecular Subgroups Reveal Tumors with Cytotoxic T Cell Infiltration. *Cell Rep*, *29*(8), 2338-2354 e2337. doi:10.1016/j.celrep.2019.10.013
- Chun, H. E., Lim, E. L., Heravi-Moussavi, A., Saberi, S., Mungall, K. L., Bilenky, M., . . . Marra, M. A. (2016). Genome-Wide Profiles of Extra-cranial Malignant Rhabdoid Tumors Reveal Heterogeneity and Dysregulated Developmental Pathways. *Cancer Cell*, *29*(3), 394-406. doi:10.1016/j.ccell.2016.02.009
- Cibulskis, K., Lawrence, M. S., Carter, S. L., Sivachenko, A., Jaffe, D., Sougnez, C., . . . Getz, G. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature Biotechnology*, *31*, 213. doi:10.1038/nbt.2514
- Clapier, C. R., Iwasa, J., Cairns, B. R., & Peterson, C. L. (2017). Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nat Rev Mol Cell Biol*, *18*(7), 407-422. doi:10.1038/nrm.2017.26
- Clark, J., Rocques, P. J., Crew, A. J., Gill, S., Shipley, J., Chan, A. M., . . . Cooper, C. S. (1994). Identification of novel genes, SYT and SSX, involved in the t(X;18)(p11.2;q11.2) translocation found in human synovial sarcoma. *Nat Genet*, *7*(4), 502-508. doi:10.1038/ng0894-502
- Coatham, M., Li, X., Karnezis, A. N., Hoang, L. N., Tessier-Cloutier, B., Meng, B., . . . Lee, C. H. (2016). Concurrent ARID1A and ARID1B inactivation in endometrial and ovarian dedifferentiated carcinomas. *Mod Pathol*, *29*(12), 1586-1593. doi:10.1038/modpathol.2016.156
- Colaprico, A., Silva, T. C., Olsen, C., Garofano, L., Cava, C., Garolini, D., . . . Noushmehr, H. (2016). TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res*, *44*(8), e71. doi:10.1093/nar/gkv1507
- Coppola, V., Musumeci, M., Patrizii, M., Cannistraci, A., Addario, A., Maugeri-Sacca, M., . . . Bonci, D. (2013). BTG2 loss and miR-21 upregulation contribute to prostate cell transformation by inducing luminal markers expression and epithelial-mesenchymal transition. *Oncogene*, *32*(14), 1843-1853. doi:10.1038/onc.2012.194
- Dagogo-Jack, I., Schrock, A. B., Kem, M., Jessop, N., Lee, J., Ali, S. M., . . . Mino-Kenudson, M. (2020). Clinicopathologic Characteristics of BRG1-Deficient NSCLC. *J Thorac Oncol*, *15*(5), 766-776. doi:10.1016/j.jtho.2020.01.002

- Davies, A. H., Beltran, H., & Zoubeidi, A. (2018). Cellular plasticity and the neuroendocrine phenotype in prostate cancer. *Nat Rev Urol*, *15*(5), 271-286. doi:10.1038/nrurol.2018.22
- de Bono, J. S., Logothetis, C. J., Molina, A., Fizazi, K., North, S., Chu, L., . . . Investigators, C.-A.-. (2011). Abiraterone and increased survival in metastatic prostate cancer. *N Engl J Med*, *364*(21), 1995-2005. doi:10.1056/NEJMoa1014618
- Deisch, J., Raisanen, J., & Rakheja, D. (2011). Immunohistochemical expression of embryonic stem cell markers in malignant rhabdoid tumors. *Pediatr Dev Pathol*, *14*(5), 353-359. doi:10.2350/10-09-0902-OA.1
- Ding, Y., Li, N., Dong, B., Guo, W., Wei, H., Chen, Q., . . . Qin, J. (2019). Chromatin remodeling ATPase BRG1 and PTEN are synthetic lethal in prostate cancer. *J Clin Invest*, *129*(2), 759-773. doi:10.1172/JCI123557
- Doan, D. N., Veal, T. M., Yan, Z., Wang, W., Jones, S. N., & Imbalzano, A. N. (2004). Loss of the INI1 tumor suppressor does not impair the expression of multiple BRG1-dependent genes or the assembly of SWI/SNF enzymes. *Oncogene*, *23*(19), 3462-3473. doi:10.1038/sj.onc.1207472
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., . . . Gingeras, T. R. (2012). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15-21. doi:10.1093/bioinformatics/bts635
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., . . . Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15-21. doi:10.1093/bioinformatics/bts635
- Dunaief, J. L., Strober, B. E., Guha, S., Khavari, P. A., Alin, K., Luban, J., . . . Goff, S. P. (1994). The retinoblastoma protein and BRG1 form a complex and cooperate to induce cell cycle arrest. *Cell*, *79*(1), 119-130. doi:10.1016/0092-8674(94)90405-7
- E.J. Small, J. Y., J. Alumkal, C. Evans, C.J. Ryan, P. Lara, T. Beer, O. Witte, R. Baertsch, J. Stuart. (2014). *Neuroendocrine Prostate Cancer in Patients with Metastatic Castration Resistant Prostate Cancer Resistant to Abiraterone or Enzalutamide: Preliminary Results from the SU2C/PCF/AACR West Coast Prostate Cancer Dream Team*. . Paper presented at the ESMO.
- Eaton, K. W., Tooke, L. S., Wainwright, L. M., Judkins, A. R., & Biegel, J. A. (2011). Spectrum of SMARCB1/INI1 mutations in familial and sporadic rhabdoid tumors. *Pediatr Blood Cancer*, *56*(1), 7-15. doi:10.1002/pbc.22831
- Ehrenhofer-Wolfer, K., Puchner, T., Schwarz, C., Rippka, J., Blaha-Ostermann, S., Strobl, U., . . . Wohrle, S. (2019). SMARCA2-deficiency confers sensitivity to targeted inhibition of SMARCA4 in esophageal squamous cell carcinoma cell lines. *Sci Rep*, *9*(1), 11661. doi:10.1038/s41598-019-48152-x
- Eichhorn, J. H., Young, R. H., & Scully, R. E. (1992). Primary ovarian small cell carcinoma of pulmonary type. A clinicopathologic, immunohistologic, and flow cytometric analysis of 11 cases. *Am J Surg Pathol*, *16*(10), 926-938. doi:10.1097/00000478-199210000-00002

- Epstein, J. I., Amin, M. B., Beltran, H., Lotan, T. L., Mosquera, J. M., Reuter, V. E., . . . Rubin, M. A. (2014). Proposed morphologic classification of prostate cancer with neuroendocrine differentiation. *Am J Surg Pathol*, 38(6), 756-767. doi:10.1097/PAS.0000000000000208
- Epstein, J. I., Zelefsky, M. J., Sjoberg, D. D., Nelson, J. B., Egevad, L., Magi-Galluzzi, C., . . . Klein, E. A. (2015). A Contemporary Prostate Cancer Grading System: A Validated Alternative to the Gleason Score. *Eur Urol*. doi:10.1016/j.eururo.2015.06.046
- Fahiminiya, S., Witkowski, L., Nadaf, J., Carrot-Zhang, J., Goudie, C., Hasselblatt, M., . . . Foulkes, W. D. (2016). Molecular analyses reveal close similarities between small cell carcinoma of the ovary, hypercalcemic type and atypical teratoid/rhabdoid tumor. *Oncotarget*, 7(2), 1732-1740. doi:10.18632/oncotarget.6459
- Fizazi, K., Scher, H. I., Molina, A., Logothetis, C. J., Chi, K. N., Jones, R. J., . . . Investigators, C.-A.-. (2012). Abiraterone acetate for treatment of metastatic castration-resistant prostate cancer: final overall survival analysis of the COU-AA-301 randomised, double-blind, placebo-controlled phase 3 study. *Lancet Oncol*, 13(10), 983-992. doi:10.1016/S1470-2045(12)70379-0
- Fizazi, K., Tran, N., Fein, L., Matsubara, N., Rodriguez-Antolin, A., Alekseev, B. Y., . . . Investigators, L. (2017). Abiraterone plus Prednisone in Metastatic, Castration-Sensitive Prostate Cancer. *N Engl J Med*. doi:10.1056/NEJMoa1704174
- Forster, C., Ostertag, H., Schmitt, J., & Roessner, A. (1997). Small cell carcinoma of the ovary, hypercalcemic type. A case report with immunohistochemical, ultrastructural and cytophotometric analysis and review of the literature. *Gen Diagn Pathol*, 142(5-6), 365-370.
- Fruhwald, M. C., Hasselblatt, M., Wirth, S., Kohler, G., Schneppenheim, R., Subero, J. I., . . . Vormoor, J. (2006). Non-linkage of familial rhabdoid tumors to SMARCB1 implies a second locus for the rhabdoid tumor predisposition syndrome. *Pediatr Blood Cancer*, 47(3), 273-278. doi:10.1002/pbc.20526
- Gamwell, L. F., Gambaro, K., Merzotis, M., Crane, C., Arcand, S. L., Bourada, V., . . . Vanderhyden, B. C. (2013). Small cell ovarian carcinoma: genomic stability and responsiveness to therapeutics. *Orphanet J Rare Dis*, 8, 33. doi:10.1186/1750-1172-8-33
- Gao, X., Tate, P., Hu, P., Tjian, R., Skarnes, W. C., & Wang, Z. (2008). ES cell pluripotency and germ-layer formation require the SWI/SNF chromatin remodeling component BAF250a. *Proc Natl Acad Sci U S A*, 105(18), 6656-6661. doi:10.1073/pnas.0801802105
- Gatchalian, J., Malik, S., Ho, J., Lee, D. S., Kelso, T. W. R., Shokhirev, M. N., . . . Hargreaves, D. C. (2018). A non-canonical BRD9-containing BAF chromatin remodeling complex regulates naive pluripotency in mouse embryonic stem cells. *Nat Commun*, 9(1), 5139. doi:10.1038/s41467-018-07528-9
- Genestie, C., Blanc-Durand, F., Auguste, A., Pautier, P., Dunant, A., Scoazec, J. Y., . . . Leary, A. (2020). Clinical utility of SMARCA4 testing by immunohistochemistry in rare ovarian tumours. *Br J Cancer*, 122(4), 564-568. doi:10.1038/s41416-019-0687-z

- Georger, B., Kang, H. J., Yalon-Oren, M., Marshall, L. V., Vezina, C., Pappo, A., . . . Pinto, N. (2020). Pembrolizumab in paediatric patients with advanced melanoma or a PD-L1-positive, advanced, relapsed, or refractory solid tumour or lymphoma (KEYNOTE-051): interim analysis of an open-label, single-arm, phase 1-2 trial. *Lancet Oncol*, *21*(1), 121-133. doi:10.1016/S1470-2045(19)30671-0
- Glaros, S., Cirrincione, G. M., Muchardt, C., Kleer, C. G., Michael, C. W., & Reisman, D. (2007). The reversible epigenetic silencing of BRM: implications for clinical targeted therapy. *Oncogene*, *26*(49), 7058-7066. doi:10.1038/sj.onc.1210514
- Glaros, S., Cirrincione, G. M., Palanca, A., Metzger, D., & Reisman, D. (2008). Targeted knockout of BRG1 potentiates lung cancer development. *Cancer Res*, *68*(10), 3689-3696. doi:10.1158/0008-5472.CAN-07-6652
- Guerrero-Martinez, J. A., & Reyes, J. C. (2018). High expression of SMARCA4 or SMARCA2 is frequently associated with an opposite prognosis in cancer. *Sci Rep*, *8*(1), 2043. doi:10.1038/s41598-018-20217-3
- Habegger, L., Sboner, A., Gianoulis, T. A., Rozowsky, J., Agarwal, A., Snyder, M., & Gerstein, M. (2010). RSEQtools: a modular framework to analyze RNA-Seq data using compact, anonymized data summaries. *Bioinformatics*, *27*(2), 281-283. doi:10.1093/bioinformatics/btq643
- Han, Z. Y., Richer, W., Freneaux, P., Chauvin, C., Lucchesi, C., Guillemot, D., . . . Bourdeaut, F. (2016). The occurrence of intracranial rhabdoid tumours in mice depends on temporal control of Smarcb1 inactivation. *Nat Commun*, *7*, 10421. doi:10.1038/ncomms10421
- Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell*, *144*(5), 646-674. doi:10.1016/j.cell.2011.02.013
- Hargreaves, D. C., & Crabtree, G. R. (2011). ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. *Cell Res*, *21*(3), 396-420. doi:10.1038/cr.2011.32
- Hasselblatt, M., Gesk, S., Oyen, F., Rossi, S., Viscardi, E., Giangaspero, F., . . . Paulus, W. (2011). Nonsense mutation and inactivation of SMARCA4 (BRG1) in an atypical teratoid/rhabdoid tumor showing retained SMARCB1 (INI1) expression. *Am J Surg Pathol*, *35*(6), 933-935. doi:10.1097/PAS.0b013e3182196a39
- Hasselblatt, M., Nagel, I., Oyen, F., Bartelheim, K., Russell, R. B., Schuller, U., . . . Schneppenheim, R. (2014). SMARCA4-mutated atypical teratoid/rhabdoid tumors are associated with inherited germline alterations and poor prognosis. *Acta Neuropathol*, *128*(3), 453-456. doi:10.1007/s00401-014-1323-x
- Heeboll, S., Borre, M., Ottosen, P. D., Andersen, C. L., Mansilla, F., Dyrskjot, L., . . . Topping, N. (2008). SMARCC1 expression is upregulated in prostate cancer and positively correlated with tumour recurrence and dedifferentiation. *Histol Histopathol*, *23*(9), 1069-1076. doi:10.14670/HH-23.1069
- Heerma van Voss, M. R., van Diest, P. J., Smolders, Y. H., Bart, J., van der Wall, E., & van der Groep, P. (2014). Distinct claudin expression characterizes BRCA1-related breast cancer. *Histopathology*, *65*(6), 814-827. doi:10.1111/his.12490

- Helming, K. C., Wang, X., & Roberts, C. W. M. (2014). Vulnerabilities of mutant SWI/SNF complexes in cancer. *Cancer Cell*, 26(3), 309-317. doi:10.1016/j.ccr.2014.07.018
- Helming, K. C., Wang, X., Wilson, B. G., Vazquez, F., Haswell, J. R., Manchester, H. E., . . . Roberts, C. W. (2014). ARID1B is a specific vulnerability in ARID1A-mutant cancers. *Nat Med*, 20(3), 251-254. doi:10.1038/nm.3480
- Hieronymus, H., Lamb, J., Ross, K. N., Peng, X. P., Clement, C., Rodina, A., . . . Golub, T. R. (2006). Gene expression signature-based chemical genomic prediction identifies a novel class of HSP90 pathway modulators. *Cancer Cell*, 10(4), 321-330. doi:10.1016/j.ccr.2006.09.005
- Hiramatsu, H., Kobayashi, K., Kobayashi, K., Haraguchi, T., Ino, Y., Todo, T., & Iba, H. (2017). The role of the SWI/SNF chromatin remodeling complex in maintaining the stemness of glioma initiating cells. *Sci Rep*, 7(1), 889. doi:10.1038/s41598-017-00982-3
- Ho, L., Jothi, R., Ronan, J. L., Cui, K., Zhao, K., & Crabtree, G. R. (2009). An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. *Proc Natl Acad Sci U S A*, 106(13), 5187-5191. doi:10.1073/pnas.0812888106
- Ho, L., Miller, E. L., Ronan, J. L., Ho, W. Q., Jothi, R., & Crabtree, G. R. (2011). esBAF facilitates pluripotency by conditioning the genome for LIF/STAT3 signalling and by regulating polycomb function. *Nat Cell Biol*, 13(8), 903-913. doi:10.1038/ncb2285
- Ho, L., Ronan, J. L., Wu, J., Staahl, B. T., Chen, L., Kuo, A., . . . Crabtree, G. R. (2009). An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency. *Proc Natl Acad Sci U S A*, 106(13), 5181-5186. doi:10.1073/pnas.0812889106
- Hodgson, M. C., Shao, L. J., Frolov, A., Li, R., Peterson, L. E., Ayala, G., . . . Agoulnik, I. U. (2011). Decreased expression and androgen regulation of the tumor suppressor gene INPP4B in prostate cancer. *Cancer Res*, 71(2), 572-582. doi:10.1158/0008-5472.CAN-10-2314
- Hoffman, G. R., Rahal, R., Buxton, F., Xiang, K., McAllister, G., Frias, E., . . . Jagani, Z. (2014). Functional epigenetics approach identifies BRM/SMARCA2 as a critical synthetic lethal target in BRG1-deficient cancers. *Proc Natl Acad Sci U S A*, 111(8), 3128-3133. doi:10.1073/pnas.1316793111
- Hohmann, A. F., Martin, L. J., Minder, J. L., Roe, J. S., Shi, J., Steurer, S., . . . Vakoc, C. R. (2016). Sensitivity and engineered resistance of myeloid leukemia cells to BRD9 inhibition. *Nat Chem Biol*, 12(9), 672-679. doi:10.1038/nchembio.2115
- Ingham, P. W. (1983). Differential expression of bithorax complex genes in the absence of the extra sex combs and trithorax genes. *Nature*, 306(5943), 591-593. doi:10.1038/306591a0
- Itakura, E., Tamiya, S., Morita, K., Shiratsuchi, H., Kinoshita, Y., Oshiro, Y., . . . Tsuneyoshi, M. (2001). Subcellular distribution of cytokeratin and vimentin in malignant rhabdoid tumor:

- three-dimensional imaging with confocal laser scanning microscopy and double immunofluorescence. *Mod Pathol*, 14(9), 854-861. doi:10.1038/modpathol.3880401
- Italiano, A., Soria, J. C., Toulmonde, M., Michot, J. M., Lucchesi, C., Varga, A., . . . Ribrag, V. (2018). Tazemetostat, an EZH2 inhibitor, in relapsed or refractory B-cell non-Hodgkin lymphoma and advanced solid tumours: a first-in-human, open-label, phase 1 study. *Lancet Oncol*, 19(5), 649-659. doi:10.1016/S1470-2045(18)30145-1
- James, N. D., de Bono, J. S., Spears, M. R., Clarke, N. W., Mason, M. D., Dearnaley, D. P., . . . Investigators, S. (2017). Abiraterone for Prostate Cancer Not Previously Treated with Hormone Therapy. *N Engl J Med*. doi:10.1056/NEJMoa1702900
- Jelinic, P., Mueller, J. J., Olvera, N., Dao, F., Scott, S. N., Shah, R., . . . Levine, D. A. (2014). Recurrent SMARCA4 mutations in small cell carcinoma of the ovary. *Nat Genet*, 46(5), 424-426. doi:10.1038/ng.2922
- Jelinic, P., Ricca, J., Van Oudenhove, E., Olvera, N., Merghoub, T., Levine, D. A., & Zamarin, D. (2018). Immune-Active Microenvironment in Small Cell Carcinoma of the Ovary, Hypercalcemic Type: Rationale for Immune Checkpoint Blockade. *J Natl Cancer Inst*, 110(7), 787-790. doi:10.1093/jnci/djx277
- Jiang, Z., Tang, Y., Zhao, X., Zhang, M., Donovan, D. M., & Tian, X. C. (2015). Knockdown of Brm and Baf170, Components of Chromatin Remodeling Complex, Facilitates Reprogramming of Somatic Cells. *Stem Cells Dev*, 24(19), 2328-2336. doi:10.1089/scd.2015.0069
- Jin, M. L., Kim, Y. W., & Jeong, K. W. (2018). BAF53A regulates androgen receptor-mediated gene expression and proliferation in LNCaP cells. *Biochem Biophys Res Commun*, 505(2), 618-623. doi:10.1016/j.bbrc.2018.09.149
- Jo, V. Y., & Fletcher, C. D. (2015). Epithelioid malignant peripheral nerve sheath tumor: clinicopathologic analysis of 63 cases. *Am J Surg Pathol*, 39(5), 673-682. doi:10.1097/PAS.0000000000000379
- Jo, V. Y., & Fletcher, C. D. M. (2017). SMARCB1/INI1 Loss in Epithelioid Schwannoma: A Clinicopathologic and Immunohistochemical Study of 65 Cases. *Am J Surg Pathol*, 41(8), 1013-1022. doi:10.1097/PAS.0000000000000849
- Johann, P. D., Erkek, S., Zapatka, M., Kerl, K., Buchhalter, I., Hovestadt, V., . . . Kool, M. (2016). Atypical Teratoid/Rhabdoid Tumors Are Comprised of Three Epigenetic Subgroups with Distinct Enhancer Landscapes. *Cancer Cell*, 29(3), 379-393. doi:10.1016/j.ccell.2016.02.001
- Jones, D. T., Jager, N., Kool, M., Zichner, T., Hutter, B., Sultan, M., . . . Lichter, P. (2012). Dissecting the genomic complexity underlying medulloblastoma. *Nature*, 488(7409), 100-105. doi:10.1038/nature11284
- Jones, S., Wang, T. L., Shih le, M., Mao, T. L., Nakayama, K., Roden, R., . . . Papadopoulos, N. (2010). Frequent mutations of chromatin remodeling gene ARID1A in ovarian clear cell carcinoma. *Science*, 330(6001), 228-231. doi:10.1126/science.1196333

- Jubierre, L., Soriano, A., Planells-Ferrer, L., Paris-Coderch, L., Tenbaum, S. P., Romero, O. A., . . . Segura, M. F. (2016). BRG1/SMARCA4 is essential for neuroblastoma cell viability through modulation of cell death and survival pathways. *Oncogene*, 35(39), 5179-5190. doi:10.1038/onc.2016.50
- Kadam, S., & Emerson, B. M. (2003). Transcriptional specificity of human SWI/SNF BRG1 and BRM chromatin remodeling complexes. *Mol Cell*, 11(2), 377-389. doi:10.1016/s1097-2765(03)00034-0
- Kadoch, C., & Crabtree, G. R. (2013). Reversible disruption of mSWI/SNF (BAF) complexes by the SS18-SSX oncogenic fusion in synovial sarcoma. *Cell*, 153(1), 71-85. doi:10.1016/j.cell.2013.02.036
- Kadoch, C., & Crabtree, G. R. (2015). Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics. *Sci Adv*, 1(5), e1500447. doi:10.1126/sciadv.1500447
- Kadoch, C., Hargreaves, D. C., Hodges, C., Elias, L., Ho, L., Ranish, J., & Crabtree, G. R. (2013). Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. *Nat Genet*, 45(6), 592-601. doi:10.1038/ng.2628
- Kahali, B., Gramling, S. J., Marquez, S. B., Thompson, K., Lu, L., & Reisman, D. (2014). Identifying targets for the restoration and reactivation of BRM. *Oncogene*, 33(5), 653-664. doi:10.1038/onc.2012.613
- Kahali, B., Yu, J., Marquez, S. B., Thompson, K. W., Liang, S. Y., Lu, L., & Reisman, D. (2014). The silencing of the SWI/SNF subunit and anticancer gene BRM in Rhabdoid tumors. *Oncotarget*, 5(10), 3316-3332. doi:10.18632/oncotarget.1945
- Kamoun, A., Cancel-Tassin, G., Fromont, G., Elarouci, N., Armenoult, L., Ayadi, M., . . . Cussenot, O. (2018). Comprehensive molecular classification of localized prostate adenocarcinoma reveals a tumour subtype predictive of non-aggressive disease. *Ann Oncol*, 29(8), 1814-1821. doi:10.1093/annonc/mdy224
- Karnezis, A. N., Hoang, L. N., Coatham, M., Ravn, S., Almadani, N., Tessier-Cloutier, B., . . . Lee, C. H. (2016). Loss of switch/sucrose non-fermenting complex protein expression is associated with dedifferentiation in endometrial carcinomas. *Mod Pathol*, 29(3), 302-314. doi:10.1038/modpathol.2015.155
- Karnezis, A. N., Wang, Y., Ramos, P., Hendricks, W. P., Oliva, E., D'Angelo, E., . . . Huntsman, D. G. (2016). Dual loss of the SWI/SNF complex ATPases SMARCA4/BRG1 and SMARCA2/BRM is highly sensitive and specific for small cell carcinoma of the ovary, hypercalcaemic type. *J Pathol*, 238(3), 389-400. doi:10.1002/path.4633
- Keenen, B., Qi, H., Saladi, S. V., Yeung, M., & de la Serna, I. L. (2010). Heterogeneous SWI/SNF chromatin remodeling complexes promote expression of microphthalmia-associated transcription factor target genes in melanoma. *Oncogene*, 29(1), 81-92. doi:10.1038/onc.2009.304

- Kidder, B. L., Palmer, S., & Knott, J. G. (2009). SWI/SNF-Brg1 regulates self-renewal and occupies core pluripotency-related genes in embryonic stem cells. *Stem Cells*, 27(2), 317-328. doi:10.1634/stemcells.2008-0710
- Kim, K. H., Kim, W., Howard, T. P., Vazquez, F., Tsherniak, A., Wu, J. N., . . . Roberts, C. W. (2015). SWI/SNF-mutant cancers depend on catalytic and non-catalytic activity of EZH2. *Nat Med*, 21(12), 1491-1496. doi:10.1038/nm.3968
- Kohashi, K., Oda, Y., Yamamoto, H., Tamiya, S., Oshiro, Y., Izumi, T., . . . Tsuneyoshi, M. (2008). SMARCB1/INI1 protein expression in round cell soft tissue sarcomas associated with chromosomal translocations involving EWS: a special reference to SMARCB1/INI1 negative variant extraskelatal myxoid chondrosarcoma. *Am J Surg Pathol*, 32(8), 1168-1174. doi:10.1097/PAS.0b013e318161781a
- Kohmoto, T., Masuda, K., Shoda, K., Takahashi, R., Ujio, S., Tange, S., . . . Imoto, I. (2020). Claudin-6 is a single prognostic marker and functions as a tumor-promoting gene in a subgroup of intestinal type gastric cancer. *Gastric Cancer*, 23(3), 403-417. doi:10.1007/s10120-019-01014-x
- Konieczkowski, D. J., Johannessen, C. M., & Garraway, L. A. (2018). A Convergence-Based Framework for Cancer Drug Resistance. *Cancer Cell*, 33(5), 801-815. doi:10.1016/j.ccell.2018.03.025
- Ku, S. Y., Rosario, S., Wang, Y., Mu, P., Seshadri, M., Goodrich, Z. W., . . . Goodrich, D. W. (2017). Rb1 and Trp53 cooperate to suppress prostate cancer lineage plasticity, metastasis, and antiandrogen resistance. *Science*, 355(6320), 78-83. doi:10.1126/science.aah4199
- Kupryjanczyk, J., Dansonka-Mieszkowska, A., Moes-Sosnowska, J., Plisiecka-Halasa, J., Szafron, L., Podgorska, A., . . . Spiewankiewicz, B. (2013). Ovarian small cell carcinoma of hypercalcemic type - evidence of germline origin and SMARCA4 gene inactivation. a pilot study. *Pol J Pathol*, 64(4), 238-246.
- Kurman RJ, C. M., Herrington CS, Young RH. (2014). *WHO Classification of Tumours of Female Reproductive Organs*
WHO Classification of Tumours, 4th Edition, Volume 6: IARC (Lyon).
- Kwon, Y. H., Jovanovic, A., Serfas, M. S., Kiyokawa, H., & Tyner, A. L. (2002). P21 functions to maintain quiescence of p27-deficient hepatocytes. *J Biol Chem*, 277(44), 41417-41422. doi:10.1074/jbc.M203388200
- Labrecque, M. P., Coleman, I. M., Brown, L. G., True, L. D., Kollath, L., Lakely, B., . . . Morrissey, C. (2019). Molecular profiling stratifies diverse phenotypes of treatment-refractory metastatic castration-resistant prostate cancer. *J Clin Invest*, 129(10), 4492-4505. doi:10.1172/JCI128212
- Lang, J. D., Hendricks, W. P. D., Orlando, K. A., Yin, H., Kiefer, J., Ramos, P., . . . Trent, J. M. (2018). Ponatinib Shows Potent Antitumor Activity in Small Cell Carcinoma of the Ovary Hypercalcemic Type (SCCOHT) through Multikinase Inhibition. *Clin Cancer Res*, 24(8), 1932-1943. doi:10.1158/1078-0432.CCR-17-1928

- Laurette, P., Coassolo, S., Davidson, G., Michel, I., Gambi, G., Yao, W., . . . Davidson, I. (2019). Chromatin remodellers Brg1 and Bptf are required for normal gene expression and progression of oncogenic Braf-driven mouse melanoma. *Cell Death Differ*. doi:10.1038/s41418-019-0333-6
- Laurette, P., Coassolo, S., Davidson, G., Michel, I., Gambi, G., Yao, W., . . . Davidson, I. (2020). Chromatin remodellers Brg1 and Bptf are required for normal gene expression and progression of oncogenic Braf-driven mouse melanoma. *Cell Death Differ*, 27(1), 29-43. doi:10.1038/s41418-019-0333-6
- Lawrence, M. S., Stojanov, P., Polak, P., Kryukov, G. V., Cibulskis, K., Sivachenko, A., . . . Getz, G. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*, 499(7457), 214-218. doi:10.1038/nature12213
- Le Loarer, F., Watson, S., Pierron, G., de Montpreville, V. T., Ballet, S., Firmin, N., . . . Tirode, F. (2015). SMARCA4 inactivation defines a group of undifferentiated thoracic malignancies transcriptionally related to BAF-deficient sarcomas. *Nat Genet*, 47(10), 1200-1205. doi:10.1038/ng.3399
- Le Loarer, F., Zhang, L., Fletcher, C. D., Ribeiro, A., Singer, S., Italiano, A., . . . Antonescu, C. R. (2014). Consistent SMARCB1 homozygous deletions in epithelioid sarcoma and in a subset of myoepithelial carcinomas can be reliably detected by FISH in archival material. *Genes Chromosomes Cancer*, 53(6), 475-486. doi:10.1002/gcc.22159
- Lee, E. K., Esselen, K. M., Kolin, D. L., Lee, L. J., Matulonis, U. A., & Konstantinopoulos, P. A. (2020). Combined CDK4/6 and PD-1 Inhibition in Refractory SMARCA4-Deficient Small-Cell Carcinoma of the Ovary, Hypercalcemic Type. *JCO Precis Oncol*, 4, 736-742. doi:10.1200/PO.20.00063
- LeGouy, E., Thompson, E. M., Muchardt, C., & Renard, J. P. (1998). Differential preimplantation regulation of two mouse homologues of the yeast SWI2 protein. *Dev Dyn*, 212(1), 38-48. doi:10.1002/(SICI)1097-0177(199805)212:1<38::AID-AJA4>3.0.CO;2-3
- Leis, O., Eguiara, A., Lopez-Arribillaga, E., Alberdi, M. J., Hernandez-Garcia, S., Elorriaga, K., . . . Martin, A. G. (2012). Sox2 expression in breast tumours and activation in breast cancer stem cells. *Oncogene*, 31(11), 1354-1365. doi:10.1038/onc.2011.338
- Leruste, A., Tosello, J., Ramos, R. N., Tauziede-Espariat, A., Brohard, S., Han, Z. Y., . . . Bourdeaut, F. (2019). Clonally Expanded T Cells Reveal Immunogenicity of Rhabdoid Tumors. *Cancer Cell*, 36(6), 597-612 e598. doi:10.1016/j.ccell.2019.10.008
- Lessard, J., Wu, J. I., Ranish, J. A., Wan, M., Winslow, M. M., Stahl, B. T., . . . Crabtree, G. R. (2007). An essential switch in subunit composition of a chromatin remodeling complex during neural development. *Neuron*, 55(2), 201-215. doi:10.1016/j.neuron.2007.06.019
- Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdottir, H., Tamayo, P., & Mesirov, J. P. (2011). Molecular signatures database (MSigDB) 3.0. *Bioinformatics*, 27(12), 1739-1740. doi:10.1093/bioinformatics/btr260

- Lickert, H., Takeuchi, J. K., Von Both, I., Walls, J. R., McAuliffe, F., Adamson, S. L., . . . Bruneau, B. G. (2004). Baf60c is essential for function of BAF chromatin remodelling complexes in heart development. *Nature*, *432*(7013), 107-112. doi:10.1038/nature03071
- Lin, D., Wyatt, A. W., Xue, H., Wang, Y., Dong, X., Haegert, A., . . . Wang, Y. (2014). High fidelity patient-derived xenografts for accelerating prostate cancer discovery and drug development. *Cancer Res*, *74*(4), 1272-1283. doi:10.1158/0008-5472.CAN-13-2921-T
- Lin, D. I., Allen, J. M., Hecht, J. L., Killian, J. K., Ngo, N. T., Edgerly, C., . . . Elvin, J. A. (2019). SMARCA4 inactivation defines a subset of undifferentiated uterine sarcomas with rhabdoid and small cell features and germline mutation association. *Mod Pathol*, *32*(11), 1675-1687. doi:10.1038/s41379-019-0303-z
- Lin, D. I., Chudnovsky, Y., Duggan, B., Zajchowski, D., Greenbowe, J., Ross, J. S., . . . Elvin, J. A. (2017). Comprehensive genomic profiling reveals inactivating SMARCA4 mutations and low tumor mutational burden in small cell carcinoma of the ovary, hypercalcemic-type. *Gynecol Oncol*, *147*(3), 626-633. doi:10.1016/j.ygyno.2017.09.031
- Link, K. A., Balasubramaniam, S., Sharma, A., Comstock, C. E., Godoy-Tundidor, S., Powers, N., . . . Knudsen, K. E. (2008). Targeting the BAF57 SWI/SNF subunit in prostate cancer: a novel platform to control androgen receptor activity. *Cancer Res*, *68*(12), 4551-4558. doi:10.1158/0008-5472.CAN-07-6392
- Liu, G., Cuffe, S., Liang, S., Azad, A. K., Cheng, L., Brhane, Y., . . . Reisman, D. N. (2017). BRM Promoter Polymorphisms and Survival of Advanced Non-Small Cell Lung Cancer Patients in the Princess Margaret Cohort and CCTG BR.24 Trial. *Clin Cancer Res*, *23*(10), 2460-2470. doi:10.1158/1078-0432.CCR-16-1640
- Liu, G., Gramling, S., Munoz, D., Cheng, D., Azad, A. K., Mirshams, M., . . . Reisman, D. (2011). Two novel BRM insertion promoter sequence variants are associated with loss of BRM expression and lung cancer risk. *Oncogene*, *30*(29), 3295-3304. doi:10.1038/onc.2011.81
- Love, C., Sun, Z., Jima, D., Li, G., Zhang, J., Miles, R., . . . Dave, S. S. (2012). The genetic landscape of mutations in Burkitt lymphoma. *Nat Genet*, *44*(12), 1321-1325. doi:10.1038/ng.2468
- Lu, B., & Shi, H. (2019). An In-Depth Look at Small Cell Carcinoma of the Ovary, Hypercalcemic Type (SCCOHT): Clinical Implications from Recent Molecular Findings. *J Cancer*, *10*(1), 223-237. doi:10.7150/jca.26978
- Lu, C., & Allis, C. D. (2017). SWI/SNF complex in cancer. *Nat Genet*, *49*(2), 178-179. doi:10.1038/ng.3779
- Magnani, L., & Cabot, R. A. (2009). Manipulation of SMARCA2 and SMARCA4 transcript levels in porcine embryos differentially alters development and expression of SMARCA1, SOX2, NANOG, and EIF1. *Reproduction*, *137*(1), 23-33. doi:10.1530/REP-08-0335
- Marshall, T. W., Link, K. A., Petre-Draviam, C. E., & Knudsen, K. E. (2003). Differential requirement of SWI/SNF for androgen receptor activity. *J Biol Chem*, *278*(33), 30605-30613. doi:10.1074/jbc.M304582200

- Mashtalir, N., D'Avino, A. R., Michel, B. C., Luo, J., Pan, J., Otto, J. E., . . . Kadoch, C. (2018). Modular Organization and Assembly of SWI/SNF Family Chromatin Remodeling Complexes. *Cell*, *175*(5), 1272-1288 e1220. doi:10.1016/j.cell.2018.09.032
- Maslah-Planchon, J., Bieche, I., Guinebretiere, J. M., Bourdeaut, F., & Delattre, O. (2015). SWI/SNF chromatin remodeling and human malignancies. *Annu Rev Pathol*, *10*, 145-171. doi:10.1146/annurev-pathol-012414-040445
- Maslah-Planchon, J., Machet, M. C., Freneaux, P., Jourdain, A., Mortemousque, I., Rais, K. A., . . . Bourdeaut, F. (2016). SMARCA4-Mutated Atypical Teratoid/Rhabdoid Tumor with Retained BRG1 Expression. *Pediatr Blood Cancer*, *63*(3), 568-569. doi:10.1002/pbc.25772
- McBride, M. J., Pulice, J. L., Beird, H. C., Ingram, D. R., D'Avino, A. R., Shern, J. F., . . . Kadoch, C. (2018). The SS18-SSX Fusion Oncoprotein Hijacks BAF Complex Targeting and Function to Drive Synovial Sarcoma. *Cancer Cell*, *33*(6), 1128-1141 e1127. doi:10.1016/j.ccell.2018.05.002
- McCluggage, W. G., Witkowski, L., Clarke, B. A., & Foulkes, W. D. (2017). Clinical, morphological and immunohistochemical evidence that small-cell carcinoma of the ovary of hypercalcaemic type (SCCOHT) may be a primitive germ-cell neoplasm. *Histopathology*, *70*(7), 1147-1154. doi:10.1111/his.13177
- McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., . . . Cunningham, F. (2016). The Ensembl Variant Effect Predictor. *Genome Biology*, *17*(1), 122. doi:10.1186/s13059-016-0974-4
- Medina, P. P., Carretero, J., Fraga, M. F., Esteller, M., Sidransky, D., & Sanchez-Cespedes, M. (2004). Genetic and epigenetic screening for gene alterations of the chromatin-remodeling factor, SMARCA4/BRG1, in lung tumors. *Genes Chromosomes Cancer*, *41*(2), 170-177. doi:10.1002/gcc.20068
- Memon, A. A., Sorensen, B. S., Melgard, P., Fokdal, L., Thykjaer, T., & Nexø, E. (2004). Expression of HER3, HER4 and their ligand heregulin-4 is associated with better survival in bladder cancer patients. *Br J Cancer*, *91*(12), 2034-2041. doi:10.1038/sj.bjc.6602251
- Metzger, A. L., Abel, S., Wegner, R. E., Fuhrer, R., Mao, S., Miller, R., . . . Horne, Z. D. (2019). Patterns of care and outcomes in small cell carcinoma of the prostate: A national cancer database analysis. *Prostate*, *79*(12), 1457-1461. doi:10.1002/pros.23864
- Michel, B. C., D'Avino, A. R., Cassel, S. H., Mashtalir, N., McKenzie, Z. M., McBride, M. J., . . . Kadoch, C. (2018). A non-canonical SWI/SNF complex is a synthetic lethal target in cancers driven by BAF complex perturbation. *Nat Cell Biol*, *20*(12), 1410-1420. doi:10.1038/s41556-018-0221-1
- Mittal, P., & Roberts, C. W. M. (2020). The SWI/SNF complex in cancer - biology, biomarkers and therapy. *Nat Rev Clin Oncol*. doi:10.1038/s41571-020-0357-3
- Mobley, B. C., McKenney, J. K., Bangs, C. D., Callahan, K., Yeom, K. W., Schneppenheim, R., . . . Vogel, H. (2010). Loss of SMARCB1/INI1 expression in poorly differentiated chordomas. *Acta Neuropathol*, *120*(6), 745-753. doi:10.1007/s00401-010-0767-x

- Mu, P., Zhang, Z., Benelli, M., Karthaus, W. R., Hoover, E., Chen, C. C., . . . Sawyers, C. L. (2017). SOX2 promotes lineage plasticity and antiandrogen resistance in TP53- and RB1-deficient prostate cancer. *Science*, 355(6320), 84-88. doi:10.1126/science.aah4307
- Muchardt, C., & Yaniv, M. (2001). When the SWI/SNF complex remodels...the cell cycle. *Oncogene*, 20(24), 3067-3075. doi:10.1038/sj.onc.1204331
- Muscat, A., Popovski, D., Jayasekara, W. S., Rossello, F. J., Ferguson, M., Marini, K. D., . . . Ashley, D. M. (2016). Low-Dose Histone Deacetylase Inhibitor Treatment Leads to Tumor Growth Arrest and Multi-Lineage Differentiation of Malignant Rhabdoid Tumors. *Clin Cancer Res*, 22(14), 3560-3570. doi:10.1158/1078-0432.CCR-15-2260
- Muthuswami, R., Bailey, L., Rakesh, R., Imbalzano, A. N., Nickerson, J. A., & Hockensmith, J. W. (2019). BRG1 is a prognostic indicator and a potential therapeutic target for prostate cancer. *J Cell Physiol*. doi:10.1002/jcp.28161
- Neugeborn, L., & Carlson, M. (1984). Genes affecting the regulation of SUC2 gene expression by glucose repression in *Saccharomyces cerevisiae*. *Genetics*, 108(4), 845-858.
- Nguyen, N. Y., Maxwell, M. J., Ooms, L. M., Davies, E. M., Hilton, A. A., Collinge, J. E., . . . Curtis, D. J. (2011). An ENU-induced mouse mutant of SHIP1 reveals a critical role of the stem cell isoform for suppression of macrophage activation. *Blood*, 117(20), 5362-5371. doi:10.1182/blood-2011-01-331041
- Nikolayeva, O., & Robinson, M. D. (2014). edgeR for differential RNA-seq and ChIP-seq analysis: an application to stem cell biology. *Methods Mol Biol*, 1150, 45-79. doi:10.1007/978-1-4939-0512-6_3
- O'Keefe, C., McDevitt, M. A., & Maciejewski, J. P. (2010). Copy neutral loss of heterozygosity: a novel chromosomal lesion in myeloid malignancies. *Blood*, 115(14), 2731-2739. doi:10.1182/blood-2009-10-201848
- Oike, T., Ogiwara, H., Tominaga, Y., Ito, K., Ando, O., Tsuta, K., . . . Kohno, T. (2013). A synthetic lethality-based strategy to treat cancers harboring a genetic deficiency in the chromatin remodeling factor BRG1. *Cancer Res*, 73(17), 5508-5518. doi:10.1158/0008-5472.CAN-12-4593
- Okuno, K., Ohta, S., Kato, H., Taga, T., Sugita, K., & Takeuchi, Y. (2010). Expression of neural stem cell markers in malignant rhabdoid tumor cell lines. *Oncol Rep*, 23(2), 485-492.
- Ooi, L., Belyaev, N. D., Miyake, K., Wood, I. C., & Buckley, N. J. (2006). BRG1 chromatin remodeling activity is required for efficient chromatin binding by repressor element 1-silencing transcription factor (REST) and facilitates REST-mediated repression. *J Biol Chem*, 281(51), 38974-38980. doi:10.1074/jbc.M605370200
- Pan, J., McKenzie, Z. M., D'Avino, A. R., Mashtalir, N., Lareau, C. A., St Pierre, R., . . . Kadoch, C. (2019). The ATPase module of mammalian SWI/SNF family complexes mediates subcomplex identity and catalytic activity-independent genomic targeting. *Nat Genet*, 51(4), 618-626. doi:10.1038/s41588-019-0363-5

- Panwalkar, P., Pratt, D., Chung, C., Dang, D., Le, P., Martinez, D., . . . Venneti, S. (2020). SWI/SNF complex heterogeneity is related to polyphenotypic differentiation, prognosis, and immune response in rhabdoid tumors. *Neuro Oncol*, 22(6), 785-796. doi:10.1093/neuonc/noaa004
- Pautier, P., Ribrag, V., Duvillard, P., Rey, A., Elghissassi, I., Sillet-Bach, I., . . . Lhomme, C. (2007). Results of a prospective dose-intensive regimen in 27 patients with small cell carcinoma of the ovary of the hypercalcemic type. *Ann Oncol*, 18(12), 1985-1989. doi:10.1093/annonc/mdm376
- Perret, R., Chalabreysse, L., Watson, S., Serre, I., Garcia, S., Forest, F., . . . Le Loarer, F. (2019). SMARCA4-deficient Thoracic Sarcomas: Clinicopathologic Study of 30 Cases With an Emphasis on Their Nosology and Differential Diagnoses. *Am J Surg Pathol*, 43(4), 455-465. doi:10.1097/PAS.0000000000001188
- Peterson, C. L., & Herskowitz, I. (1992). Characterization of the yeast SWI1, SWI2, and SWI3 genes, which encode a global activator of transcription. *Cell*, 68(3), 573-583. doi:10.1016/0092-8674(92)90192-f
- Piccolo, S. R., Sun, Y., Campbell, J. D., Lenburg, M. E., Bild, A. H., & Johnson, W. E. (2012). A single-sample microarray normalization method to facilitate personalized-medicine workflows. *Genomics*, 100(6), 337-344. doi:10.1016/j.ygeno.2012.08.003
- Prandi, D., Baca, S. C., Romanel, A., Barbieri, C. E., Mosquera, J.-M., Fontugne, J., . . . Demichelis, F. (2014). Unraveling the clonal hierarchy of somatic genomic aberrations. *Genome Biology*, 15(8), 439. doi:10.1186/s13059-014-0439-6
- Prasad, P., Ronnerblad, M., Arner, E., Itoh, M., Kawaji, H., Lassmann, T., . . . consortium, F. (2014). High-throughput transcription profiling identifies putative epigenetic regulators of hematopoiesis. *Blood*, 123(17), e46-57. doi:10.1182/blood-2013-02-483537
- Prensner, J. R., Chen, W., Han, S., Iyer, M. K., Cao, Q., Kothari, V., . . . Feng, F. Y. (2014). The long non-coding RNA PCAT-1 promotes prostate cancer cell proliferation through cMyc. *Neoplasia*, 16(11), 900-908. doi:10.1016/j.neo.2014.09.001
- Prensner, J. R., Iyer, M. K., Sahu, A., Asangani, I. A., Cao, Q., Patel, L., . . . Chinnaiyan, A. M. (2013). The long noncoding RNA SCHLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. *Nat Genet*, 45(11), 1392-1398. doi:10.1038/ng.2771
- Puca, L., Bareja, R., Prandi, D., Shaw, R., Benelli, M., Karthaus, W. R., . . . Beltran, H. (2018). Patient derived organoids to model rare prostate cancer phenotypes. *Nat Commun*, 9(1), 2404. doi:10.1038/s41467-018-04495-z
- Pugh, T. J., Weeraratne, S. D., Archer, T. C., Pomeranz Krummel, D. A., Auclair, D., Bochicchio, J., . . . Cho, Y. J. (2012). Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations. *Nature*, 488(7409), 106-110. doi:10.1038/nature11329
- Raab, J. R., Smith, K. N., Spear, C. C., Manner, C. J., Calabrese, J. M., & Magnuson, T. (2019). SWI/SNF remains localized to chromatin in the presence of SCHLAP1. *Nat Genet*, 51(1), 26-29. doi:10.1038/s41588-018-0272-z

- Rago, F., DiMare, M. T., Elliott, G., Ruddy, D. A., Sovath, S., Kerr, G., . . . Jagani, Z. (2019). Degron mediated BRM/SMARCA2 depletion uncovers novel combination partners for treatment of BRG1/SMARCA4-mutant cancers. *Biochem Biophys Res Commun*, 508(1), 109-116. doi:10.1016/j.bbrc.2018.09.009
- Ramalingam, P., Croce, S., & McCluggage, W. G. (2017). Loss of expression of SMARCA4 (BRG1), SMARCA2 (BRM) and SMARCB1 (INI1) in undifferentiated carcinoma of the endometrium is not uncommon and is not always associated with rhabdoid morphology. *Histopathology*, 70(3), 359-366. doi:10.1111/his.13091
- Ramos, P., Karnezis, A. N., Craig, D. W., Sekulic, A., Russell, M. L., Hendricks, W. P., . . . Trent, J. M. (2014). Small cell carcinoma of the ovary, hypercalcemic type, displays frequent inactivating germline and somatic mutations in SMARCA4. *Nat Genet*, 46(5), 427-429. doi:10.1038/ng.2928
- Rawla, P. (2019). Epidemiology of Prostate Cancer. *World J Oncol*, 10(2), 63-89. doi:10.14740/wjon1191
- Reyes, J. C., Barra, J., Muchardt, C., Camus, A., Babinet, C., & Yaniv, M. (1998). Altered control of cellular proliferation in the absence of mammalian brahma (SNF2alpha). *EMBO J*, 17(23), 6979-6991. doi:10.1093/emboj/17.23.6979
- Rickman, D. S., Beltran, H., Demichelis, F., & Rubin, M. A. (2017). Biology and evolution of poorly differentiated neuroendocrine tumors. *Nat Med*, 23(6), 1-10. doi:10.1038/nm.4341
- Roberts, C. W., Galusha, S. A., McMenamin, M. E., Fletcher, C. D., & Orkin, S. H. (2000). Haploinsufficiency of Snf5 (integrase interactor 1) predisposes to malignant rhabdoid tumors in mice. *Proc Natl Acad Sci U S A*, 97(25), 13796-13800. doi:10.1073/pnas.250492697
- Roberts, C. W., Leroux, M. M., Fleming, M. D., & Orkin, S. H. (2002). Highly penetrant, rapid tumorigenesis through conditional inversion of the tumor suppressor gene Snf5. *Cancer Cell*, 2(5), 415-425. doi:10.1016/s1535-6108(02)00185-x
- Robinson, D., Van Allen, E. M., Wu, Y. M., Schultz, N., Lonigro, R. J., Mosquera, J. M., . . . Chinnaiyan, A. M. (2015). Integrative Clinical Genomics of Advanced Prostate Cancer. *Cell*, 162(2), 454. doi:10.1016/j.cell.2015.06.053
- Ross, A. E., Johnson, M. H., Yousefi, K., Davicioni, E., Netto, G. J., Marchionni, L., . . . Schaeffer, E. M. (2016). Tissue-based Genomics Augments Post-prostatectomy Risk Stratification in a Natural History Cohort of Intermediate- and High-Risk Men. *Eur Urol*, 69(1), 157-165. doi:10.1016/j.eururo.2015.05.042
- Roy, N., Malik, S., Villanueva, K. E., Urano, A., Lu, X., Von Figura, G., . . . Hebrok, M. (2015). Brg1 promotes both tumor-suppressive and oncogenic activities at distinct stages of pancreatic cancer formation. *Genes Dev*, 29(6), 658-671. doi:10.1101/gad.256628.114
- Rudin, C. M., Poirier, J. T., Byers, L. A., Dive, C., Dowlati, A., George, J., . . . Gazdar, A. F. (2019). Molecular subtypes of small cell lung cancer: a synthesis of human and mouse model data. *Nat Rev Cancer*, 19(5), 289-297. doi:10.1038/s41568-019-0133-9

- Sandoval, G. J., Pulice, J. L., Pakula, H., Schenone, M., Takeda, D. Y., Pop, M., . . . Kadoch, C. (2018). Binding of TMPRSS2-ERG to BAF Chromatin Remodeling Complexes Mediates Prostate Oncogenesis. *Mol Cell*, *71*(4), 554-566 e557. doi:10.1016/j.molcel.2018.06.040
- Santen, G. W., Aten, E., Sun, Y., Almomani, R., Gilissen, C., Nielsen, M., . . . Kriek, M. (2012). Mutations in SWI/SNF chromatin remodeling complex gene ARID1B cause Coffin-Siris syndrome. *Nat Genet*, *44*(4), 379-380. doi:10.1038/ng.2217
- Saunders, J., Ingley, K., Wang, X. Q., Harvey, M., Armstrong, L., Ng, T., . . . Bush, J. (2020). Loss of BRG1 (SMARCA4) Immunoexpression in a Pediatric Non-Central Nervous System Tumor Cohort. *Pediatr Dev Pathol*, *23*(2), 132-138. doi:10.1177/1093526619869154
- Schaefer, I. M., & Hornick, J. L. (2020). SWI/SNF complex-deficient soft tissue neoplasms: An update. *Semin Diagn Pathol*. doi:10.1053/j.semdp.2020.05.005
- Scher, H. I., Fizazi, K., Saad, F., Taplin, M. E., Sternberg, C. N., Miller, K., . . . Investigators, A. (2012). Increased survival with enzalutamide in prostate cancer after chemotherapy. *N Engl J Med*, *367*(13), 1187-1197. doi:10.1056/NEJMoa1207506
- Schneppenheim, R., Fruhwald, M. C., Gesk, S., Hasselblatt, M., Jeibmann, A., Kordes, U., . . . Siebert, R. (2010). Germline nonsense mutation and somatic inactivation of SMARCA4/BRG1 in a family with rhabdoid tumor predisposition syndrome. *Am J Hum Genet*, *86*(2), 279-284. doi:10.1016/j.ajhg.2010.01.013
- Schoffski P., A. M., Stracchiotti S., Davis L.E., Villalobos V.M., . . . , Ho P.T.C. and Goudner M.M. (2017). Phase 2 multicenter study of the EZH2 inhibitor tazemetostat in adults with synovial sarcoma (NCT02601950). *Journal of Clinical Oncology*, *35*:15_suppl, 11057-11057.
- Sekiguchi, F., Tsurusaki, Y., Okamoto, N., Teik, K. W., Mizuno, S., Suzumura, H., . . . Matsumoto, N. (2019). Genetic abnormalities in a large cohort of Coffin-Siris syndrome patients. *J Hum Genet*, *64*(12), 1173-1186. doi:10.1038/s10038-019-0667-4
- Sequist, L. V., Waltman, B. A., Dias-Santagata, D., Digumarthy, S., Turke, A. B., Fidias, P., . . . Engelman, J. A. (2011). Genotypic and histological evolution of lung cancers acquiring resistance to EGFR inhibitors. *Sci Transl Med*, *3*(75), 75ra26. doi:10.1126/scitranslmed.3002003
- Shain, A. H., & Pollack, J. R. (2013). The spectrum of SWI/SNF mutations, ubiquitous in human cancers. *PLoS One*, *8*(1), e55119. doi:10.1371/journal.pone.0055119
- Shen, H., Powers, N., Saini, N., Comstock, C. E., Sharma, A., Weaver, K., . . . Knudsen, K. E. (2008). The SWI/SNF ATPase Brm is a gatekeeper of proliferative control in prostate cancer. *Cancer Res*, *68*(24), 10154-10162. doi:10.1158/0008-5472.CAN-08-1794
- Siegel, R. L., Miller, K. D., & Jemal, A. (2019). Cancer statistics, 2019. *CA Cancer J Clin*, *69*(1), 7-34. doi:10.3322/caac.21551

- Singhal, N., Graumann, J., Wu, G., Arauzo-Bravo, M. J., Han, D. W., Greber, B., . . . Scholer, H. R. (2010). Chromatin-Remodeling Components of the BAF Complex Facilitate Reprogramming. *Cell*, *141*(6), 943-955. doi:10.1016/j.cell.2010.04.037
- Small, E. J. e. a. (2020). *Intermediate atypical carcinoma (IAC): A discrete subtype of metastatic castration-resistant prostate cancer (mCRPC) suggesting that treatment-associated small cell/neuroendocrine prostate cancer (t-SCNC) may evolve from mCRPC adenocarcinoma (adeno)—Results from the SU2C/PCF/AACR West Coast Prostate Cancer Dream Team (WCDT).*
- Smith, M. J., O'Sullivan, J., Bhaskar, S. S., Hadfield, K. D., Poke, G., Caird, J., . . . Evans, D. G. (2013). Loss-of-function mutations in SMARCE1 cause an inherited disorder of multiple spinal meningiomas. *Nat Genet*, *45*(3), 295-298. doi:10.1038/ng.2552
- Soldi, R., Ghosh Halder, T., Weston, A., Thode, T., Drenner, K., Lewis, R., . . . Sharma, S. (2020). The novel reversible LSD1 inhibitor SP-2577 promotes anti-tumor immunity in SWI/SNF complex mutated ovarian cancer. *PLoS One*, *15*(7), e0235705. doi:10.1371/journal.pone.0235705
- Spahn, M., Kneitz, S., Scholz, C. J., Stenger, N., Rudiger, T., Strobel, P., . . . Kneitz, B. (2010). Expression of microRNA-221 is progressively reduced in aggressive prostate cancer and metastasis and predicts clinical recurrence. *Int J Cancer*, *127*(2), 394-403. doi:10.1002/ijc.24715
- Spratt, D. E., Yousefi, K., Deheshi, S., Ross, A. E., Den, R. B., Schaeffer, E. M., . . . Feng, F. Y. (2017). Individual Patient-Level Meta-Analysis of the Performance of the Decipher Genomic Classifier in High-Risk Men After Prostatectomy to Predict Development of Metastatic Disease. *J Clin Oncol*, *35*(18), 1991-1998. doi:10.1200/JCO.2016.70.2811
- Stahl, B. T., Tang, J., Wu, W., Sun, A., Gitler, A. D., Yoo, A. S., & Crabtree, G. R. (2013). Kinetic analysis of npBAF to nBAF switching reveals exchange of SS18 with CREST and integration with neural developmental pathways. *J Neurosci*, *33*(25), 10348-10361. doi:10.1523/JNEUROSCI.1258-13.2013
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., . . . Mesirov, J. P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*, *102*(43), 15545-15550. doi:10.1073/pnas.0506580102
- Sun, X., Wang, S. C., Wei, Y., Luo, X., Jia, Y., Li, L., . . . Zhu, H. (2017). Arid1a Has Context-Dependent Oncogenic and Tumor Suppressor Functions in Liver Cancer. *Cancer Cell*, *32*(5), 574-589 e576. doi:10.1016/j.ccell.2017.10.007
- Takahashi, K., & Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, *126*(4), 663-676. doi:10.1016/j.cell.2006.07.024
- Talevich, E., Shain, A. H., Botton, T., & Bastian, B. C. (2016). CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *PLOS Computational Biology*, *12*(4), e1004873. doi:10.1371/journal.pcbi.1004873

- Tamkun, J. W., Deuring, R., Scott, M. P., Kissinger, M., Pattatucci, A. M., Kaufman, T. C., & Kennison, J. A. (1992). *brhma*: a regulator of *Drosophila* homeotic genes structurally related to the yeast transcriptional activator SNF2/SWI2. *Cell*, *68*(3), 561-572. doi:10.1016/0092-8674(92)90191-e
- Tang, A., Gao, K., Chu, L., Zhang, R., Yang, J., & Zheng, J. (2017). Aurora kinases: novel therapy targets in cancers. *Oncotarget*, *8*(14), 23937-23954. doi:10.18632/oncotarget.14893
- Tessier-Cloutier, B., Schaeffer, D. F., Bacani, J., Marginean, C. E., Kalloger, S., Kobel, M., & Lee, C. H. (2020). Loss of switch/sucrose non-fermenting complex protein expression in undifferentiated gastrointestinal and pancreatic carcinomas. *Histopathology*, *77*(1), 46-54. doi:10.1111/his.14096
- Theruvath, J., Sotillo, E., Mount, C. W., Graef, C. M., Delaidelli, A., Heitzeneder, S., . . . Mackall, C. L. (2020). Locoregionally administered B7-H3-targeted CAR T cells for treatment of atypical teratoid/rhabdoid tumors. *Nat Med*, *26*(5), 712-719. doi:10.1038/s41591-020-0821-8
- Thompson, K. W., Marquez, S. B., Lu, L., & Reisman, D. (2015). Induction of functional Brm protein from Brm knockout mice. *Oncoscience*, *2*(4), 349-361. doi:10.18632/oncoscience.153
- Tischkowitz, M., Huang, S., Banerjee, S., Hague, J., Hendricks, W. P. D., Huntsman, D. G., . . . Weissman, B. E. (2020). Small Cell Carcinoma of the Ovary, Hypercalcaemic Type - genetics, new treatment targets and current management guidelines. *Clin Cancer Res*. doi:10.1158/1078-0432.CCR-19-3797
- Trouche, D., Le Chalony, C., Muchardt, C., Yaniv, M., & Kouzarides, T. (1997). RB and hbrm cooperate to repress the activation functions of E2F1. *Proc Natl Acad Sci U S A*, *94*(21), 11268-11273. doi:10.1073/pnas.94.21.11268
- Tsurusaki, Y., Okamoto, N., Ohashi, H., Kosho, T., Imai, Y., Hibi-Ko, Y., . . . Matsumoto, N. (2012). Mutations affecting components of the SWI/SNF complex cause Coffin-Siris syndrome. *Nat Genet*, *44*(4), 376-378. doi:10.1038/ng.2219
- Ugur Sahin, D. J., Frederik Marme, Athanasios Mavratzas, Jürgen Krauss, Jacques De Greve, Ignace Vergote, and Oezlem Tureci. (2015). First-in-human phase I/II dose-escalation study of IMAB027 in patients with recurrent advanced ovarian cancer (OVAR): Preliminary data of phase I part. *Journal of Clinical Oncology*, *33*:15_suppl, 5537-5537.
- Van Houdt, J. K., Nowakowska, B. A., Sousa, S. B., van Schaik, B. D., Seuntjens, E., Avonce, N., . . . Vermeesch, J. R. (2012). Heterozygous missense mutations in SMARCA2 cause Nicolaides-Baraitser syndrome. *Nat Genet*, *44*(4), 445-449, S441. doi:10.1038/ng.1105
- Varela, I., Tarpey, P., Raine, K., Huang, D., Ong, C. K., Stephens, P., . . . Futreal, P. A. (2011). Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. *Nature*, *469*(7331), 539-542. doi:10.1038/nature09639
- Velcheti, V., Wong, K. K., & Sauntharajah, Y. (2019). EZH2 Inhibitors: Take It EZy, It Is All About Context. *Cancer Discov*, *9*(4), 472-475. doi:10.1158/2159-8290.CD-19-0090

- Venneti, S., Le, P., Martinez, D., Xie, S. X., Sullivan, L. M., Rorke-Adams, L. B., . . . Judkins, A. R. (2011). Malignant rhabdoid tumors express stem cell factors, which relate to the expression of EZH2 and Id proteins. *Am J Surg Pathol*, 35(10), 1463-1472. doi:10.1097/PAS.0b013e318224d2cd
- Versteeg, I., Sevenet, N., Lange, J., Rousseau-Merck, M. F., Ambros, P., Handgretinger, R., . . . Delattre, O. (1998). Truncating mutations of hSNF5/INI1 in aggressive paediatric cancer. *Nature*, 394(6689), 203-206. doi:10.1038/28212
- Vitte, J., Gao, F., Coppola, G., Judkins, A. R., & Giovannini, M. (2017). Timing of Smarcb1 and Nf2 inactivation determines schwannoma versus rhabdoid tumor development. *Nat Commun*, 8(1), 300. doi:10.1038/s41467-017-00346-5
- Vogel-Ciernia, A., Matheos, D. P., Barrett, R. M., Kramar, E. A., Azzawi, S., Chen, Y., . . . Wood, M. A. (2013). The neuron-specific chromatin regulatory subunit BAF53b is necessary for synaptic plasticity and memory. *Nat Neurosci*, 16(5), 552-561. doi:10.1038/nn.3359
- Wang, H. T., Yao, Y. H., Li, B. G., Tang, Y., Chang, J. W., & Zhang, J. (2014). Neuroendocrine Prostate Cancer (NEPC) progressing from conventional prostatic adenocarcinoma: factors associated with time to development of NEPC and survival from NEPC diagnosis-a systematic review and pooled analysis. *J Clin Oncol*, 32(30), 3383-3390. doi:10.1200/JCO.2013.54.3553
- Wang, J., Yin, J., Yang, Q., Ding, F., Chen, X., Li, B., & Tian, X. (2016). Human epidermal growth factor receptor 4 (HER4) is a favorable prognostic marker of breast cancer: a systematic review and meta-analysis. *Oncotarget*, 7(47), 76693-76703. doi:10.18632/oncotarget.12485
- Wang, J. R., Gramling, S. J., Goldstein, D. P., Cheng, D., Chen, D., Azad, A. K., . . . Liu, G. (2013). Association of two BRM promoter polymorphisms with head and neck squamous cell carcinoma risk. *Carcinogenesis*, 34(5), 1012-1017. doi:10.1093/carcin/bgt008
- Wang, L., Jin, X., Lin, D., Liu, Z., Zhang, X., Lu, Y., . . . Quan, C. (2013). Clinicopathologic significance of claudin-6, occludin, and matrix metalloproteinases -2 expression in ovarian carcinoma. *Diagn Pathol*, 8, 190. doi:10.1186/1746-1596-8-190
- Wang, L., Xue, Y., Shen, Y., Li, W., Cheng, Y., Yan, X., . . . Zeng, F. (2012). Claudin 6: a novel surface marker for characterizing mouse pluripotent stem cells. *Cell Res*, 22(6), 1082-1085. doi:10.1038/cr.2012.77
- Wang, X., Sansam, C. G., Thom, C. S., Metzger, D., Evans, J. A., Nguyen, P. T., & Roberts, C. W. (2009). Oncogenesis caused by loss of the SNF5 tumor suppressor is dependent on activity of BRG1, the ATPase of the SWI/SNF chromatin remodeling complex. *Cancer Res*, 69(20), 8094-8101. doi:10.1158/0008-5472.CAN-09-0733
- Wang, X., Wang, S., Troisi, E. C., Howard, T. P., Haswell, J. R., Wolf, B. K., . . . Roberts, C. W. M. (2019). BRD9 defines a SWI/SNF sub-complex and constitutes a specific vulnerability in malignant rhabdoid tumors. *Nat Commun*, 10(1), 1881. doi:10.1038/s41467-019-09891-7

- Wang, Y., Chen, S. Y., Karnezis, A. N., Colborne, S., Santos, N. D., Lang, J. D., . . . Huntsman, D. G. (2017). The histone methyltransferase EZH2 is a therapeutic target in small cell carcinoma of the ovary, hypercalcaemic type. *J Pathol*, *242*(3), 371-383. doi:10.1002/path.4912
- Wiegand, K. C., Shah, S. P., Al-Agha, O. M., Zhao, Y., Tse, K., Zeng, T., . . . Huntsman, D. G. (2010). ARID1A mutations in endometriosis-associated ovarian carcinomas. *N Engl J Med*, *363*(16), 1532-1543. doi:10.1056/NEJMoa1008433
- Wilson, B. G., Helming, K. C., Wang, X., Kim, Y., Vazquez, F., Jagani, Z., . . . Roberts, C. W. (2014). Residual complexes containing SMARCA2 (BRM) underlie the oncogenic drive of SMARCA4 (BRG1) mutation. *Mol Cell Biol*, *34*(6), 1136-1144. doi:10.1128/MCB.01372-13
- Wilson, B. G., Wang, X., Shen, X., McKenna, E. S., Lemieux, M. E., Cho, Y. J., . . . Roberts, C. W. (2010). Epigenetic antagonism between polycomb and SWI/SNF complexes during oncogenic transformation. *Cancer Cell*, *18*(4), 316-328. doi:10.1016/j.ccr.2010.09.006
- Witkowski, L., Carrot-Zhang, J., Albrecht, S., Fahiminiya, S., Hamel, N., Tomiak, E., . . . Foulkes, W. D. (2014). Germline and somatic SMARCA4 mutations characterize small cell carcinoma of the ovary, hypercalcemic type. *Nat Genet*, *46*(5), 438-443. doi:10.1038/ng.2931
- Witkowski, L., Goudie, C., Ramos, P., Boshari, T., Brunet, J. S., Karnezis, A. N., . . . Foulkes, W. D. (2016). The influence of clinical and genetic factors on patient outcome in small cell carcinoma of the ovary, hypercalcemic type. *Gynecol Oncol*, *141*(3), 454-460. doi:10.1016/j.ygyno.2016.03.013
- Witkowski, L., Lalonde, E., Zhang, J., Albrecht, S., Hamel, N., Cavallone, L., . . . Foulkes, W. D. (2013). Familial rhabdoid tumour 'avant la lettre'--from pathology review to exome sequencing and back again. *J Pathol*, *231*(1), 35-43. doi:10.1002/path.4225
- Wong, A. K., Shanahan, F., Chen, Y., Lian, L., Ha, P., Hendricks, K., . . . Lees, E. (2000). BRG1, a component of the SWI-SNF complex, is mutated in multiple human tumor cell lines. *Cancer Res*, *60*(21), 6171-6177.
- Wu, J. I., Lessard, J., Olave, I. A., Qiu, Z., Ghosh, A., Graef, I. A., & Crabtree, G. R. (2007). Regulation of dendritic development by neuron-specific chromatin remodeling complexes. *Neuron*, *56*(1), 94-108. doi:10.1016/j.neuron.2007.08.021
- Wu, Q., Madany, P., Dobson, J. R., Schnabl, J. M., Sharma, S., Smith, T. C., . . . Nickerson, J. A. (2016). The BRG1 chromatin remodeling enzyme links cancer cell metabolism and proliferation. *Oncotarget*, *7*(25), 38270-38281. doi:10.18632/oncotarget.9505
- Wu, Q., Sharma, S., Cui, H., LeBlanc, S. E., Zhang, H., Muthuswami, R., . . . Imbalzano, A. N. (2016). Targeting the chromatin remodeling enzyme BRG1 increases the efficacy of chemotherapy drugs in breast cancer cells. *Oncotarget*, *7*(19), 27158-27175. doi:10.18632/oncotarget.8384
- Xue, Y., Meehan, B., Macdonald, E., Venneti, S., Wang, X. Q. D., Witkowski, L., . . . Huang, S. (2019). CDK4/6 inhibitors target SMARCA4-determined cyclin D1 deficiency in

- hypercalcemic small cell carcinoma of the ovary. *Nat Commun*, 10(1), 558. doi:10.1038/s41467-018-06958-9
- Yamamichi, N., Yamamichi-Nishina, M., Mizutani, T., Watanabe, H., Minoguchi, S., Kobayashi, N., . . . Iba, H. (2005). The Brm gene suppressed at the post-transcriptional level in various human cell lines is inducible by transient HDAC inhibitor treatment, which exhibits antioncogenic potential. *Oncogene*, 24(35), 5471-5481. doi:10.1038/sj.onc.1208716
- Yan, Z., Wang, Z., Sharova, L., Sharov, A. A., Ling, C., Piao, Y., . . . Ko, M. S. (2008). BAF250B-associated SWI/SNF chromatin-remodeling complex is required to maintain undifferentiated mouse embryonic stem cells. *Stem Cells*, 26(5), 1155-1165. doi:10.1634/stemcells.2007-0846
- Yao, J., Chen, Y., Nguyen, D. T., Thompson, Z. J., Eroshkin, A. M., Nerlakanti, N., . . . Mahajan, K. (2019). The Homeobox gene, HOXB13, Regulates a Mitotic Protein-Kinase Interaction Network in Metastatic Prostate Cancers. *Sci Rep*, 9(1), 9715. doi:10.1038/s41598-019-46064-4
- Yoo, A. S., Staahl, B. T., Chen, L., & Crabtree, G. R. (2009). MicroRNA-mediated switching of chromatin-remodelling complexes in neural development. *Nature*, 460(7255), 642-646. doi:10.1038/nature08139
- Yoo, A. S., Sun, A. X., Li, L., Shcheglovitov, A., Portmann, T., Li, Y., . . . Crabtree, G. R. (2011). MicroRNA-mediated conversion of human fibroblasts to neurons. *Nature*, 476(7359), 228-231. doi:10.1038/nature10323
- Yoshida, A., Kobayashi, E., Kubo, T., Kodaira, M., Motoi, T., Motoi, N., . . . Hiraoka, N. (2017). Clinicopathological and molecular characterization of SMARCA4-deficient thoracic sarcomas with comparison to potentially related entities. *Mod Pathol*, 30(6), 797-809. doi:10.1038/modpathol.2017.11
- Young, R. H., Oliva, E., & Scully, R. E. (1994). Small cell carcinoma of the ovary, hypercalcemic type. A clinicopathological analysis of 150 cases. *Am J Surg Pathol*, 18(11), 1102-1116. doi:10.1097/00000478-199411000-00004
- Zehir, A., Benayed, R., Shah, R. H., Syed, A., Middha, S., Kim, H. R., . . . Berger, M. F. (2017). Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med*, 23(6), 703-713. doi:10.1038/nm.4333
- Zhang, X., Coleman, I. M., Brown, L. G., True, L. D., Kollath, L., Lucas, J. M., . . . Morrissey, C. (2015). SRRM4 Expression and the Loss of REST Activity May Promote the Emergence of the Neuroendocrine Phenotype in Castration-Resistant Prostate Cancer. *Clin Cancer Res*, 21(20), 4698-4708. doi:10.1158/1078-0432.CCR-15-0157
- Zhu, P., Wang, Y., Wu, J., Huang, G., Liu, B., Ye, B., . . . Fan, Z. (2016). LncBRM initiates YAP1 signalling activation to drive self-renewal of liver cancer stem cells. *Nat Commun*, 7, 13608. doi:10.1038/ncomms13608

- Zhu, Y., Liu, C., Cui, Y., Nadiminty, N., Lou, W., & Gao, A. C. (2014). Interleukin-6 induces neuroendocrine differentiation (NED) through suppression of RE-1 silencing transcription factor (REST). *Prostate*, 74(11), 1086-1094. doi:10.1002/pros.22819
- Zou, M., Toivanen, R., Mitrofanova, A., Floch, N., Hayati, S., Sun, Y., . . . Abate-Shen, C. (2017). Transdifferentiation as a Mechanism of Treatment Resistance in a Mouse Model of Castration-Resistant Prostate Cancer. *Cancer Discov*, 7(7), 736-749. doi:10.1158/2159-8290.CD-16-1174

Appendix I: Publications

1) Role of Specialized Composition of SWI/SNF Complexes in Prostate Cancer Lineage Plasticity.

Joanna Cyrta*, Anke Augspach*, Maria Rosaria de Filippo, Davide Prandi, Phillip Thienger, Matteo Benelli, Victoria Cooley, Rohan Bareja, David Wilkes, Sung-Suk Chae, Paola Cavaliere, Noah Dephoure, Anne-Christine Uldry, Sophie Braga Lagache, Sandra Cohen, Muriel Jaquet, Laura P. Brandt, Mohammed Alshalalfa, Andrea Sboner, Felix Feng, Shangqian Wang, Himisha Beltran, Tamara Lotan, Martin Spahn, Marianna Kruithof-de Julio, Yu Chen, Karla V. Ballman, Francesca Demichelis, Salvatore Piscuoglio and Mark A. Rubin.

* first co-authorship.

Nature Communications. 2020 Nov 3;11(1):5549.

2) Genomic correlates of clinical outcome in advanced prostate cancer.

Wassim Abida*, Joanna Cyrta*, Glenn Heller, Davide Prandi, Joshua Armenia, Ilsa Coleman, Marcin Cieslik, Matteo Benelli, Dan Robinson, Eliezer M. Van Allen, Andrea Sboner, Tarcisio Fedrizzi, Juan Miguel Mosquera, Brian D. Robinson, Navonil De Sarkar, Lakshmi P. Kunju, Scott Tomlins, Yi Mi Wu, Daniel Nava Rodrigues, Massimo Loda, Anuradha Gopalan, Victor E. Reuter, Colin C. Pritchard, Joaquin Mateo, Diletta Bianchini, Susana Miranda, Suzanne Carreira, Pasquale Rescigno, Julie Filipenko, Jacob Vinson, Robert B. Montgomery, Himisha Beltran, Elisabeth I. Heath, Howard I. Scher, Philip W. Kantoff, Mary-Ellen Taplin, Nikolaus Schultz, Johann S. deBono, Francesca Demichelis, Peter S. Nelson, Mark A. Rubin, Arul M. Chinnaiyan, and Charles L. Sawyers

* first co-authorship.

PNAS. June 4, 2019. 116 (23) 11428-11436

3) Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer.

Himisha Beltran, Davide Prandi, Juan Miguel Mosquera, Matteo Benelli, Loredana Puca, Joanna Cyrta, Clarisse Marotz, Eugenia Giannopoulou, Balabhadrapatruni V.S.K. Chakravarthi, Sooryanarayana Varambally, Scott A. Tomlins, David M. Nanus, Scott T. Tagawa, Eliezer M. Van Allen, Olivier Elemento, Andrea Sboner, Levi A. Garraway, Mark A. Rubin, and Francesca Demichelis.

Nature Medicine. 2016 Mar; 22(3): 298–305.

4) Small Cell Carcinoma of the Ovary, Hypercalcemic Type (SCCOHT) beyond *SMARCA4* Mutations: A Comprehensive Genomic Analysis.

Aur lie Auguste, F lix Blanc-Durand, Marc Deloger, Audrey Le Formal, Rohan Bareja, David C. Wilkes, Catherine Richon, B atrice Brunn, Olivier Caron, Mojgan Devouassoux-Shisheboran, S bastien Gouy, Philippe Morice, Enrica Bentivegna, Andrea Sboner, Olivier Elemento, Mark A. Rubin, Patricia Pautier, Catherine Genestie, Joanna Cyrta and Alexandra Leary.

Cells 2020. 9(6), 1496.

Appendix II: Other paper contributions during this PhD

1) Aberrant activation of a gastrointestinal transcriptional circuit in prostate cancer mediates castration resistance.

Shipra Shukla, Joanna Cyrta, Devan A. Murphy, Edward G. Walczak, Leili Ran, Praveen Agrawal, Yuanyuan Xie, Yuedan Chen, Shangqian Wang, Yu Zhan, Dan Li, Wai Pung E. Wong, Andrea Sboner, Himisha Beltran, Juan Miguel Mosquera, Jessica Sher, Zhen Cao, John Wongvipat, Richard P. Koche, Anuradha Gopalan, Deyou Zheng, Mark A. Rubin, Howard I. Scher, Ping Chi, and Yu Chen.

Cancer Cell. 2017 Dec 11; 32(6): 792–806.e7.

2) NSD2 is a conserved driver of metastatic prostate cancer progression.

Alvaro Aytes, Arianna Giacobbe, Antonina Mitrofanova, Katia Ruggero, Joanna Cyrta, Juan Arriaga, Luis Palomero, Sonia Farran-Matas, Mark A. Rubin, Michael M. Shen, Andrea Califano, and Cory Abate-Shen.

Nature Communications. 2018; 9: 5201.

3) Patient derived organoids to model rare prostate cancer phenotypes.

Loredana Puca, Rohan Bareja, Davide Prandi, Reid Shaw, Matteo Benelli, Wouter R. Karthaus, Judy Hess, Michael Sigouros, Adam Donoghue, Myriam Kossai, Dong Gao, Joanna Cyrta, Verena Sailer, Aram Vosoughi, Chantal Pauli, Yelena Churakova, Cynthia Cheung, Lesa Dayal Deonarine, Terra J. McNary, Rachele Rosati, Scott T. Tagawa, David M. Nanus, Juan Miguel Mosquera, Charles L. Sawyers, Yu Chen, Giorgio Inghirami, Rema A. Rao, Carla Grandori, Olivier Elemento, Andrea Sboner, Francesca Demichelis, Mark A. Rubin, and Himisha Beltran.

Nature Communications. 2018; 9: 2404.

4) Bone biopsy protocol for advanced prostate cancer in the era of precision medicine.

Verena Sailer, Marc H Schiffman, Myriam Kossai, Joanna Cyrta, Shaham Beg, Brian Sullivan, Bradley B Pua, Kyungmouk Steve Lee, Adam D Talenfeld, David M Nanus, Scott T Tagawa, Brian D Robinson, Rema A Rao, Chantal Pauli, Rohan Bareja, Luis S Beltran, Alexandros Sigaras, Kenneth Wa Eng, Olivier Elemento, Andrea Sboner, Mark A Rubin, Himisha Beltran, Juan Miguel Mosquera.

Cancer. 2018 Mar 1;124(5):1008-1015.

Role of specialized composition of SWI/SNF complexes in prostate cancer lineage plasticity

Joanna Cyrta et al.[#]

Advanced prostate cancer initially responds to hormonal treatment, but ultimately becomes resistant and requires more potent therapies. One mechanism of resistance observed in around 10–20% of these patients is lineage plasticity, which manifests in a partial or complete small cell or neuroendocrine prostate cancer (NEPC) phenotype. Here, we investigate the role of the mammalian SWI/SNF (mSWI/SNF) chromatin remodeling complex in NEPC. Using large patient datasets, patient-derived organoids and cancer cell lines, we identify mSWI/SNF subunits that are deregulated in NEPC and demonstrate that SMARCA4 (BRG1) overexpression is associated with aggressive disease. We also show that SWI/SNF complexes interact with different lineage-specific factors in NEPC compared to prostate adenocarcinoma. These data point to a role for mSWI/SNF complexes in therapy-related lineage plasticity, which may also be relevant for other solid tumors.

[#]A list of authors and their affiliations appears at the end of the paper.

Prostate cancer (PCa) is the second most commonly diagnosed cancer and the fifth cause of cancer-related death in men worldwide^{1,2}. Although most men are effectively treated by local therapies (surgery and/or radiotherapy), some develop metastatic recurrence or present with metastases at initial diagnosis. The mainstay of treatment for metastatic PCa is androgen deprivation therapy (ADT), but resistance ultimately develops with progression to castration-resistant prostate cancer (CRPC), which typically harbors a “luminal” (adenocarcinoma) phenotype (CRPC-Adeno) with continued dependence on androgen receptor (AR) signaling^{3–5}. Improved, more potent androgen receptor signaling inhibitors (ARSi) have been developed to treat patients that are not responsive to these therapeutics, yet acquired resistance to these drugs ultimately develops as well. In CRPC, indifference to AR signaling may manifest with a distinct histomorphology and expression of neural-like markers, leading to neuroendocrine or small cell prostate cancer (CRPC-NE)^{5–7}. Approximately 10–20% of CRPC cases treated with ARSi display a neuroendocrine phenotype^{5,8,9}. CRPC-NE no longer responds to ARSi and carries a dismal prognosis, with a mean overall survival of 12 months and no specific standard of care treatment options available¹⁰. There is mounting evidence that CRPC-Adeno can transdifferentiate to an AR-indifferent state through a mechanism of lineage plasticity under specific genomic conditions, including but not limited to *TP53*, *RBI*, and *PTEN* loss^{4,11–13}. Epigenetic regulators, such as *EZH2*, are also critical in this process^{4,12,13}. Although the mammalian Switch Sucrose Non-Fermenting (mSWI/SNF) complex is another major chromatin regulator well known for its role in physiological processes and frequently altered in cancer^{14–16}, its putative implication in NEPC lineage plasticity is unknown.

Mammalian SWI/SNF complexes, also known as Brg/Brahma-associated factor (BAF) complexes, are a heterogeneous family of ATP-dependent chromatin remodeling complexes composed of about 11–15 protein subunits and generally considered as positive mediators of chromatin accessibility¹⁶. These complexes are evolutionarily conserved in eukaryotes and required for normal embryonic development^{16,17}. Specialized complex assemblies with distinct functions have been identified at different stages of embryogenesis and during tissue maturation^{18–22}. Over 20% of human malignancies carry a genomic alteration involving at least one of the SWI/SNF subunit genes^{14–16}, including malignant rhabdoid tumors²³, synovial sarcoma²⁴, small cell carcinoma of the ovary hypercalcemic type, ovarian clear cell carcinoma, endometrioid carcinoma, bladder cancer, renal cell carcinoma, and lung adenocarcinoma, among others^{14,23,25–27}.

To date, SWI/SNF alterations have not been studied in the context of advanced PCa. In this study, we show that SWI/SNF composition is altered in the setting of CRPC-NE and that in contrast to many of the above-cited tumor types, SWI/SNF can have tumor-promoting functions in PCa. We also provide evidence that SWI/SNF interacts with different lineage-specific partners throughout PCa transdifferentiation. Collectively, these findings suggest that specialized SWI/SNF complexes are associated with PCa disease progression and may play a role in therapy resistance.

Results

SWI/SNF subunit expression is altered in CRPC-NE. To define somatic mutation frequencies of genes encoding SWI/SNF subunits across the entire spectrum of PCa, we conducted a comprehensive analysis of whole exome sequencing (WES) data from 600 PCa patients representing a wide range of the disease spectrum, including 56 CRPC-NE cases (Fig. 1a, Supplementary Data 1, Supplementary Data 2, Supplementary Data 3). No

recurrent SWI/SNF somatic mutations were observed and there was a low overall rate of point mutations and insertions/deletions in those genes (59 samples, 9.8% of all cases) (Fig. 1b). We observed an increased percentage of loss-of-heterozygosity (LOH) by hemizygous deletion or copy number neutral LOH (CNL), in 27 out of 28 genes (significant for 15 genes, proportion test, $\alpha = 0.05$), when comparing localized hormone treatment-naïve PCa vs. CRPC-Adeno (Supplementary Fig. 1, Supplementary Data 1). A similar result was obtained when comparing localized hormone treatment-naïve PCa and CRPC-NE cases (26 out of 28 genes with higher LOH frequency in CRPC-NE). Conversely, there were fewer differences when comparing CRPC-Adeno and CRPC-NE. A significant increase in the fraction of LOH in CRPC-NE as compared to CRPC-Adeno (proportion test, $\alpha = 0.05$) was only noted for three genes: *BRD7* (51% vs. 30%, respectively, $p = 0.005$), *SMARCD1* (11% vs. 3%, $p = 0.04$), and *PBRM1* (18% vs. 8%, $p = 0.049$) (Fig. 1b). However, this was not accompanied by a decrease in *SMARCD1* or *PBRM1* expression in CRPC-NE (Supplementary Fig. 2). Expression levels of *BRD7* were significantly lower in CRPC-NE compared to CRPC-Adeno, but not in CRPC-NE compared to localized PCa. This is in line with a previous study in which *BRD7* loss was identified as part of a larger heterozygous deletion event enriched in CRPC-NE and centered around the *CYLD* gene⁴. Collectively, these observations suggest that the increased fractions of LOH observed in CRPC-NE for *BRD7*, *PBRM1*, and *SMARCD1* are unlikely to carry functional significance.

Given the modest differential abundance of genomic lesions, we next queried the expression levels of SWI/SNF subunits by examining RNA-seq data of 572 unique PCa patients, including 20 CRPC-NE cases^{4,5} (Supplementary Data 4). The *SMARCA4* ATPase subunit was significantly upregulated, with accompanying downregulation of its mutually exclusive paralogue *SMARCA2*^{16,28} in CRPC-NE ($n = 20$) compared to CRPC-Adeno ($n = 120$) with a mean difference of 0.55 ($p = 0.015$) (averaged $\log_2(\text{FPKM} + 1)$) for *SMARCA4* and mean difference of -0.60 ($p = 0.02$) for *SMARCA2*, respectively (Fig. 1c). A concordant result was observed when comparing *SMARCA4*/*SMARCA2* expression ratios per patient in CRPC-Adeno (median ratio = 1.07) and in CRPC-NE (median ratio = 3.06, $p = 0.007$) (Supplementary Fig. 3). To validate that these transcriptomic findings translated into differences in protein expression, we performed immunohistochemistry (IHC) on patient samples and confirmed higher *SMARCA4* (BRG1) and lower *SMARCA2* (BRM) expression with increasing PCa disease progression, with highest *SMARCA4* expression observed in CRPC-NE (Fig. 1d and Supplementary Fig. 4).

Importantly, we also identified strong upregulation of neuron-specific SWI/SNF subunit genes in CRPC-NE: *ACTL6B* (BAF53B), *DPF1* (BAF45B), and *SS18L1* (CREST) (mean $\log_2[\text{FPKM} + 1]$ values: 2.79, 1.19, and 3.58, respectively) compared to CRPC-Adeno (mean 0.24, $p = 4.86\text{e}-06$; mean 0.35, $p = 0.0016$; and mean 2.76, $p = 6.85\text{e}-05$, respectively) (Fig. 1c). These subunits are expressed in post-mitotic neurons, serving instructive functions in neuronal differentiation²². By IHC, BAF53B, and BAF45B were highly expressed in CRPC-NE, but absent from benign prostate, localized PCa or CRPC-Adeno samples (Fig. 1d), demonstrating high specificity for the neuroendocrine phenotype.

We also noted intra-tumor heterogeneity in the expression of SWI/SNF subunits, as illustrated by IHC in patient specimens with a mixed phenotype (combining areas with adenocarcinoma and neuroendocrine differentiation) (Supplementary Figs. 4 and 5) and in 3D CRPC-NE organoid cultures (Supplementary Fig. 6). In the latter, we identified distinct cell clusters with high expression of the neural stem cell factor *SOX2*, low expression

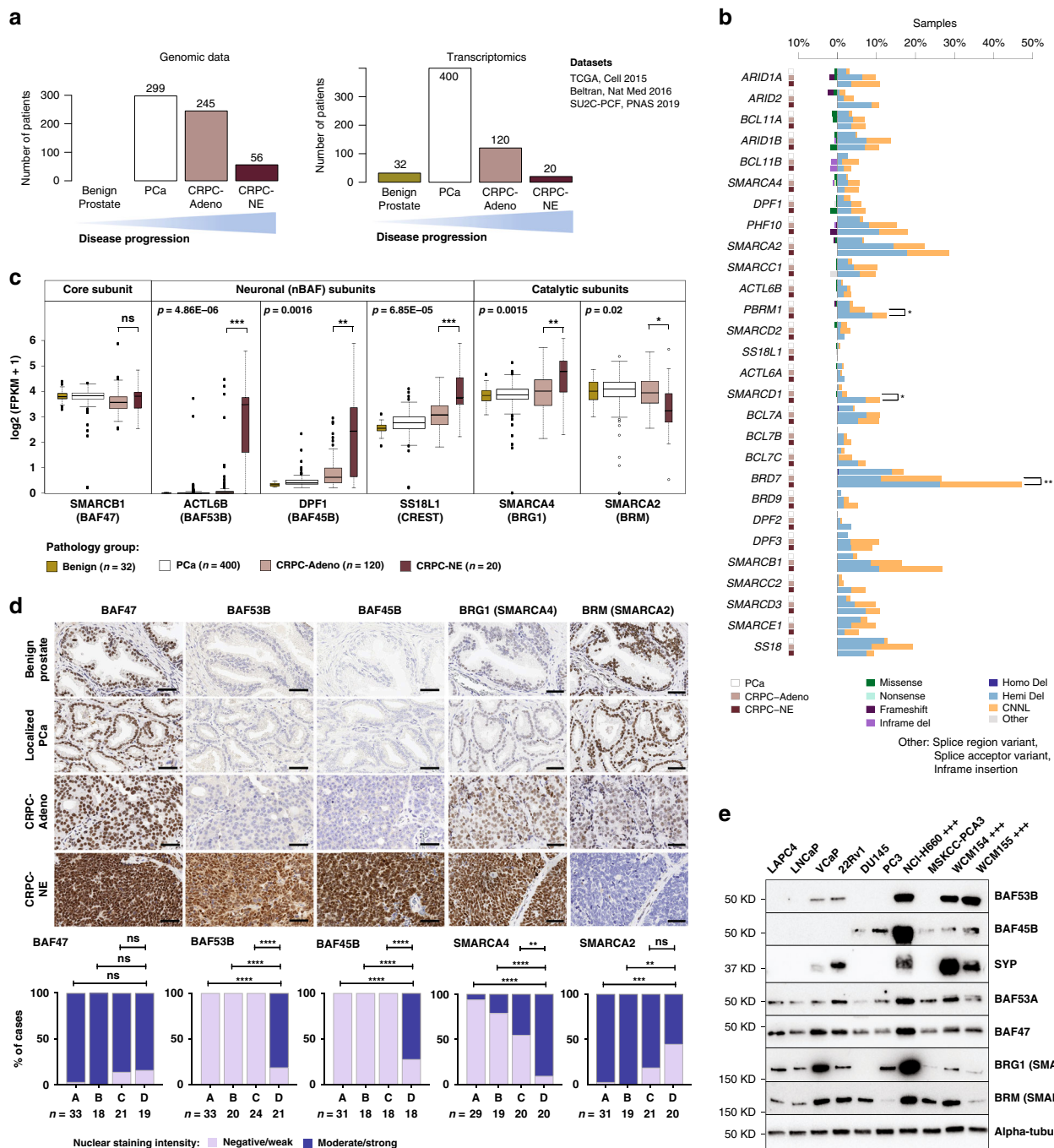


Fig. 1 Identification of SWI/SNF subunits deregulated in CRPC-NE. **a** Summary of the number of patients analyzed by whole exome sequencing (WES) and RNA-seq for each disease state. **b** WES results for SWI/SNF genes in 600 samples from unique PCa patients. For each gene, three consecutive bars represent alteration frequency in localized hormone treatment-naïve PCa, CRPC-Adeno and CRPC-NE, respectively. **c** RNA-seq analysis of gene expression levels in 572 unique patient samples from four studies, showing selected genes (*ACTL6B*: $p = 4.86E-06$, *DPF1*: $p = 0.0016$, *SS18L1*: $p = 6.85E-05$, *SMARCA4*: $p = 0.0015$, *SMARCA2*: $p = 0.02$) significantly deregulated in CRPC-NE. The core subunit *SMARCB1* is shown for comparison. The box plots represent the median values and the lower and upper interquartile range (IQR); the upper whisker = $\min(\max(x), Q3 + 1.5 \times IQR)$ and lower whisker = $\max(\min(x), Q1 - 1.5 \times IQR)$, and the outliers are plotted as individual points. **d** Representative immunostainings against BAF47 (*SMARCB1*), BAF53B (*ACTL6B*), BAF45B (*DPF1*), BRG1 (*SMARCA4*) and BRM (*SMARCA2*), and statistical analysis of staining intensity in patient samples. A-benign prostate glands, B-hormone treatment-naïve localized PCa, C-CRPC-Adeno, D-CRPC-NE. ** $p < 0.01$ ($p = 0.0057$ for BRG1, $p = 0.0012$ for BRM), *** $p < 0.001$ ($p = 0.0004$ for BRM) and **** $p < 0.0001$, ns indicates not significant (two-sided Fisher's exact test). Scale bars, 50 μm . **e** Immunoblot showing expression levels of selected SWI/SNF subunits in PCa cell lines (+++ designates CRPC-NE cell lines). Benign: benign prostatic tissue, PCa: localized hormone treatment-naïve prostate cancer, CRPC-Adeno: Castration resistant prostate cancer, adenocarcinoma subtype, CRPC-NE: Castration resistant prostate cancer, neuroendocrine subtype. Source data are provided in the Source Data file.

of the terminal neuronal marker synaptophysin, and higher expression of SMARCA4 (BRG1) and SMARCC1 (BAF155) than in the rest of the cell population (Supplementary Fig. 6). Overall, these observations suggest a relationship between expression of specific SWI/SNF subunits and different phenotype states, which can be seen even in a clonal tumor population.

BAF53B and BAF45B protein expression was confirmed in CRPC-NE cell lines and organoids (NCL-H660, WCM154, and WCM155²⁹) (Fig. 1e). BAF53B was also detected, albeit at lower levels, in two synaptophysin-positive PCa cell lines VCaP and 22Rv1, which bear some degree of transcriptomic similarity to neuroendocrine PCa cell lines⁹. BAF45B, on the other hand, was detected in some CRPC-Adeno cell lines and organoids (DU145, PC3, and MSKCC-PCA3). Unlike what we observed in patient samples, we did not observe high SMARCA4 (BRG1) and low SMARCA2 (BRM) expression in CRPC-NE cell lines, which could in part be due to marked differences in cell growth rates among different cell lines (Fig. 1e).

Although in neurons, BAF53B has been characterized as a mutually exclusive paralog to BAF53A, our data revealed that in CRPC-NE, BAF53A expression is maintained (Fig. 1e, Supplementary Fig. 7). BAF53B expression in neurons is known to be mediated by the downregulation of the RE1-Silencing Transcription factor (REST), a negative regulator of neuron-specific genes²⁰. In prostate adenocarcinoma cells, we observed that short-term REST knock-down led to an increase of BAF53B (*ACTL6B*) mRNA and protein levels, but the effect was modest, while other neuronal genes known to be negatively controlled by REST (e.g., synaptophysin) were highly upregulated (Supplementary Fig. 8).

To understand whether high SMARCA4 expression in CRPC-NE was related to other characteristics of CRPC-NE, such as acquisition of pluripotent stem cell-like features, and not only to the expression of terminal neural markers, we performed single-cell RNA-seq on two CRPC-NE organoids in 3D culture (MSKCC PCa1 and 16) and confirmed that SMARCA4 expression was significantly higher in cells with high expression of the pluripotent stem cell marker SOX2 (Supplementary Fig. 9), consistent with our IHC findings (Supplementary Fig. 6). Bulk RNA-seq data from 18 PCa organoids (CRPC-Adeno and CRPC-NE) (Supplementary Fig. 10) revealed that SMARCA4 expression was positively correlated with the expression of synaptophysin (a terminal neuronal marker), but also showed a tendency towards positive correlation with SOX2; conversely, there was a trend towards an inverse correlation between SMARCA2 and SOX2 (Pearson correlation analysis). Of note, some organoids (including MSKCC PCa1 and PCa16) classified as CRPC-NE based on their transcriptomic NEPC score⁴ showed high expression of SOX2, but low expression of terminal neural markers, such as synaptophysin (*SYN*) (Supplementary Fig. 10). These results suggest that high SMARCA4 expression may be related to pluripotent stem cell-like features and/or to proliferation at least in some CRPC-NE, rather than just to the expression of terminal neuronal markers.

Taken together, the above observations suggest that specialized SWI/SNF composition varies with PCa lineage plasticity to small cell or neuroendocrine states.

High SMARCA4 (BRG1) expression is associated with aggressive PCa. We posited that high SMARCA4 expression is associated with a more aggressive clinical course. To address this, we interrogated protein expression of SMARCA4 (BRG1) by IHC in a cohort of 203 men operated for localized hormone-treatment naïve PCa (demographics previously described in Spahn et al.³⁰). High SMARCA4 protein expression in primary PCa was

associated with a significantly shorter overall survival (HR = 2.17 [95% CI: 1.07–4.42], $p = 0.028$) (Fig. 2a). This relationship remained significant after adjustment for single covariates that have known association with PCa outcome (Supplementary Table 1). Patients with high tumor SMARCA2 (BRM) protein expression showed a trend towards a better overall survival, although this relationship did not reach statistical significance. Taken together, the above findings suggest that high SMARCA4 expression is associated with more aggressive cases of PCa.

We next sought to determine the effects of SMARCA4 and SMARCA2 depletion in PCa cell lines. We performed siRNA-mediated knock-down of SMARCA4 and SMARCA2 in an androgen-sensitive (LNCaP) cell line and in a CRPC-Adeno cell line (22Rv1) and compared global transcriptional alterations using RNA-seq. As expected, given its posited dominant role, SMARCA4 depletion demonstrated a stronger effect on the transcriptome of both cell lines, while SMARCA2 depletion led to only modest transcriptional alterations (Fig. 2b, Supplementary Figs. 11 and 12). Among the genes most significantly deregulated upon SMARCA4 knock-down were several of known significance in PCa progression, including: upregulation of cell cycle regulators *CDKN1A* (p21) and *BTG2* (in both LNCaP and 22Rv1 cell lines), downregulation of E2F targets (in both cell lines), downregulation of *EZH2*, and downregulation of the oncogenic long non-coding RNA *PCAT-1* (both significant in LNCaP only)^{4,31,32} (Fig. 2c–e, Supplementary Fig. 12, Supplementary Data 5, Supplementary Data 6). We also observed a significant enrichment in gene sets related to *EZH2* knock-down, suggesting that knock-down of SMARCA4 and knock-down of *EZH2* can have partly overlapping effects in PCa cells (Supplementary Fig. 13). Expression of *REST* was not altered by SMARCA4 knock-down (Supplementary Fig. 14).

The observed changes in cell cycle-related pathways led us to explore the requirement for SMARCA4 and SMARCA2 for PCa cell growth. Depletion of SMARCA4, but not of SMARCA2, significantly reduced proliferation of the adenocarcinoma cell line LNCaP and the LNCaP-derived androgen-independent CRPC-Adeno cell line C4-2 (Fig. 2f), in line with previous findings^{33,34}. Knock-down of SMARCA4, but not of SMARCA2, in PCa cells resulted in a decrease of other SWI/SNF subunits, including SMARCC1 (BAF155) and *ACTL6A* (BAF53A), at the protein level, but not at the transcript level (Supplementary Fig. 15). Accordingly, both LNCaP and C4-2 cells proved to be highly sensitive to depletion of BAF155 (SMARCC1) (Supplementary Fig. 16). Recent work has shown that sensitivity of PCa cells to SMARCA4 knock-down may be dependent on *PTEN* loss, via a mechanism of synthetic lethality³³. To expand upon these findings, we performed knock-down of BAF155 (SMARCC1) in two *PTEN* wild-type cell lines, 22Rv1 (CRPC-Adeno) and WCM154 (CRPC-NE), and observed a significant decrease in cell growth (Supplementary Fig. 17). This suggests that *PTEN*-competent PCa cells can still be sensitive to SWI/SNF disruption, even though they may be differentially responsive to depletion of different subunits.

Given that loss of *TP53* and/or *RBI* has been suggested to confer a poised pluripotent state required for neuroendocrine transdifferentiation^{11,12}, we also tested the effect of SMARCA4 knock-down in LNCaP cells having undergone CRISPR-Cas9 mediated knock-out of *TP53*, *RBI*, or both genes. The effect of SMARCA4 knock-down on cell proliferation was not entirely abrogated by the absence of functional p53 and/or Rb (Supplementary Fig. 18).

To strengthen the above observations of a putative tumor-promoting function of SMARCA4 (BRG1) in PCa, we also sought to study the effects of SMARCA4 overexpression in PCa cells.

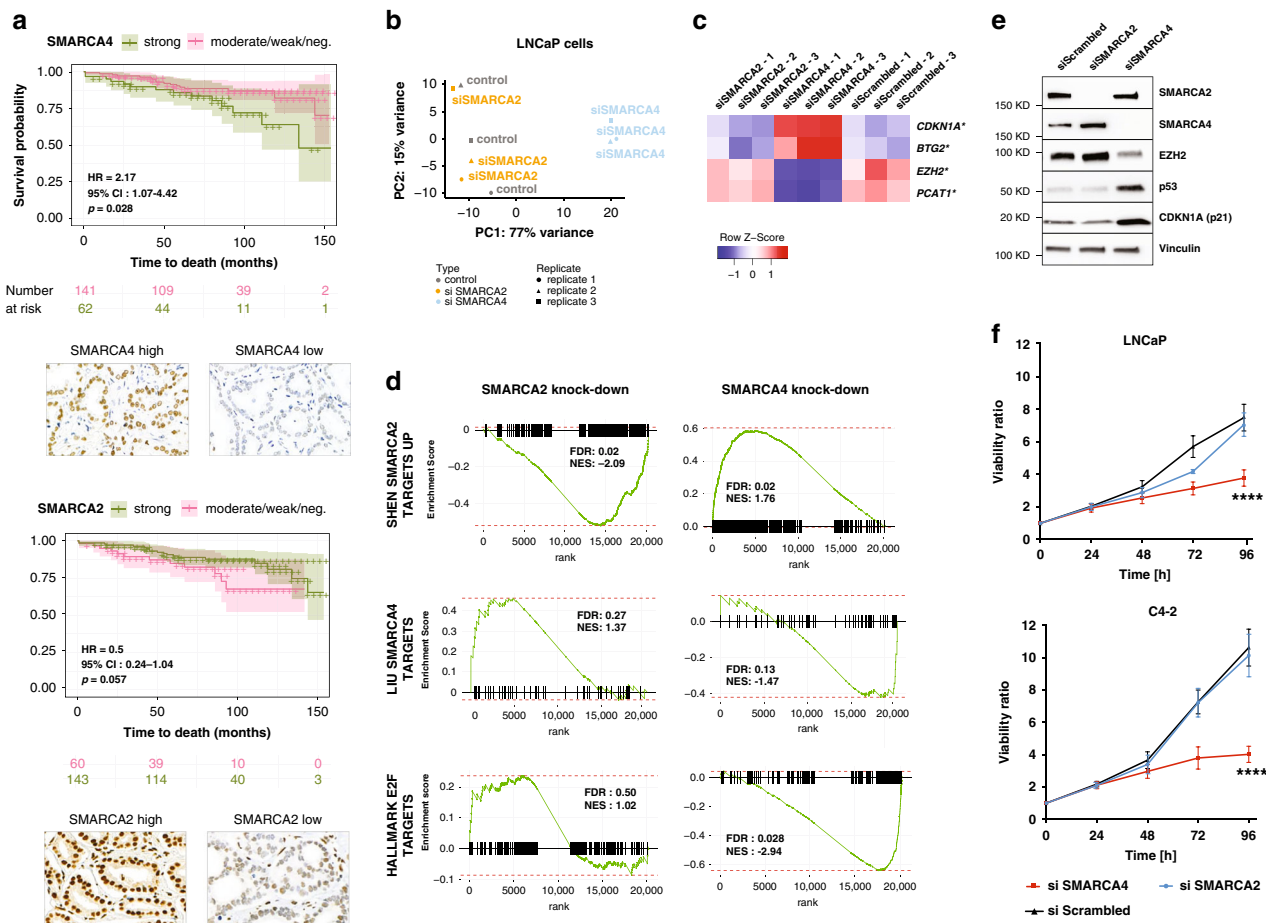


Fig. 2 SWI/SNF *SMARCA4* and *SMARCA2* expression in prostate cancer. **a** Kaplan–Meier curves showing the association between overall survival and *SMARCA4* (BRG1) IHC expression ($p = 0.028$, Log-rank test) or *SMARCA2* (BRM) IHC expression (not significant), in 203 patients with localized PCa. **b** Principal component analysis (PCA) of RNA-seq data for prostate adenocarcinoma (LNCaP) cells 72 h after *SMARCA4* or *SMARCA2* knock-down. **c** Expression levels (RNA-seq) of selected genes upon *SMARCA4* and *SMARCA2* knock-down in LNCaP cells; *FDR < 0.05. **d** Gene Set Enrichment Analysis based on RNA-seq gene expression analysis in LNCaP cells with *SMARCA4* or *SMARCA2* knock-down. **e** Immunoblot showing selected deregulated proteins upon *SMARCA4* and *SMARCA2* knock-down in LNCaP cells. **f** Effect of *SMARCA4* or *SMARCA2* knock-down on cell proliferation of prostatic adenocarcinoma (LNCaP) and CRPC-Adeno (C4-2) cells. $N = 3$ independent experiments. Data are presented as mean values \pm SEM and analyzed using two-way Anova (**** $p < 0.0001$). Statistical significance was evaluated at 0.05 alpha level with GraphPadPrism, version 8.2.1, Mac. Source data are provided in the Source Data file.

22Rv1 cells were stably transduced with lentiviral vectors designed to overexpress either *SMARCA4* or *SMARCA2* or with a matched empty control vector, and sorted based on the expression of the fluorescent reporter. Despite strong expression of the reporters, we did not observe an increase in *SMARCA4* or *SMARCA2* at the protein level (Supplementary Fig. 19). However, after an additional 24 h treatment with the proteasome inhibitor MG-132, *SMARCA4*, and *SMARCA2* overexpression was readily detected. These findings hint towards a tight and context-dependent regulation of catalytic SWI/SNF subunits, as forced isolated overexpression of a single subunit seems to provoke rapid degradation of the excess protein. Thus, it is possible that *SMARCA4* overexpression may be necessary, but not sufficient, to promote an aggressive phenotype in prostate cancer cells.

To understand whether BAF53B and BAF45B—two other subunits overexpressed in CRPC-NE—potentially regulated similar gene expression programs as *SMARCA4*, we performed shRNA-mediated knock-down of these subunits in the CRPC-NE organoid line WCM155. Neither BAF53B nor BAF45B knock-down had an effect on CRPC-NE cell proliferation (Supplementary Fig. 20) Therefore, it appears that BAF53B and BAF45B expression may be specific for the CRPC-NE phenotype, but not a

critical mediator of CRPC-NE aggressiveness. Collectively, the above genomic, transcriptomic, and functional findings support a tumor-promoting role of *SMARCA4*-containing mSWI/SNF complexes in PCa.

Aggressive prostate cancer anti-correlates with *SMARCA4* knock-down signature. Based on the association of *SMARCA4* expression with poor clinical outcome, and observations that *SMARCA4* knock-down leads to a significant decrease in PCa cell growth in line with previous studies^{33,34}, we posited that a *SMARCA4* knock-down signature (composed of genes deregulated upon *SMARCA4* depletion) would be associated with more indolent PCa. To address this, we interrogated RNA-seq data of several large clinical cohorts using a *SMARCA4* knock-down signature derived from the LNCaP PCa cell line (see “Methods” section) and composed of the top 419 deregulated genes. A high *SMARCA4* knock-down signature score was, indeed, associated with more indolent disease. In contrast, a low *SMARCA4* knock-down signature score was associated with more aggressive PCa.

As expected, a low *SMARCA4* knock-down signature score was also strongly associated with a CRPC-NE phenotype. We

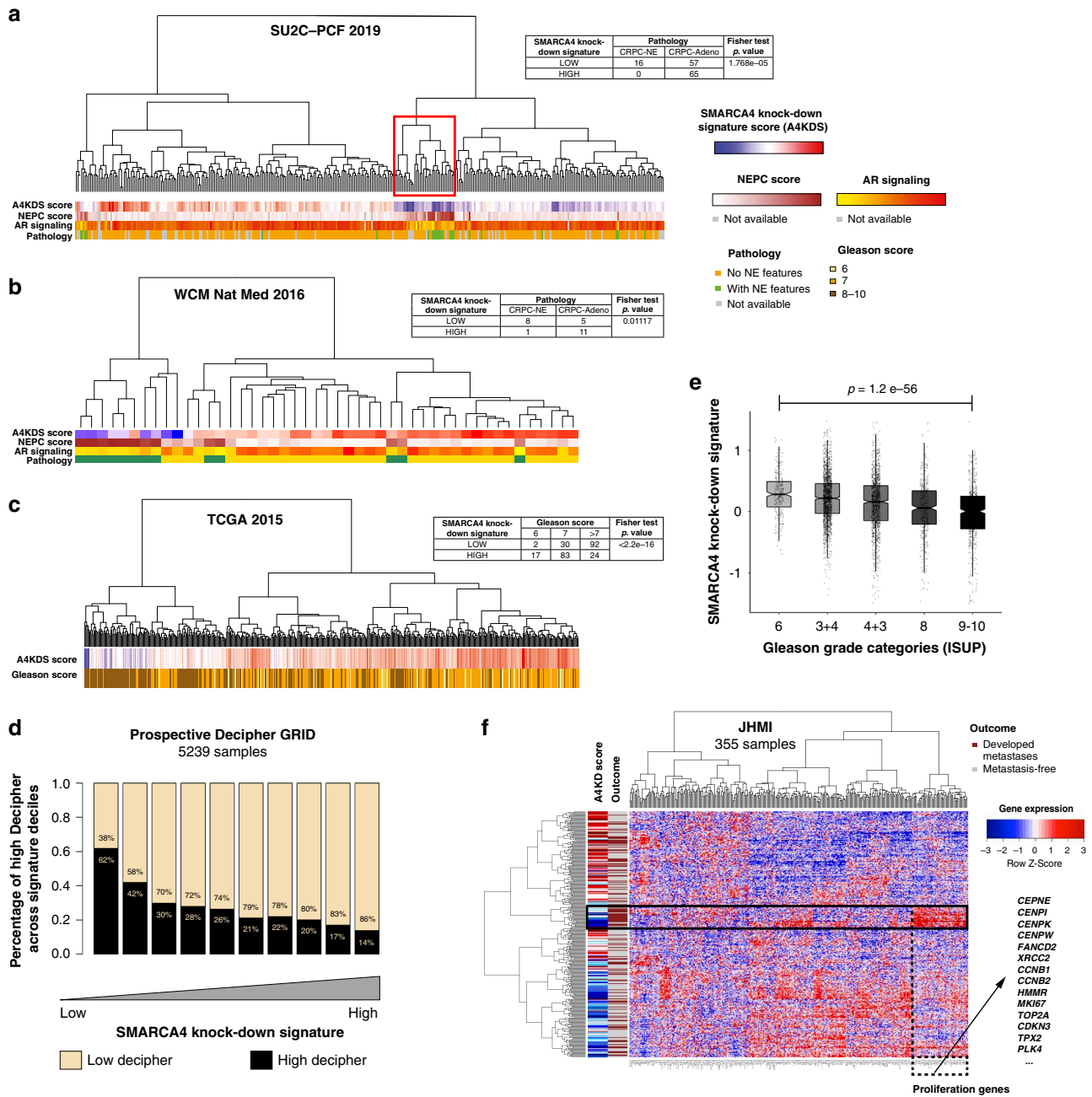


Fig. 3 Transcriptomic *SMARCA4* knock-down signature in PCa cohorts. **a** 332 cases of CRPC from the SU2C-PCF cohort (table: $n = 138$). **b** 47 cases of CRPC from the WCM cohort (table: $n = 25$). **c** 495 cases of localized PCa from the TCGA cohort (table: $n = 248$); p value = 1.21×10^{-18} (**a–c** represent two-paired tests). **d** Low *SMARCA4* knock-down signature scores are associated with high Decipher scores (surrogate for risk of metastasis) in 5239 primary PCa samples from the Prospective Decipher GRID (Mann-Whitney U test). **e** Low *SMARCA4* knock-down signature scores are associated with higher Gleason score in the same Decipher GRID cohort (Mann-Kendall trend test). The center of each boxplot represents the mean, lower bound represents the 25th percentile, the top bound represents the 75th percentile, the whiskers represent the 95% CI. p value = 1.2×10^{-56} . **f** Unsupervised clustering of patients from in the JHMI natural history PCa cohort (Johns Hopkins Medical Institute, $n = 355$) based on the downregulated genes from the *SMARCA4* knock-down signature, and compared to metastatic outcome (brown: metastatic recurrence, gray: metastasis-free). Overexpression of a subset of genes, many of which are related to proliferation, is seen in a cluster of patients who presented metastatic outcome (black box).

examined two CRPC cohorts consisting of 332 patients from the Stand Up To Cancer-Prostate Cancer Foundation (SU2C-PCF) trial treated with ARSi⁵ and 47 patients from the Weill Cornell Medicine (WCM) cohort⁴. In the SU2C-PCF cohort, when considering patients from the highest (top 25%) and lowest (bottom 25%) quartiles of *SMARCA4* knock-down signature scores ($n = 138$), low *SMARCA4* knock-down signature scores were significantly more often observed in CRPC-NE cases ($n = 16$

or 100%) than in CRPC-Adeno cases ($n = 57$ or 46.7%) ($p = 1.77 \times 10^{-5}$) (Fig. 3a). A similar result was obtained in the WCM cohort ($n = 25$): low *SMARCA4* knock-down signature scores were seen in 89% ($n = 8$) of CRPC-NE cases vs. 31% ($n = 5$) of CRPC-Adeno cases ($p = 0.011$) (Fig. 3b). Furthermore, low *SMARCA4* knock-down signature was associated with a higher NEPC4 and a lower AR signaling score³⁵ in both cohorts (Supplementary Table 2). One particularly informative cluster

was found to show low *SMARCA4* knock-down signature scores, high CRPC-NE scores, and low AR signaling scores (Fig. 3a, red box). Of note, *SMARCA4* mRNA levels were consistent with the predicted signature score in all analyzed cohorts (Supplementary Fig. 21).

We next queried if the *SMARCA4* knock-down signature was associated with higher tumor grade, referred to as Gleason score risk groups in localized PCa³⁶. We first explored 248 patients from The Cancer Genome Atlas (TCGA) PCa cohort with localized, hormone treatment-naïve PCa³⁷. Tumors in the highest Gleason score risk groups (IV and V) more often displayed low *SMARCA4* knock-down signature scores ($p < 2.2 \times 10^{-16}$) (Fig. 3c).

As high tumor grade is associated with risk of metastatic progression, we decided to validate these findings in other independent clinical cohorts annotated with clinical survival data. We calculated *SMARCA4* knock-down signature scores for 5239 prospectively collected radical prostatectomy samples from men with localized PCa and analyzed with the Decipher GRID transcriptomic platform³⁸. Samples with a low *SMARCA4* knock-down signature (lowest 10%) were significantly enriched (62%) with high Decipher score, which is a strong surrogate of metastasis prediction³⁸ (Fig. 3d), compared to 14% in samples with high *SMARCA4* knock-down signature (highest 10%). In this patient population and consistent with TCGA results, we observed an association between *SMARCA4* knock-down signature and Gleason score risk categories: signature scores in the Gleason 9–10 group (mean = -0.13) were significantly lower compared to the Gleason 6 group (mean = 0.29 , $p = 1.2 \times 10^{-56}$) (Fig. 3e). We next explored an independent retrospective cohort from Johns Hopkins Medical Institution (JHMI)³⁹. In the JHMI cohort, patients with low *SMARCA4* knock-down signature showed a trend towards higher metastasis frequency, the strongest surrogate for lethal disease progression (Supplementary Fig. 22). When clustering patients based on the downregulated genes (Fig. 3f) or on all genes (Supplementary Fig. 23) that make up the *SMARCA4* knock-down signature, overexpression of a subset of genes involved in cell proliferation was associated with a cluster of patients enriched with metastatic outcome (Fig. 3f, box). In summary, these results from large patient cohorts confirm that the lowest *SMARCA4* knock-down signatures are observed in the most aggressive PCa.

The SWI/SNF complex has distinct lineage-specific interaction partners in CRPC-NE and in prostate adenocarcinoma cells. To gain insight into the potential effectors of NEPC-specific epigenetic regulation, we next sought to identify interactors of mSWI/SNF in the context of CRPC-NE and prostate adenocarcinoma cell lines. To this end, we performed co-IP with an antibody directed against the core SWI/SNF subunit BAF155 (*SMARCC1*) at low stringency (see “Methods” section) followed by mass spectrometry (MS) in NCI-H660 (a CRPC-NE cell line) and in LNCaP-AR cells (LNCaP cells engineered to overexpress the androgen receptor⁴⁰). Proteins that immunoprecipitated with BAF155 in CRPC-NE cells, but not in adenocarcinoma cells, (Fig. 4a, b) included BAF53B (*ACTL6B*) and BAF45B (*DPPF1*) subunits, as anticipated from results described above, as well as several factors specific to neural differentiation, such as the transcription factor NKX2.1 (*TTF-1*), the microtubule-associated factor MAP2 and the growth factor VGF. Moreover, we found several members of the NuRD chromatin remodeling complex, such as MTA1 and CHD4, to immunoprecipitate with BAF155. This is in line with previous findings of a potential interaction of those two chromatin remodeling complexes (Fig. 4a, b)^{41,42}. A considerable amount of CRPC-NE specific SWI/SNF interactors were proteins involved in chromatin regulation or DNA repair

(Fig. 4a, b, Supplementary Data 7, Supplementary Data 8). Conversely, proteins that immunoprecipitated with BAF155 in adenocarcinoma cells, but not in CRPC-NE, included HOXB13, a homeobox transcription factor involved in AR signaling⁴³ (Fig. 4b). In line with these findings, genes encoding most of the above factors were differentially expressed between CRPC-NE and adenocarcinoma cell lines and organoids (Fig. 4c, Supplementary Fig. 24). Further, we confirmed unique interaction of factors NKX2.1, CHD4, MTA1, and VGF with BAF155 in NCI-H660 by immunoblotting, while these interactions were absent in LNCaP-AR cells (Supplementary Fig. 25a). Interaction of HOXB13 with BAF155 in LNCaP-AR cells was also confirmed by immunoblotting (Supplementary Fig. 25b). The co-IP experiment also showed an enrichment of proteins negatively associated with REST signaling in NCI-H660 cells, such as HMG20A, a chromatin-associated protein known to overcome the repressive effects of REST and induce activation of neuronal genes⁴⁴. Loss of expression or altered splicing of REST has been associated with neural-like lineage plasticity in PCa in multiple studies^{45–51}. An independent co-IP experiment using an antibody directed against *SMARCA4* followed by MS in NCI-H660 and in LNCaP cells found similar results for BAF53B, BAF45B, NKX2.1, and HOXB13 (Supplementary Fig. 26, Supplementary Data 9).

As a proof-of-principle, we compared genome occupancy of *SMARCC1*, HOXB13, the active chromatin histone mark H3K27ac and the inactive chromatin mark H3K27me3 in LNCaP cells, using published ChIP-seq datasets (Supplementary Fig. 27). *SMARCC1* and HOXB13 colocalized at active chromatin sites (11,824 sites), while there was almost no overlap between *SMARCC1* and HOXB13 at inactive chromatin sites, thus suggesting a functional nature of this interaction. Collectively, the above observations suggest that the set of SWI/SNF interaction partners in CRPC-NE is quite distinct from the one in prostatic adenocarcinoma.

Discussion

Whereas neuroendocrine PCa is rarely present at diagnosis in hormone-treatment naïve PCa patients (de novo neuroendocrine PCa, <1% of cases)⁵², recent work supports the hypothesis that acquisition of a CRPC-NE phenotype in PCa is a more common mechanism of resistance to ARSi^{4,5,8,13,53}. Based on a recent review of 440 CRPC patients, CRPC-NE was seen in 11% of CRPC patients that underwent biopsy^{5,8,9}. There is increasing evidence that CRPC-NE can directly arise from CRPC-Adeno cells through lineage plasticity (Supplementary Fig. 28), which is supported by lineage tracing experiments in a genetically engineered mouse model of PCa with combined *Trp53* and *Pten* loss⁵⁴. Moreover, mouse models with *Trp53* and *Rb1* genomic loss show lineage plasticity, but epigenetic therapy can sensitize those tumors towards ARSi treatment¹¹. In patient cohorts, CRPC-NE are characterized by an overexpression of several epigenetic regulators (such as EZH2) and a specific DNA methylation profile^{4,13,29}. Overall, these data support the idea that PCa progression through lineage plasticity is regulated by epigenetic changes in a specific genomic context^{12,55}.

Given that mSWI/SNF complexes are major epigenetic regulators in physiological cell differentiation, we posited that they may play a role in CRPC-NE lineage plasticity. Specialized assemblies of the SWI/SNF complex with distinct functions are observed at different stages of embryonic development and tissue maturation^{18,19}. The most notable changes in SWI/SNF composition described to date occur during neuronal differentiation. Cells committed to the neural lineage initially express a neural progenitor form of the complex (termed npBAF), which incorporates among others the BAF53A, BAF45A/D, and

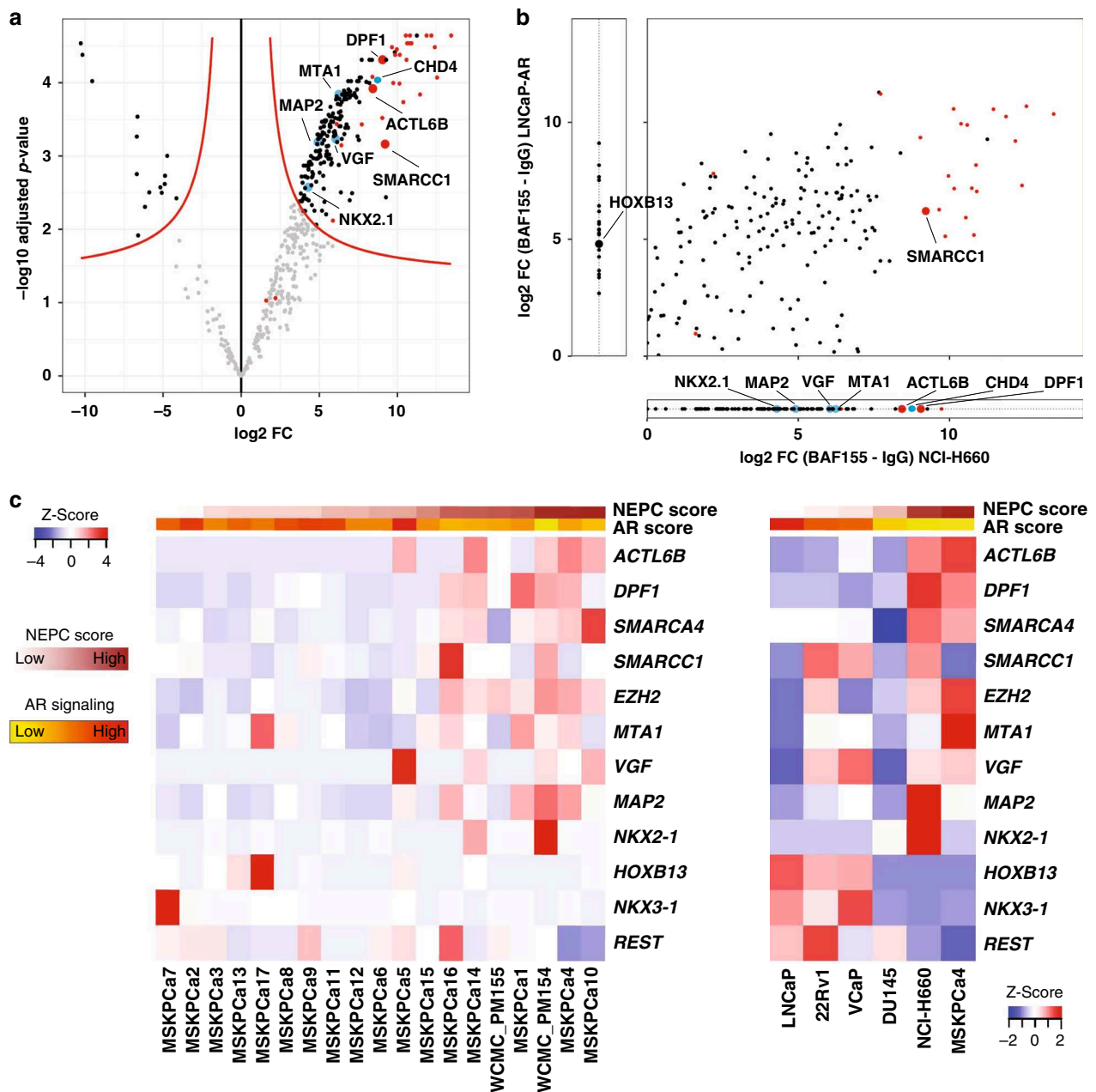


Fig. 4 SWI/SNF associates with different transcriptional regulators in CRPC-NE and in adenocarcinoma cells. **a** Volcano plot showing proteins most significantly represented (upper right) in the co-IP using an anti-BAF155 antibody, as compared to IgG isotype control in NCI-H660 (CRPC-NE) cells (pooled data from 3 co-IP replicates). The x-axis represents log₂ fold change (FC) values, the y-axis represents $-\log_{10}$ of adjusted p-values. Each dot represents a protein; red dots represent SWI/SNF members, blue dots indicate notable findings. **b** A qualitative representation comparing proteins associated with SWI/SNF in NCI-H660 (CRPC-NE) and in LNCaP-AR (adenocarcinoma) cells (averaged data from two co-IP experiments). Plotted are log₂ fold change values between BAF155 IP and IgG IP in NCI-H660 cells (x-axis) and in LNCaP-AR cells (y-axis), for proteins present in both cell lines with sufficient evidence in each cell line (i.e., if present in two replicates of at least one condition). Proteins plotted outside of the main field represent proteins that were detected exclusively in one of the cell lines. **c** Heatmap showing RNA-seq expression (FPKM) of prostate cancer 3D organoids (left) and 2D cell lines (right), ordered by increasing NEPC score.

SS18 subunits^{20–22}. However, upon differentiation to post-mitotic neurons, the complex undergoes a switch to the neural variant and incorporates the respective paralogs of these subunits (i.e., BAF53B, BAF45B/C, and SS18L1). This switch is mediated by repression of BAF53A by micro-RNAs in response to down-regulation of REST²⁰. In this study, we observed for the first time the presence of “neuronal” SWI/SNF subunits outside of the nervous system, characterized by the expression of BAF53B and BAF45B in CRPC-NE. Although expression of these subunits was

highly specific of CRPC-NE, it remains unclear whether they play a role in activating neural-like gene programs, or are simply expressed as a consequence of this process. Additional studies are warranted to assess the putative utility of BAF53B and BAF45B as CRPC-NE biomarkers or as predictors of patients at risk of developing CRPC-NE from CRPC-Adeno while on ARSi. Of note, expression of the BAF53A paralogue is retained in CRPC-NE, pointing to potential differences in the way SWI/SNF complexes assemble in post-mitotic neurons and in neuroendocrine

cancer cells, and to possible co-existence of different forms of the complex within the same tumor.

This study supports a pleiotropic role for the SWI/SNF chromatin remodeling complex in cancer, which may depend on the genomic and/or the epigenetic context—a paradigm which has been gaining support both in regards to SWI/SNF and to other epigenetic regulators^{56–58}. Although the complex has been described as a tumor suppressor in many cancer types^{14,23,25,59}, there is increasing evidence for possible tumor-promoting functions of SWI/SNF in other malignancies, including leukemia, breast, liver and pancreas cancer melanoma, glioblastoma, neuroblastoma and synovial sarcoma^{24,60–65}. In PCa, the role of SWI/SNF has long remained insufficiently characterized. Our study provides novel evidence that it can have tumor-promoting functions in PCa, including its most aggressive forms. Based on prior studies and on the current analysis, mutations in SWI/SNF genes are very rare in PCa^{4,5,34,37,66–68} (see Fig. 1b), in contrast to some other cancers types^{14,15}. From the functional perspective, inhibition of the SWI/SNF subunits BAF57 (*SMARCE1*) or BAF53A (*ACTL6A*) in PCa cells has been shown to abrogate androgen-dependent cell proliferation^{69,70}. Similarly, Sandoval et al. reported that SWI/SNF interacts with ERG in PCa cells harboring the *TMPRSS2:ERG* gene fusion and is required to activate specific gene programs to maintain cell growth⁷¹. Although on the contrary, Prensner et al. had suggested that SWI/SNF acts as a tumor suppressor in PCa, by demonstrating an antagonistic relationship between the pro-oncogenic long non-coding RNA *SChLAP1* and the SWI/SNF core subunit BAF47⁷², a subsequent study failed to confirm that *SChLAP1*-SWI/SNF interaction leads to depletion of SWI/SNF from the genome⁷³. Most recently, two studies demonstrated that *SMARCA4* was required for growth of prostatic adenocarcinoma cells^{33,34}, as also confirmed by our results (Fig. 2). Accordingly, localized PCa has been reported to show higher *SMARCA4* and lower *SMARCA2* expression than benign prostate tissue^{33,34,74,75}. We confirm these results and further report an overexpression of *SMARCA4* in CRPC and especially in CRPC-NE, in contrast to lower expression in early PCa. In addition, we show that a low *SMARCA4* knock-down gene signature score is associated with aggressive PCa, and with a CRPC-NE phenotype.

Recent work by Ding et al. specifically proposed a synthetic lethal association between PTEN and *SMARCA4* in PCa, identified through a CRISPR-Cas9 screen³³. They showed that in vitro, *SMARCA4* knock-down leads to decreased cell proliferation in PTEN-negative cell lines, and confirmed these findings in a mouse model. In our study, knock-down of the core SWI/SNF subunit BAF155 (*SMARCC1*) and BAF170 (*SMARCC2*) inhibited growth of both PTEN-deficient and PTEN-competent PCa adenocarcinoma cells (Supplementary Figs. 16 and 17), and the PTEN-competent CRPC-NE cell line WCM154 was sensitive to ablation of BAF155, but not of BAF170. This suggests that even if PTEN-competent cells are not sensitive to *SMARCA4* loss, they may still be vulnerable to SWI/SNF disruption through depletion of other critical subunits. Taken together, our and previously published findings indicate that PCa expands the spectrum of cancer types in which SWI/SNF can display tumor-promoting functions.

In addition, we observed that SWI/SNF composition in prostate cancer is not a hard-set feature; instead, specialized forms of SWI/SNF may assemble in cancer cells depending on their phenotype (Fig. 5). There is increasing evidence that de-repression of “terminal” neuronal genes in PCa cells is not sufficient to model other critical steps of neuroendocrine lineage plasticity in CRPC-NE⁷⁶. As such, the distinct phenotype of CRPC-NE is not limited to the expression of terminal neuronal markers, but involves other key characteristics, such as dedifferentiation, AR signaling

indifference, acquisition of stem cell-like features and/or high proliferation¹³. In line with this, we show that some patient-derived PCa organoids that are classified as CRPC-NE using a transcriptome-based NEPC score⁴ (Supplementary Fig. 10), do not all show high expression of terminal neural markers such as synaptophysin, but instead may highly express factors related to “stemness” (e.g., SOX2). Based on our observations, it is possible that specific forms of SWI/SNF are implicated in various above-mentioned cellular processes, rather than only in the expression of terminal neuronal markers. One possible hypothesis is that an equivalent of the embryonic stem cell form of the complex (esBAF), which is known to exclusively incorporate BRG1 (*SMARCA4*), BAF53A and BAF155 (*SMARCC1*) subunits and not their paralogs^{18,19}, could exist in cancers cells with pluripotent stem cell-like features, and possibly explain the over-expression and/or the functional requirement for these subunits. Similarly, neural-like forms of the complex, including BAF53B and/or BAF45B, could be more specific of cancer cells with a more terminal neural-like phenotype. Further studies are needed to determine whether variants of SWI/SNF can co-exist within the same cell or whether they define distinct tumor sub-populations, in line with what we have observed in 3D CRPC-NE organoid cultures (Supplementary Fig. 6).

One of the ways in which SWI/SNF might contribute to CRPC-NE transdifferentiation is by cooperating with other transcriptional regulators in a context-dependent manner. To this end, we showed that SWI/SNF interacts with different lineage-specific proteins in CRPC-NE than in adenocarcinoma cells (Fig. 4b, Supplementary Fig. 25). In particular, SWI/SNF interacts with the transcription factor NK2 homeobox 1 (NKX2.1/TTF-1) in CRPC-NE cells, but not in adenocarcinoma cells (Fig. 4c, Supplementary Fig. 25). TTF-1 is a master regulator critical for the development of lung and thyroid, but also of specific parts of the brain^{77–79} and is known to be expressed in neuroendocrine neoplasms, including CRPC-NE⁷⁶. We also observed SWI/SNF interaction with Metastasis-associated Protein 1 (MTA1), a member of the nucleosome-remodeling and deacetylation complex (NuRD), which is overexpressed in metastatic prostate cancer⁸⁰ (Fig. 4c). Conversely, we found HOXB13 to be specifically associated with SWI/SNF in adenocarcinoma cells, and not in CRPC-NE. HOXB13 is a homeobox transcription factor involved in prostate development and displays context-dependent roles in PCa: it can act as a collaborator or a negative regulator of AR signaling^{43,81}, it cooperates with the AR-V7 splice variant found in a subset of CRPC-Adeno⁸², and germline gain-of-function G84E *HOXB13* mutations are associated with increased prostate cancer risk⁸³. The fact that by ChIP-seq, SWI/SNF colocalizes with HOXB13 at active chromatin sites in prostatic adenocarcinoma cells, further supports the hypothesis that interaction between SWI/SNF and lineage-specific factors in PCa may be meaningful at the functional level.

In conclusion, this work confirms that SWI/SNF has tumor-promoting functions in PCa, including the lethal CRPC-NE. Our findings provide a rationale to further study selected SWI/SNF subunits as potential therapeutic targets in PCa.

Methods

Genomic analysis. Matched tumor and normal WES data of localized and advanced prostate cancer from The Cancer Genome Atlas⁸⁴, SU2C-PCF⁵ and from the Weill Cornell Medicine cohort⁴ were uniformly analyzed for somatic copy number aberrations (SCNA) with CNVkit⁸⁵, and for single nucleotide variations (SNVs) and indels with MuTect2⁸⁶. SNVs and Indels were annotated with variant effect predictor (VEP)⁸⁷ and only mutations with HIGH or MODERATE predicted impact on a transcript or protein (https://www.ensembl.org/info/genome/variation/prediction/predicted_data.html) were retained. All samples with tumor ploidy and purity estimated using CLONET⁸⁸ were retained in the analyses and processed for allele specific characterization. The integrated dataset includes 299

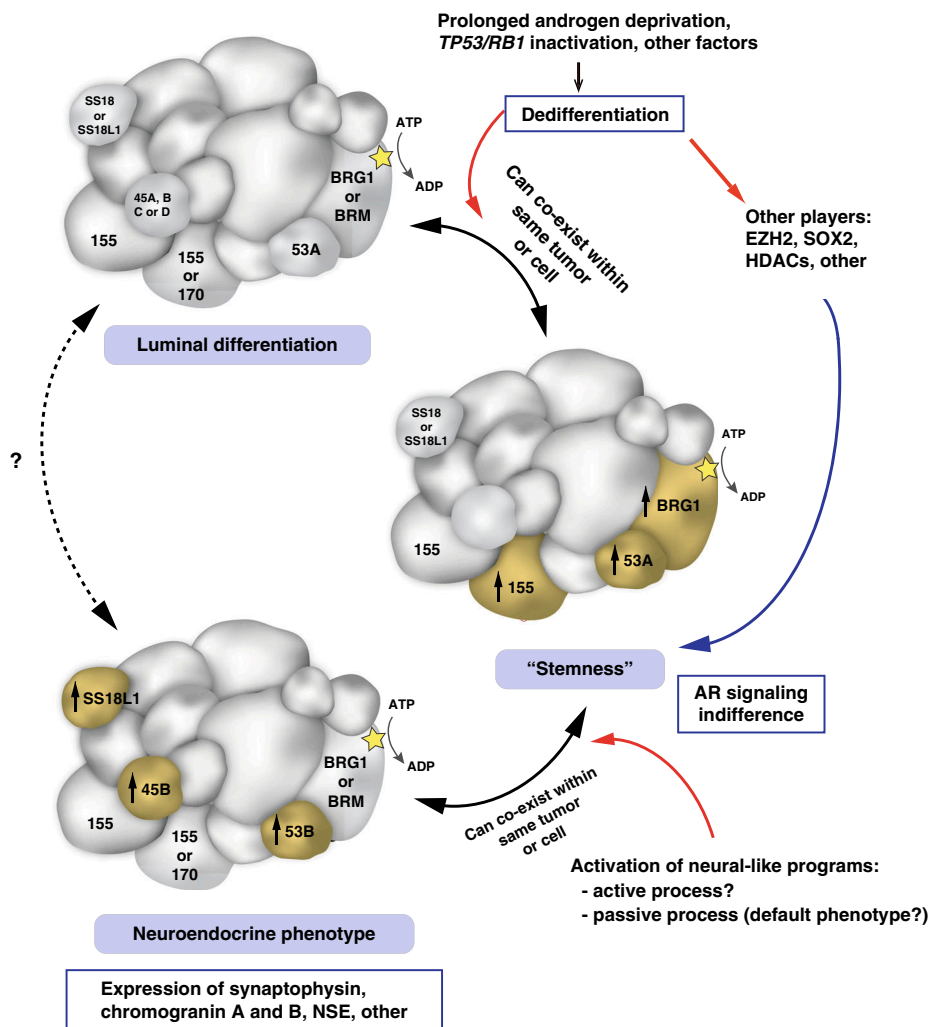


Fig. 5 Schematic representation of putative specialized SWI/SNF assemblies in prostate cancer cells. Hypothetical SWI/SNF assemblies are shown in the context of current knowledge about prostate cancer phenotype plasticity. Subunits of particular interest are annotated with their names. Two names within a subunit indicate possible incorporation of either one of the two paralogs. Subunit sizes are approximately indicative of their molecular weights. 155: BAF155, 170: BAF170, 53A: BAF53A, 53B: BAF53B, 45B: BAF45B, AR: Androgen receptor.

unique hormone treatment-naïve prostatic adenocarcinoma (Adeno), 245 castration resistant prostate adenocarcinoma (CRPC-Adeno), and 56 castration resistant neuroendocrine prostate carcinoma (CRPC-NE) patients. Two-tailed proportion test has been used to check enrichment of hemizygous deletion and copy number neutral loss.

RNA-seq data analysis of human samples. RNA-seq data from 32 normal prostate samples^{89,90}, 400 localized PCa^{37,89,90} and 120 CRPC-Adenos and 20 CRPC-NE patients^{4,5} were utilized for the initial investigation of the SWI-SNF complex units levels and were processed as follows. Reads (FASTQ files) were mapped to the human genome reference sequence (hg19/GRC37) using STAR v2.3.0e⁹¹, and the resulting alignment files were converted into Mapped Read Format (MRF) for gene expression quantification using RSEQtools⁹² and GENCODE v19 (<http://www.genecodegenes.org/releases/19.html>) as reference gene annotation set. A composite model of genes based on the union of all exonic regions from all gene transcripts was used, resulting in a set of 20,345 protein-coding genes. Normalized expression levels were estimated as FPKM. After converting the FPKM via $\log_2(\text{FPKM} + 1)$, differential expression analysis was performed using Mann-Whitney Wilcoxon test. RNA-seq data of the SU2C-PCF cohort were downloaded from original study⁵. NEPC score and AR signaling score were inferred as previously described⁵. Gleason scores of the TCGA PCas were retrieved from the original study³⁷. RNA-seq data and Gleason score from the TCGA PCa dataset were retrieved from the TCGA data portal using TCGAbiolinks R package v2.12.2⁹³.

Immunohistochemistry. Immunohistochemistry (IHC) was performed on sections of formalin-fixed paraffin-embedded patient tissue (FFPE) using a Bond III

automated immunostainer and the Bond Polymer Refine detection system (Leica Microsystems, IL, USA). Slides were de-paraffinized and heat-mediated antigen retrieval using the Bond Epitope Retrieval 1 solution at pH6 (H1) or Bond Epitope Retrieval 2 solution at pH9 (H2) or enzyme-mediated antigen retrieval (E1) was performed. All antibodies, dilutions and conditions used are listed in Supplementary Table 3.

The intensity of nuclear immunostaining for SWI/SNF subunits was evaluated on tissue micro-arrays (TMAs) and whole slide sections by a pathologist (J.C.) blinded to additional pathological and clinical data, and was scored as negative (score 0), weak (score 1), moderate (score 2), or strong (score 3). Association between disease state and staining intensity (negative/weak vs. moderate/strong) was examined using the two-tailed Fisher's exact test.

Analysis of SMARCA4 and SMARCA2 expression in localized PCa vs. clinical outcome. The patient cohort with localized PCa and available clinical and follow-up information has been previously described³⁰. IHC for SMARCA4 and SMARCA2 was performed on TMAs constructed from these patients' prostatectomy specimens. Staining intensity was scored by a pathologist (J.C.) blinded to the clinical data, using the digital online TMA scoring tool Scorenado (University of Bern, Switzerland). The Kaplan-Meier method was used to estimate patients' overall survival. The association between SMARCA4 and SMARCA2 expression (strong vs. moderate/weak/negative) and overall survival was examined using the log-rank test and multivariable Cox proportional hazards regression models. Ninety-five percent confidence intervals were calculated to assess the precision of the obtained hazard ratios. All *p*-values were two-sided, and statistical significance was evaluated at the 0.05 alpha level. All analyses were performed in R (3.5.1) for Windows.

Development of a SMARCA4 knock-down signature. We defined the SMARCA4 knock-down signature by selecting a list of differentially expressed genes between SMARCA4 siRNA-mediated knock-down and Scrambled control in the LNCaP cell line with a log fold change of 1.5 and an FDR < 0.01. For each sample, gene expression data were first normalized by z-score transformation. Then signature score was calculated as a weighted sum of normalized expression of the genes in the signature and was finally re-scaled with the 2.5% and 97.5% quantiles equaled -1 and +1, respectively. We defined samples with low SMARCA4 knock-down signature score as the 25% of cases with the lowest scores, and samples with high signature score as the 25% of cases with the highest scores.

Validation of SMARCA4 knock-down signature in multiple clinical cohorts.

SMARCA4 knock-down generated signature was applied to two CRPC cohorts consisting of 332 patients from the Stand Up To Cancer-Prostate Cancer Foundation (SU2C-PCF) trial treated with ARSi (recently published by Abida et al.⁵) and 47 patients from the Weil Cornell Medicine (WCM) cohort (published by Beltran et al.⁴) and on one cohort of localized, hormone treatment-naïve PCa consisting of 495 patients from The Cancer Genome Atlas (TCGA).

Results from the signature was then correlated with NEPC score and AR signaling scores for the SU2C-PCF and the WCM dataset and with Gleason score for the TCGA dataset.

Decipher GRID analysis. For prospective Decipher GRID and JHMI cohort, tumor RNA was extracted from FFPE blocks or slides after macrodissection guided by a histologic review of the tumor lesion by a GU pathologist. RNA extraction and microarray hybridization were all done in a Clinical Laboratory Improvement Amendments (CLIA)-certified laboratory facility (GenomeDx Biosciences, San Diego, CA, USA). Total RNA was amplified and hybridized to Human Exon 1.0 ST GeneChips (Thermo-Fisher, Carlsbad, CA). All data was normalized using the Single Channel Array Normalization (SCAN) algorithm⁹⁴. Decipher scores were calculated based on the predefined 22-markers³⁸. Patients with high Decipher (>0.7) were categorized as genomically high risk patients. Mann-Whitney U test was used to assess score differences across Gleason score groups and Mann-Kendall trend test was used to test the association between the percentage of high Decipher scores across deciles of the SMARCA4 knock-down signature. Kaplan-Meier analysis and Cox proportional hazard model was used to associate SMARCA4 knock-down signature with time to metastasis in the JHMI cohort.

Cell culture. Commercially available PCa cell lines (RWPE-1, LNCaP, 22Rv1, VCaP, LAPC4, PC3, DU145, NCI-H660, C4-2) were purchased from ATCC and maintained according to ATCC protocols. WCM154 and WCM155 CRPC-NE cell lines have been previously established and were maintained in two-dimensional monolayer culture according to the previously described protocol²⁹. LNCaP-AR cells were a kind gift from Dr. Sawyers and Dr. Mu (Memorial Sloan Kettering Cancer Center) and were cultured as previously described¹². MSKCC-PCa3 CRPC-Adeno cells were a kind gift from Dr. Chen (Memorial Sloan Kettering Cancer Center) and were maintained identically to WCM154 and WCM155 cells. All cell lines used and their phenotype are listed in Supplementary Table 4. Cell cultures were regularly tested for *Mycoplasma* contamination and confirmed to be negative.

Cell transfection and siRNA-mediated knock-down. ON-TARGET plus siRNA SMARTpool siRNAs against SMARCA4, SMARCA2, SMARCC1, SMARCC2, and REST were purchased from Dharmacon. Transfection was performed overnight on attached cells growing in 6-well plates using the Lipofectamine 3000 reagent (Thermo Fisher Scientific) to the proportions of 10 μ L of 20 μ M siRNA per well. Cells were harvested for protein and RNA extraction 72 h after transfection.

Cell infection, shRNA-mediated knock-down and gene overexpression. The *ACTL6B* shRNA and the matching Scrambled shRNA control were a kind gift from Dr. Cigall Kadoch (Dana Farber Cancer Institute). The vector was pGIPZ and the target sequence was: sh#1-TGGATCACACCTACAGCAA. The *DPF1* shRNA and the corresponding Scrambled shRNA control were purchased from Genecopoeia. The vector was psi-LVRU6GP and the target sequences were: sh#1-GAATTAAC TGTCTGTGTAT, Scrambled control-GCTTCGCGCCGTAGTCTTA. For infection, WCM155 cells were collected, resuspended in media containing Polybrene (Millipore) and lentiviral particles, and centrifuged at 800 \times g at room temperature for 60 min. Both vectors included a GFP reporter and infection efficiency was confirmed by green fluorescence. Cells were harvested for protein and RNA extraction 72 h after transfection. Given the short-term nature of the experiments, selection was not performed. For the SMARCA4 or SMARCA2 overexpression experiment, lentiviral particles were prepared as described above using the pEZ-Lv203 vector (SMARCA4 gene, eGFP reporter), the pEZLv216 vector (SMARCA2 gene, mCherry reporter) (all vectors Genecopoeia, MD, USA; all sequence-verified). 22Rv1 cells were infected as described above, cultured and sorted based on the expression of the fluorescent reporter.

Immunoblotting. Cells were lysed in RIPA buffer with protease and phosphatase inhibitors (Thermo Fisher Scientific) and total protein concentration was measured

using the DC Protein Assay (Bio-Rad). Protein samples were resolved in SDS-PAGE, transferred onto a nitrocellulose membrane using the iBlot 2 dry blotting system (Thermo Fisher Scientific) and incubated overnight at 4 °C with primary antibodies dissolved in 5% Blotting-Grade Blocker (Bio-Rad). All primary antibodies and dilutions used are listed in Supplementary Table 3. After 3 washes, the membrane was incubated with secondary antibody conjugated to horseradish peroxidase for 1 h at room temperature. After 3 washes, signal was visualized by chemiluminescence using the Luminata Forte substrate (Thermo Fisher Scientific) and images were acquired with the ChemiDoc™ Touch Imaging System (Bio-Rad, Hercules, CA). When blotting of a single membrane for different proteins was necessary, the membrane was stripped using the Restore PLUS Stripping Buffer (Thermo Fisher Scientific) according to producer's instructions and the immunoblotting process was repeated.

RNA extraction from cells, RNA sequencing and analysis, qPCR. Total RNA was extracted from cells using the Maxwell 16 LEV simplyRNA Purification Kit and the Maxwell 16 Instrument. RNA integrity was verified using the Agilent Bioanalyzer 2100 (Agilent Technologies). cDNA was synthesized from total RNA using Superscript III (Invitrogen). Library preparation was performed using TruSeq RNA Library Preparation Kit v2. RNA sequencing was performed on the HiSeq 2500 sequencer to generate 2 \times 75 bp paired-end reads.

Sequence reads were aligned using STAR two-pass⁹⁵ to the human reference genome GRCh37. Gene counts were quantified using the "GeneCounts" option. Per-gene counts-per-million (CPM) were computed and log₂-transformed adding a pseudo-count of 1 to avoid transforming 0. Genes with log₂-CPM < 1 in more than three samples were removed. Unsupervised clustering was performed using the top 500 most variable genes, Euclidean distance as the distance metric and the Ward clustering algorithm. When required, the batch effect was removed using the function removeBatchEffect from the limma R package for data visualization. For differential expression the batch factor was included in the design matrix.

Differential expression analysis between knock-down cells and control samples was performed using the edgeR v3.28.1 package⁹⁶. Normalization was performed using the "TMM" (weighted trimmed mean) method and differential expression was assessed using the quasi-likelihood F-test.

Genes with FDR < 0.05 and >2-fold were considered significantly differentially expressed.

Gene Set Enrichment Analysis (GSEA) was performed using the Preranked tool⁹⁷ for C2 (canonical pathways) and H (hallmark gene sets)⁹⁸. Genes were ranked based on the T-statistic from the differential expression analysis.

Primer sequences used for qPCR are available in Supplementary Table 5.

Cell growth experiments. Cells were treated with siRNA (3 pmol) against SMARCA4, SMARCA2, SMARCC1, SMARCC2 or with a scrambled control for 24 h. LNCaP and C4-2 cells were then seeded in Poly-L-Lysine coated 96-well plates (2000 cells/well) and WCM154 cells were seeded in a collagen-coated 96-well plates (5000 cells/well). Cell viability was determined after 24, 48, 72, and 96 h with a Tecan Infinite M200PRO reader using the CellTiter-Glo® Luminescent Cell Viability Assay according to manufacturer's directions (Promega). Cell confluence was determined using the Incucyte S3 instrument and the IncuCyte S3 2018B software (Essen Bioscience, Germany). Values were calculated as x-fold of cells transfected with siRNA for 0 h.

Co-immunoprecipitation and mass spectrometry analysis. For the co-immunoprecipitation (co-IP) using an anti-BAF155 antibody, nuclear fractions of LNCaP-AR and NCI-H660 cells were isolated using the Universal CoIP Kit (Activ Motif). Chromatin of the nuclear fraction was mechanically sheared using a Dounce homogenizer. Nuclear membrane and debris were pelleted by centrifugation and protein concentration of the cleared lysate was determined with the Pierce BCA Protein Assay Kit (Thermo Fisher Scientific). 2 μ g of the anti-BAF155 antibody (ab172638, Abcam) and 2 μ g of rabbit IgG Isotype Control antibody (026102, Thermo Fisher Scientific) were incubated with 2 mg protein supernatant overnight at 4 °C with gentle rotation. The following morning, 30 μ l of Protein G Magnetic Beads (Active Motif) were washed twice with 500 μ l CoIP buffer and incubated with Antibody-containing lysate for 1 h at 4 °C with gentle rotation. Bead-bound SWI/SNF complexes were washed 3 times with CoIP buffer and twice with a buffer containing 150 mM NaCl, 50 mM Tris-HCl (pH 8) and Protease and Phosphatase inhibitors. Air-dried and frozen (-20 °C) beads were subjected to mass spectrometry (MS) analysis. Proteins on the affinity pulldown beads were re-suspended in 8 M Urea/50 mM Tris-HCl pH 8, reduced 30 min at 37 °C with DTT 0.1 M/100 mM Tris-HCl pH 8, alkylated 30 min at 37 °C in the dark with IAA 0.5 M/100 mM Tris-HCl pH 8, diluted with 4 volumes of 20 mM Tris-HCl pH 8/2 mM CaCl₂ prior to overnight digestion at room temperature with 100 ng sequencing grade trypsin (Promega). Samples were centrifuged and the magnetic beads trapped by a magnet holder in order to extract the peptides in the supernatant.

The digests were analyzed by liquid chromatography (LC)-MS/MS (PROXEON coupled to a QExactive HF mass spectrometer, ThermoFisher Scientific) with three injections of 5 μ l digests. Peptides were trapped on a μ Precolumn C18 PepMap100 (5 μ m, 100 Å, 300 μ m \times 5 mm, ThermoFisher Scientific, Reinach, Switzerland) and

separated by backflush on a C18 column (5 μm , 100 \AA , 75 μm \times 15 cm, C18) by applying a 60-min gradient of 5% acetonitrile to 40% in water, 0.1% formic acid, at a flow rate of 350 nl/min. The Full Scan method was set with resolution at 60,000 with an automatic gain control (AGC) target of 1E06 and maximum ion injection time of 50 ms. The data-dependent method for precursor ion fragmentation was applied with the following settings: resolution 15,000, AGC of 1E05, maximum ion time of 110 ms, mass window 1.6 m/z , collision energy 28, under fill ratio 1%, charge exclusion of unassigned and 1+ ions, and peptide match preferred, respectively.

MS data was interpreted with MaxQuant (version 1.6.1.0) against a SwissProt human database (release 2019_02) using the default MaxQuant settings, allowed mass deviation for precursor ions of 10 ppm for the first search, maximum peptide mass of 5500 Da, match between runs activated with a matching time window of 0.7 min and the use of non-consecutive fractions for the different pulldowns to prevent over-fitting. Settings that differed from the default setting included: strict trypsin cleavage rule allowing for 3 missed cleavages, fixed carbamidomethylation of cysteines, variable oxidation of methionines and acetylation of protein N-termini.

Protein intensities are reported as MaxQuant's Label Free Quantification (LFQ) values, as well as Top3 values (sum of the intensities of the three most intense peptides); for the latter, variance stabilization was used for the peptide normalization, and missing peptide intensities were imputed in the following manner: if there was at least two evidences in one group of replicates, the missing value was drawn from a Gaussian distribution of width 0.3 centered at the sample distribution mean minus 1.8x the sample standard deviation. Imputation at protein level for both LFQ and Top3 values was performed if there were at least two measured intensities in at least one group of replicates; missing values in this case were drawn from a Gaussian distribution of width 0.2 centered at the sample distribution mean minus 2.5x the sample standard deviation. Differential expression tests were performed using the moderated *t*-test *empirical Bayes* (R function EBAYes from the limma package version 3.40.6) on imputed LFQ and Top3 protein intensities. The Benjamini and Hochberg method was further applied to correct for multiple testing. The criterion for statistically significant differential expression is that the maximum adjusted *p*-value for large fold changes is 0.05, and that this maximum decreases asymptotically to 0 as the \log_2 fold change of 1 is approached (with a curve parameter of one time the overall standard deviation).

Please see below, the description of the methods for the second Co-IP (validation experiment) using an anti-BRG1 antibody in LNCaP and NCI-H660 cells.

CRISPR-Cas9 mediated TP53 and RB1 knock-out. To generate the stable p53 and RB1 knockout cells, all-in-one CRISPR plasmids with mCherry reporter were purchased from Genecopia (Cat # HCP218175-CG01, HCP216131-CG01). Cells were transfected with CRISPR plasmids, selected with puromycin and sorted for mCherry positivity. TP53 gRNA sequences used: TCGACGCTAGGATCTGACTG, CGTCGAGCCCCCTCTGAGTC, CCATTGTTCAATATCGTCCG. RB1 gRNA sequences used: CGGTGGCGGCCGTTTTCGG, CGGTGCCGGGGTCCGC GG, CGGAGGACCTGCCTCTCGTC. Control gRNA sequence: GGGTCCGGCC GTAGTCTTA.

Single cell RNA-sequencing (scRNAseq). scRNAseq was performed for two CRPC-NE organoids in 3D culture: MSK PCa1 and MSK PCa16. Cell counting and viability tests were conducted using a Moxi Go II Flow Cytometer (Orflo Technologies) with trypan blue and Propidium Iodide staining. Subsequently, GEM generation and barcoding, reverse transcription, cDNA amplification and 3' Gene Expression library generation steps were all performed according to the Chromium Single Cell 3' Reagent Kits v3 user Guide (10x Genomics CG000183 Rev B). Specifically, 32.0, 11.4, and 40.0 μL of PCa1, PCa8, and PCa16 cell suspension (100, 750, and 200 cells/ μL) were used for a targeted cell recovery of 2000, 5000, and 2000 cells, respectively. GEM generation was followed by a GEM-reverse transcription incubation, a clean-up step and 12 cycles of cDNA amplification. The resulting cDNA was assessed for quantity and quality using fluorometry and capillary electrophoresis, respectively. The cDNA libraries were pooled and sequenced paired-end and single indexed on an illumina NovaSeq 6000 sequencer with a shared NovaSeq 6000 S2 Reagent Kit (100 cycles). The read-set up was as follows: read 1 = 28 cycles, i7 index = 8 cycles, i5 = 0 cycles and read 2 = 91 cycles. An average of 300,753,777 reads/library were obtained, equating to an average of 111, 978 reads/cell. All steps were performed at the Next Generation Sequencing Platform, University of Bern. Data demultiplexing was performed using SEURAT v. 3.1.5 package (PMID 29608179). Low quality cells and multiplets were excluded by removing cells with unique feature counts over 5500 or less than 1000. Cells containing mitochondrial gene counts greater than 25% were also removed. Data were then scaled to 10,000 and log transformed. Only cells expressing SOX2 and SMARCA4 genes were included. Boxplots were drawn using GGPlot2 3.3.0 (<https://ggplot2.tidyverse.org>) and *p*-values were calculated using Wilcoxon test.

ChIP-sequencing data analysis. ChIP-seq peaks for SMARCC1 and HOXB13 in LNCaP cells were downloaded from GEO: GSE110655 and GSE94682, respectively. ChIP-seq peaks for H3K27ac and H3K27me3 in LNCaP cells were from data

published by Sandoval et al.⁷¹. Peak comparison was performed using BEDTOOLS v2-29.0 (<https://bedtools.readthedocs.io/en/latest/#>).

Co-immunoprecipitation using the anti-SMARCA4 antibody and mass spectrometry analysis. For the second Co-IP (validation experiment) using an anti-SMARCA4 antibody (results shown in Supplementary Fig. 26 and Supplementary Data 9), SWI/SNF complexes were isolated from the nuclear fraction of LNCaP (adenocarcinoma) or NCI-H660 (CRPC-NE) cells, which was prepared using the Universal CoIP Kit (Active Motif). Briefly, anti-Brg-1 antibodies (H-10, Santa Cruz Biotechnology) were cross-linked using Dimethyl pimelimidate dihydrochloride (Sigma-Aldrich) to Protein G conjugated magnetic beads (Bio-Rad). 30 μg of cross-linked antibodies were incubated with 0.8–1 mg of nuclear lysates overnight. Bead-bound BAF complexes were washed and eluted using 8 M urea buffer. The obtained protein complexes were subjected to immunoblotting and MS analysis.

For MS analysis, the eluted proteins were precipitated with trichloroacetic acid (TCA, 20% w/v), rinsed three times with acetone, and dried at room temperature. The pellets were re-suspended in 50 μL resuspension buffer (8 M urea, 50 mM ammonium bicarbonate, and 5 mM DTT) and subjected to reduction and alkylation by adding 15 mM iodoacetamide to each sample for 30 min in the dark at room temperature, followed by addition of 5 mM DTT to quench the reaction. Samples were diluted to a final concentration of 2 M urea and digested with LysC at room temperature overnight, and then diluted further to 1 M urea and digested with Trypsin at 37 $^{\circ}\text{C}$ overnight (for each enzyme a ratio of 1:125 enzyme:protein was used).

Samples were labeled using reductive dimethylation. Labeling was done while the peptides were bound to the solid phase C18 resin in self-packed STAGE Tip micro-columns⁹⁹. Stage tips were washed with methanol, acetonitrile (ACN) 70% v/v and formic acid (FA) 1% v/v. Samples were acidified by adding 100% FA to a final concentration of 2% FA before loading. After sample loading, stage tips were washed with 1% FA and phosphate/citrate buffer (0.23 M sodium phosphate and 86.4 mM citric acid [pH 5.5]). At this point, the "light" solution (0.4% CH_2O and 60 mM NaBH_3CN), or "heavy" solution (0.4% CD_2O and 60 mM NaBD_3CN) was added twice on each stage tip to label the peptides. A final wash with 1% FA was performed prior to elution with 70% ACN and 1% FA. Samples were dried under vacuum, resuspended in 5% FA, and mixed together in equal amounts for analysis using an Orbitrap Fusion Mass Spectrometer. Peptides were introduced into the mass spectrometer by nano-electrospray as they eluted off a self-packed 40 cm, 75 μm (ID) reverse-phase column packed with 1.8 μm , 120 \AA pore size, SEPAX C18 resin. Peptides were separated with a gradient of 5–25% buffer B (99.9% ACN, 0.1% FA) with a flow rate of 350 nl/min for 65 min. For each scan cycle, one high mass resolution full MS scan was acquired in the Orbitrap mass analyzer at a resolution of 120 K, AGC value of 500,000, in a m/z scan range of 375–1400, max acquisition time of 100 ms and up to 20 parent ions were chosen based on their intensity for collision induced dissociation (normalized collision energy = 35%) and MS/MS fragment ion scans at low mass resolution in the linear ion trap. Dynamic exclusion was enabled to exclude ions that had already been selected for MS/MS in the previous 40 s. Ions with a charge of +1 and those whose charge state could not be assigned were also excluded. All scans were collected in centroid mode. Two biological replicates for each condition were processed and analyzed.

MS2 spectra were searched using SEQUEST (version 28 revision 13) against a composite database containing all Swiss-Prot reviewed human protein sequences (20,193 target sequences, downloaded from www.uniprot.org March 18, 2016) and their reversed complement, using the following parameters: a precursor mass tolerance of ± 25 ppm; 1.0 Da product ion mass tolerance; tryptic digestion; up to two missed cleavages; static modifications of carbamidomethylation on cysteine (+57.0214) and dimethylation on n-termini and lysines (+28.0313); dynamic modifications of methionine oxidation (+15.9949) and heavy dimethylation on N-termini and lysines (+6.03766). Peptide spectral matches (PSMs) were filtered to 1% FDR using the target-decoy strategy¹⁰⁰ combined with linear discriminant analysis (LDA)¹⁰¹ using several different parameters including Xcorr, ΔCn , precursor mass error, observed ion charge state, and predicted solution charge state. Linear discriminant models were calculated for each LC-MS/MS run using peptide matches to forward and reversed protein sequences as positive and negative training data. PSMs within each run were sorted in descending order by discriminant score and filtered to a 1% FDR as revealed by the number of decoy sequences remaining in the data set. The data were further filtered to control protein level FDRs. Peptides were combined and assembled into proteins. Protein scores were derived from the product of all LDA peptide probabilities, sorted by rank, and filtered to 1% FDR as described for peptides. The FDR of the remaining peptides fell dramatically after protein filtering. The data were further filtered to require a minimum of 8 PSMs per protein. All peptides were required to have a sum of heavy and light signal-to-noise (SN) ≥ 10 . Protein ratios were calculated as the \log_2 ratio of the total SN of all experimental sample peptide values over that for IgG control sample peptides. For a small number of the most highly enriched proteins, the control value was zero (this is the theoretical ideal). In these cases, we imputed a value of one for ratio calculations. Subsequent visualization and statistical analysis was done with Perseus and R program¹⁰².

SMARCC1 Co-IP immunoblotting. For the co-immunoprecipitation (co-IP), using an anti-BAF155 antibody, nuclear fractions of LNCaP-AR and NCI-H660 cells were isolated using the Universal CoIP Kit (Active Motif). Chromatin of the nuclear fraction was mechanically sheared using a Dounce homogenizer. Nuclear membrane and debris were pelleted by centrifugation and protein concentration of the cleared lysate was determined with the Pierce BCA Protein Assay Kit (Thermo Fisher Scientific). One microgram of the rabbit anti-BAF155 antibody (ab172638, Abcam) and 1 µg of rabbit IgG Isotype Control antibody (026102, Thermo Fisher Scientific) were incubated with 1 mg protein supernatant overnight at 4 °C with gentle rotation. The following morning, 30 µl of Protein G Magnetic Beads (Active Motif) were washed twice with 500 µl CoIP buffer and incubated with Antibody-containing lysate for 2 h at 4 °C with gentle rotation. Bead-bound SWI/SNF complexes were washed twice with CoIP buffer and twice with a buffer containing 150 mM NaCl, 50 mM Tris-HCl (pH 8) and Protease and Phosphatase inhibitors. Washing procedure was executed at 4 °C with gentle rotation. Bead-bound protein and Input controls are reduced and denatured in 40 µl Laemmli buffer containing DTT through boiling for 5 min at 95 °C. Magnetic beads are removed from solution and 20 µl of reduce protein is loaded on an SDS-PAGE gel with subsequent immunoblotting using iBlot (Life Technologies). Membranes were blocked in 5% dry-milk solution and then incubated over night with respective antibodies against targets of interest. Protein signal was detected using HRP-labeled native anti-rabbit IgG antibody (CST, #5127) and ECL substrate solution (Merck Millipore) using the Fusion FX.

RNA isolation and qPCR. Cells were first seeded in 10cm-petridish and grown until they reached a confluency of approx. 90%. The cells were then harvested for RNA isolation using the RNeasy Mini Kit (Qiagen). Synthesis of complementary DNAs (cDNAs) using FIREScript RT cDNA Synthesis Kit (Solis BioDyne) and real-time reverse transcription PCR (RT-PCR) assays using HOT FIREPol Eva-Green qPCR Mix Plus (Solis BioDyne) were performed using and applying the manufacturer protocols. Relative mRNA levels of each gene shown were normalized to the expression of the average of housekeeping genes GAPDH and ACTB. The sequences of the primers for qRT-PCR assays can be found in Supplementary Table 5.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Data generated during this study have been submitted on the European Genome-phenome Archive under the accession EGAS00001004177 (<https://ega-archive.org/datasets/EGAD00001005800>). The mass spectrometry proteomics data that support the findings of this study have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository with the dataset identifier PXD016861. Source data are provided with this paper.

Received: 8 February 2020; Accepted: 7 October 2020;
Published online: 03 November 2020

References

- Bray, F. et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **68**, 394–424 (2018).
- Rawla, P. Epidemiology of prostate cancer. *World J. Oncol.* **10**, 63–89 (2019).
- Robinson, D. et al. Integrative clinical genomics of advanced prostate. *Cancer Cell* **162**, 454 (2015).
- Beltran, H. et al. Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. *Nat. Med.* **22**, 298–305 (2016).
- Abida, W. et al. Genomic correlates of clinical outcome in advanced prostate cancer. *Proc. Natl Acad. Sci. USA* **116**, 11428–11436 (2019).
- Epstein, J. I. et al. Proposed morphologic classification of prostate cancer with neuroendocrine differentiation. *Am. J. Surg. Pathol.* **38**, 756–767 (2014).
- Rickman, D. S., Beltran, H., Demichelis, F. & Rubin, M. A. Biology and evolution of poorly differentiated neuroendocrine tumors. *Nat. Med.* **23**, 1–10 (2017).
- Aggarwal, R. et al. Clinical and genomic characterization of treatment-emergent small-cell neuroendocrine prostate cancer: a multi-institutional prospective study. *J. Clin. Oncol.* **36**, 2492–2503 (2018).
- Bluemn, E. G. et al. Androgen receptor pathway-independent prostate cancer is sustained through FGF signaling. *Cancer Cell* **32**, 474–489 e6 (2017).
- Metzger, A. L. et al. Patterns of care and outcomes in small cell carcinoma of the prostate: a national cancer database analysis. *Prostate* **79**, 1457–1461 (2019).
- Ku, S. Y. et al. Rb1 and Trp53 cooperate to suppress prostate cancer lineage plasticity, metastasis, and antiandrogen resistance. *Science* **355**, 78–83 (2017).
- Mu, P. et al. SOX2 promotes lineage plasticity and antiandrogen resistance in TP53- and RB1-deficient prostate cancer. *Science* **355**, 84–88 (2017).
- Davies, A. H., Beltran, H. & Zoubeidi, A. Cellular plasticity and the neuroendocrine phenotype in prostate cancer. *Nat. Rev. Urol.* **15**, 271–286 (2018).
- Kadoch, C. et al. Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. *Nat. Genet.* **45**, 592–601 (2013).
- Shain, A. H. & Pollack, J. R. The spectrum of SWI/SNF mutations, ubiquitous in human cancers. *PLoS ONE* **8**, e55119 (2013).
- Kadoch, C. & Crabtree, G. R. Mammalian SWI/SNF chromatin remodeling complexes and cancer: mechanistic insights gained from human genomics. *Sci. Adv.* **1**, e1500447 (2015).
- Hargreaves, D. C. & Crabtree, G. R. ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. *Cell Res.* **21**, 396–420 (2011).
- Ho, L. et al. An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. *Proc. Natl Acad. Sci. USA* **106**, 5187–5191 (2009).
- Ho, L. et al. esBAF facilitates pluripotency by conditioning the genome for LIF/STAT3 signalling and by regulating polycomb function. *Nat. Cell Biol.* **13**, 903–913 (2011).
- Yoo, A. S., Staahl, B. T., Chen, L. & Crabtree, G. R. MicroRNA-mediated switching of chromatin-remodelling complexes in neural development. *Nature* **460**, 642–646 (2009).
- Wu, J. I. et al. Regulation of dendritic development by neuron-specific chromatin remodeling complexes. *Neuron* **56**, 94–108 (2007).
- Lessard, J. et al. An essential switch in subunit composition of a chromatin remodeling complex during neural development. *Neuron* **55**, 201–215 (2007).
- Versteeg, I. et al. Truncating mutations of hSNF5/INI1 in aggressive paediatric cancer. *Nature* **394**, 203–206 (1998).
- Kadoch, C. & Crabtree, G. R. Reversible disruption of mSWI/SNF (BAF) complexes by the SS18-SSX oncogenic fusion in synovial sarcoma. *Cell* **153**, 71–85 (2013).
- Witkowski, L. et al. Germline and somatic SMARCA4 mutations characterize small cell carcinoma of the ovary, hypercalcemic type. *Nat. Genet.* **46**, 438–443 (2014).
- Le Loarer, F. et al. SMARCA4 inactivation defines a group of undifferentiated thoracic malignancies transcriptionally related to BAF-deficient sarcomas. *Nat. Genet.* **47**, 1200–1205 (2015).
- Biegel, J. A., Busse, T. M. & Weissman, B. E. SWI/SNF chromatin remodeling complexes and cancer. *Am. J. Med. Genet. C Semin. Med. Genet.* **166C**, 350–366 (2014).
- Hoffman, G. R. et al. Functional epigenetics approach identifies BRM/SMARCA2 as a critical synthetic lethal target in BRG1-deficient cancers. *Proc. Natl Acad. Sci. USA* **111**, 3128–3133 (2014).
- Puca, L. et al. Patient derived organoids to model rare prostate cancer phenotypes. *Nat. Commun.* **9**, 2404 (2018).
- Spahn, M. et al. Expression of microRNA-221 is progressively reduced in aggressive prostate cancer and metastasis and predicts clinical recurrence. *Int. J. Cancer* **127**, 394–403 (2010).
- Prensner, J. R. et al. The long non-coding RNA PCAT-1 promotes prostate cancer cell proliferation through cMyc. *Neoplasia* **16**, 900–908 (2014).
- Coppola, V. et al. BTG2 loss and miR-21 upregulation contribute to prostate cell transformation by inducing luminal markers expression and epithelial-mesenchymal transition. *Oncogene* **32**, 1843–1853 (2013).
- Ding, Y. et al. Chromatin remodeling ATPase BRG1 and PTEN are synthetic lethal in prostate cancer. *J. Clin. Invest.* **129**, 759–773 (2019).
- Muthuswami, R. et al. BRG1 is a prognostic indicator and a potential therapeutic target for prostate cancer. *J. Cell Physiol.* **234**, 15194–15205 (2019).
- Hieronymus, H. et al. Gene expression signature-based chemical genomic prediction identifies a novel class of HSP90 pathway modulators. *Cancer Cell* **10**, 321–330 (2006).
- Epstein, J. I. et al. A contemporary prostate cancer grading system: a validated alternative to the Gleason score. *Eur. Urol.* **69**, 428–435 (2015).
- Cancer Genome Atlas Research, N. The molecular taxonomy of primary prostate cancer. *Cell* **163**, 1011–1025 (2015).
- Spratt, D. E. et al. Individual patient-level meta-analysis of the performance of the decipher genomic classifier in high-risk men after prostatectomy to predict development of metastatic disease. *J. Clin. Oncol.* **35**, 1991–1998 (2017).
- Ross, A. E. et al. Tissue-based genomics augments post-prostatectomy risk stratification in a natural history cohort of intermediate- and high-risk men. *Eur. Urol.* **69**, 157–165 (2016).
- Chen, C. D. et al. Molecular determinants of resistance to antiandrogen therapy. *Nat. Med.* **10**, 33–39 (2004).
- Raupach, E. A. et al. Novel functional insights revealed by distinct protein-protein interactions of the residual SWI/SNF complex in SMARCA4-deficient small cell carcinoma of the ovary, hypercalcemic type. *bioRxiv* 794776 (2019).

42. Chatterjee, S. S., Biswas, M. & Sengupta, A. Transcriptional cooperativity between SWI/SNF and nurd chromatin remodelers in acute myeloid leukemia. *Exp. Hematol.* **64**, S58 (2018).
43. Yao, J. et al. The homeobox gene, HOXB13, regulates a mitotic protein-kinase interaction network in metastatic prostate cancers. *Sci. Rep.* **9**, 9715 (2019).
44. Artegianni, B. et al. The interaction with HMG20a/b proteins suggests a potential role for beta-dystrobrevin in neuronal differentiation. *J. Biol. Chem.* **285**, 24740–24750 (2010).
45. Gupta, K. & Gupta, S. Neuroendocrine differentiation in prostate cancer: key epigenetic players. *Transl. Cancer Res.* **6**, S104–s108 (2017).
46. Li, Y. et al. SRRM4 drives neuroendocrine transdifferentiation of prostate adenocarcinoma under androgen receptor pathway inhibition. *Eur. Urol.* **71**, 68–78 (2017).
47. Fraser, J. A., Sutton, J. E., Tazayoni, S., Bruce, I. & Poole, A. V. hASH1 nuclear localization persists in neuroendocrine transdifferentiated prostate cancer cells, even upon reintroduction of androgen. *Sci. Rep.* **9**, 19076 (2019).
48. Zhu, Y. et al. Interleukin-6 induces neuroendocrine differentiation (NED) through suppression of RE-1 silencing transcription factor (REST). *Prostate* **74**, 1086–1094 (2014).
49. Zhang, X. et al. SRRM4 expression and the loss of REST activity may promote the emergence of the neuroendocrine phenotype in castration-resistant prostate cancer. *Clin. Cancer Res.* **21**, 4698–4708 (2015).
50. Chen, R., Li, Y., Buttyan, R. & Dong, X. Implications of PI3K/AKT inhibition on REST protein stability and neuroendocrine phenotype acquisition in prostate cancer cells. *Oncotarget* **8**, 84863–84876 (2017).
51. Chang, Y. T. et al. REST is a crucial regulator for acquiring EMT-like and stemness phenotypes in hormone-refractory prostate cancer. *Sci. Rep.* **7**, 42795 (2017).
52. Small, E. J. et al. Neuroendocrine prostate cancer in patients with metastatic castration resistant prostate cancer resistant to abiraterone or enzalutamide: preliminary results from the SU2C/PCF/AACR west coast prostate cancer dream team. In *ESMO Vol. 25*, iv255–iv279 (Annals of Oncology, 2014).
53. Khalaf, D. J. et al. Optimal sequencing of enzalutamide and abiraterone acetate plus prednisone in metastatic castration-resistant prostate cancer: a multicentre, randomised, open-label, phase 2, crossover trial. *Lancet Oncol.* **20**, 1730–1739 (2019).
54. Zou, M. et al. Transdifferentiation as a mechanism of treatment resistance in a mouse model of castration-resistant prostate cancer. *Cancer Discov.* **7**, 736–749 (2017).
55. Zehir, A. et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat. Med.* **23**, 703–713 (2017).
56. Roy, N. et al. Brg1 promotes both tumor-suppressive and oncogenic activities at distinct stages of pancreatic cancer formation. *Genes Dev.* **29**, 658–671 (2015).
57. Velcheti, V., Wong, K. K. & Sauntharajah, Y. EZH2 Inhibitors: take it EZy, it is all about context. *Cancer Discov.* **9**, 472–475 (2019).
58. Sun, X. et al. Arid1a has context-dependent oncogenic and tumor suppressor functions in liver cancer. *Cancer Cell* **32**, 574–589 e6 (2017).
59. Jones, S. et al. Frequent mutations of chromatin remodeling gene ARID1A in ovarian clear cell carcinoma. *Science* **330**, 228–231 (2010).
60. Buscariello, M. et al. Essential role of BRG, the ATPase subunit of BAF chromatin remodeling complexes, in leukemia maintenance. *Blood* **123**, 1720–1728 (2014).
61. Wu, Q. et al. The BRG1 chromatin remodeling enzyme links cancer cell metabolism and proliferation. *Oncotarget* **7**, 38270–38281 (2016).
62. Hiramatsu, H. et al. The role of the SWI/SNF chromatin remodeling complex in maintaining the stemness of glioma initiating cells. *Sci. Rep.* **7**, 889 (2017).
63. Clark, J. et al. Identification of novel genes, SYT and SSX, involved in the t(X;18)(p11.2;q11.2) translocation found in human synovial sarcoma. *Nat. Genet.* **7**, 502–508 (1994).
64. Jubierre, L. et al. BRG1/SMARCA4 is essential for neuroblastoma cell viability through modulation of cell death and survival pathways. *Oncogene* **35**, 5179–5190 (2016).
65. Laurette, P. et al. Chromatin remodellers Brg1 and Bptf are required for normal gene expression and progression of oncogenic Braf-driven mouse melanoma. *Cell Death Differ.* **27**, 29–43 (2019).
66. Baca, S. C. & Garraway, L. A. The genomic landscape of prostate cancer. *Front. Endocrinol.* **3**, 69 (2012).
67. Barbieri, C. E. et al. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nat. Genet.* **44**, 685–689 (2012).
68. Armenia, J. et al. The long tail of oncogenic drivers in prostate cancer. *Nat. Genet.* **50**, 645–651 (2018).
69. Link, K. A. et al. Targeting the BAF57 SWI/SNF subunit in prostate cancer: a novel platform to control androgen receptor activity. *Cancer Res.* **68**, 4551–4558 (2008).
70. Jin, M. L., Kim, Y. W. & Jeong, K. W. BAF53A regulates androgen receptor-mediated gene expression and proliferation in LNCaP cells. *Biochem. Biophys. Res. Commun.* **505**, 618–623 (2018).
71. Sandoval, G. J. et al. Binding of TMPRSS2-ERG to BAF chromatin remodeling complexes mediates prostate oncogenesis. *Mol. Cell* **71**, 554–566 e7 (2018).
72. Prensner, J. R. et al. The long noncoding RNA SCHLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. *Nat. Genet.* **45**, 1392–1398 (2013).
73. Raab, J. R. et al. SWI/SNF remains localized to chromatin in the presence of SCHLAP1. *Nat. Genet.* **51**, 26–29 (2019).
74. Sun, A. et al. Aberrant expression of SWI/SNF catalytic subunits BRG1/BRM is associated with tumor development and increased invasiveness in prostate cancers. *Prostate* **67**, 203–213 (2007).
75. Shen, H. et al. The SWI/SNF ATPase Brm is a gatekeeper of proliferative control in prostate cancer. *Cancer Res.* **68**, 10154–10162 (2008).
76. Labrecque, M. P. et al. Molecular profiling stratifies diverse phenotypes of treatment-refractory metastatic castration-resistant prostate cancer. *J. Clin. Invest.* **130**, 4492–4505 (2019).
77. Lazzaro, D., Price, M., de Felice, M. & Di Lauro, R. The transcription factor TTF-1 is expressed at the onset of thyroid and lung morphogenesis and in restricted regions of the foetal brain. *Development* **113**, 1093–1104 (1991).
78. Sussel, L., Marin, O., Kimura, S. & Rubenstein, J. L. Loss of Nkx2.1 homeobox gene function results in a ventral to dorsal molecular respecification within the basal telencephalon: evidence for a transformation of the pallidum into the striatum. *Development* **126**, 3359–3370 (1999).
79. Minocha, S. & Herr, W. Cortical and commissural defects upon HCF-1 loss in Nkx2.1-derived embryonic neurons and glia. *Dev. Neurobiol.* **79**, 578–595 (2019).
80. Hofer, M. D. et al. The role of metastasis-associated protein 1 in prostate cancer progression. *Cancer Res.* **64**, 825–829 (2004).
81. Norris, J. D. et al. The homeodomain protein HOXB13 regulates the cellular response to androgens. *Mol. Cell* **36**, 405–416 (2009).
82. Chen, Z. et al. Diverse AR-V7 cistromes in castration-resistant prostate cancer are governed by HoxB13. *Proc. Natl Acad. Sci. USA* **115**, 6810–6815 (2018).
83. Ewing, C. M. et al. Germline mutations in HOXB13 and prostate-cancer risk. *N. Engl. J. Med.* **366**, 141–149 (2012).
84. Abeshouse, A. et al. The molecular taxonomy of primary prostate. *Cancer Cell* **163**, 1011–1025 (2015).
85. Talevich, E., Shain, A. H., Botton, T. & Bastian, B. C. CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. *PLoS Comput. Biol.* **12**, e1004873 (2016).
86. Cibulskis, K. et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213 (2013).
87. McLaren, W. et al. The ensembl variant effect predictor. *Genome Biol.* **17**, 122 (2016).
88. Prandi, D. et al. Unraveling the clonal hierarchy of somatic genomic aberrations. *Genome Biol.* **15**, 439 (2014).
89. Beltran, H. et al. Molecular characterization of neuroendocrine prostate cancer and identification of new drug targets. *Cancer Discov.* **1**, 487–495 (2011).
90. Chakravarty, D. et al. The oestrogen receptor alpha-regulated lncRNA NEAT1 is a critical modulator of prostate cancer. *Nat. Commun.* **5**, 5383 (2014).
91. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2012).
92. Habegger, L. et al. RSEQtools: a modular framework to analyze RNA-Seq data using compact, anonymized data summaries. *Bioinformatics* **27**, 281–283 (2010).
93. Colaprico, A. et al. TCGAAbioblinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* **44**, e71 (2016).
94. Piccolo, S. R. et al. A single-sample microarray normalization method to facilitate personalized-medicine workflows. *Genomics* **100**, 337–344 (2012).
95. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
96. Nikolayeva, O. & Robinson, M. D. edgeR for differential RNA-seq and ChIP-seq analysis: an application to stem cell biology. *Methods Mol. Biol.* **1150**, 45–79 (2014).
97. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
98. Liberzon, A. et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).
99. Rappsilber, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nano-electrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75**, 663–670 (2003).
100. Elias, J. E. & Gygi, S. P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **4**, 207–214 (2007).
101. Huttlin, E. L. et al. A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell* **143**, 1174–1189 (2010).
102. Tyanova, S. et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731–740 (2016).

Acknowledgements

We thank patients and their families for participating in genomics, transcriptomics, and precision cancer care studies. The authors would like to thank Cigall Kadoch from Dana Farber Cancer Institute for her important insights into SWI/SNF biology. We thank current and previous members of the Rubin lab for valuable discussions, Charles Sawyers and Ping Mu at Memorial Sloan-Kettering Cancer Center for sharing the LNCaP-AR cell line. We are grateful to Juan Miguel Mosquera, Brian Robinson, and Verena Sailer for their prostate cancer tissue contributions. We are thankful for expert assistance from the Translational Research Program at WCMC Pathology and Laboratory Medicine (Bing He, Leticia Dizon, Yifang Liu, Mai Ho) and the WCM CLC Genomics Core Facility (Jenny Xiang). We thank Inti Zlobec and Micha D. Eichmann at the University of Bern for their assistance with the Scorenado TMA scoring platform. We thank Elai Davicioni (Decipher Biosciences, Cam, USA) for providing data for gene expression analyses. We further thank the Next Generation Platform (Institute of Genetics, Bern) and Irene Keller (DBMR, Bern) for their assistance with the single cell experiments. We acknowledge expert assistance from Mariana Ricca at the University of Bern in preparing the manuscript for submission. This project has received funding from the Nuovo-Soldati Foundation (J.C.), the NIH/NCI WCM SPORE in Prostate Cancer P50-CA211024 (V.C., H.B., K.B., F.D., M.A.R.), the Swiss National Science Foundation Project grants number 310030_189149 and 31003A_169352 (M.K.-d.J.) and Ambizione grant number PZ00P3_168165 (S.P.), the NCI grants number 1R01CA233650 (P.C.), 1R01CA233650 (N.D.), R37 CA241486 (H.B.), and R01 CA125612 (F.D.) and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 648670) (F.D.).

Author contributions

J.C., A.A., and M.A.R. designed the study and the experiments. J.C., A.A., P.T., D.W., S.-S.C., S.C., M.J., and L.B. performed experiments. M.R.d.F., S.P., D.P., M.B., and F.D. performed genomic and transcriptomic analyses of patient data. M.R.d.F., S.P., R.B., and A.S. performed RNA-seq analysis of experimental data. L.R. and S.P. performed scRNA-seq and ChIP-seq analysis. M.R.d.F., S.P., and M.A. performed analyses on Decipher GRID and JHMI cohorts. V.C. and K.V.B. performed survival analyses. J.C. and M.A.R. performed pathology review and immunohistochemical evaluation. P.C. and N.D. performed mass spectrometry experiments and analyzed respective data. A.-C.U., S.B.L. performed mass spectrometry analyses. F.F., T.L.L., M.S. and M.K.-d.J. established annotated patient cohorts and provided clinical data. H.B. provided CRPC-NE patient samples, L.P., S.W., and Y.C. established patient-derived organoid models. M.A.R. provided administrative, technical and material support. J.C., A.A., and M.A.R. wrote the initial draft of the manuscript and all authors contributed to the final version.

Competing interests

H.B. has received research funding from Janssen, Astellas, Abbvie, Millennium, and Eli Lilly and consulting with Janssen, Astellas, Sanofi Genzyme, Astra Zeneca, Pfizer. L.P. is now an employee of Loxo Oncology at Lilly. M.A.R. has received research funding from Novartis, Roche, Ventana, Janssen, Astellas, Millennium, and Eli Lilly. M.A.R. is on the SAB of Neogenomics. T.L.L. has received research funding from DeepBio, Decipher, Ventana/Roche. Cornell and Bern Universities have filed a patent application on SWI/SNF diagnostic and therapeutic fields with A.A., J.C., and M.A.R. listed as inventors. The remaining authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41467-020-19328-1>.

Correspondence and requests for materials should be addressed to M.A.R.

Peer review information *Nature Communications* thanks Gong-Hong Wei and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020

Joanna Cyrta^{1,2,29}, Anke Augspach^{1,29}, Maria Rosaria De Filippo^{3,4}, Davide Prandi⁵, Phillip Thienger¹, Matteo Benelli^{5,6}, Victoria Cooley⁷, Rohan Bareja^{2,8}, David Wilkes², Sung-Suk Chae⁹, Paola Cavaliere¹⁰, Noah Dephoure^{10,11}, Anne-Christine Uldry¹², Sophie Braga Lagache¹², Luca Roma⁴, Sandra Cohen⁹, Muriel Jaquet¹, Laura P. Brandt¹, Mohammed Alshalalfa¹³, Loredana Puca¹⁴, Andrea Sboner^{2,8,15,16}, Felix Feng¹², Shangqian Wang¹⁷, Himisha Beltran^{14,18}, Tamara Lotan^{19,20,21}, Martin Spahn^{22,23}, Marianna Kruihof-de Julio^{1,3,24}, Yu Chen¹⁴, Karla V. Ballman⁷, Francesca Demichelis^{2,5}, Salvatore Piscuoglio^{4,25,26,30} & Mark A. Rubin^{1,27,28,30}✉

¹Department for BioMedical Research, University of Bern, 3008 Bern, Switzerland. ²The Caryl and Israel Englander Institute for Precision Medicine, Weill Cornell Medicine, New York, NY 10021, USA. ³Department for BioMedical Research, Urology Research Laboratory, University of Bern, 3008 Bern, Switzerland. ⁴Institute of Pathology and Medical Genetics, University Hospital Basel, University of Basel, 4051 Basel, Switzerland.

⁵Department of Cellular, Computational and Integrative Biology (CIBIO), University of Trento, 38122 Trento, Italy. ⁶Bioinformatics Unit, Hospital of Prato, 59100 Prato, Italy. ⁷Department of Healthcare Policy and Research, Division of Biostatistics and Epidemiology, Weill Cornell Medicine, New York, NY 10021, USA. ⁸Institute for Computational Biomedicine, Weill Cornell Medicine, New York, NY 10021, USA. ⁹Department of Laboratory Medicine and Pathology, Weill Cornell Medicine, New York, NY 10021, USA. ¹⁰Meyer Cancer Center, Weill Cornell Medicine, New York, NY 10021, USA. ¹¹Department of Biochemistry, Weill Cornell Medicine, New York, NY 10021, USA. ¹²Proteomics Mass Spectrometry Core Facility, University of Bern, 3010 Bern, Switzerland. ¹³Department of Radiation Oncology, Helen Diller Family Comprehensive Cancer Center, University of California at San Francisco, San Francisco, CA, USA. ¹⁴Department of Medicine, Division of Medical Oncology, Weill Cornell Medicine, New York, NY 10021, USA. ¹⁵HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, Weill Cornell Medicine, New York, NY 10021, USA. ¹⁶Meyer Cancer Center, Weill Cornell Medicine, New York, NY 10065, USA. ¹⁷Human Oncology and Pathogenesis Program and Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, NY 10065, USA. ¹⁸Department of Medical Oncology, Dana Farber Cancer

Institute, Boston, MA, USA. ¹⁹Department of Urology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA. ²⁰Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA. ²¹Department of Oncology, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA. ²²Lindenhofspital Bern, Prostate Center Bern, 3012 Bern, Switzerland. ²³Department of Urology, Essen University Hospital, University of Duisburg-Essen, 47057 Essen, Germany. ²⁴Department of Urology, Inselspital, 3010 Bern, Switzerland. ²⁵Visceral Surgery Research Laboratory, Clarunis, Department of Biomedicine, University of Basel, 4051 Basel, Switzerland. ²⁶Clarunis Universitäres Bauchzentrum Basel, 4002 Basel, Switzerland. ²⁷Inselspital, 3010 Bern, Switzerland. ²⁸Bern Center for Precision Medicine, 3008 Bern, Switzerland. ²⁹These authors contributed equally: Joanna Cyrta, Anke Augspach. ³⁰These authors jointly supervised this work: Salvatore Piscuoglio, Mark A. Rubin. ✉email: mark.rubin@dbmr.unibe.ch



Genomic correlates of clinical outcome in advanced prostate cancer

Wassim Abida^{a,1}, Joanna Cytra^{b,c,1,2}, Glenn Heller^{d,1}, Davide Prandi^e, Joshua Armenia^{f,3}, Ilsa Coleman^g, Marcin Cieslik^h, Matteo Benelli^{e,4}, Dan Robinson^h, Eliezer M. Van Allen^{ij}, Andrea Sboner^b, Tarcisio Fedrizzi^e, Juan Miguel Mosquera^b, Brian D. Robinson^b, Navonil De Sarkar^g, Lakshmi P. Kunju^h, Scott Tomlins^h, Yi Mi Wu^h, Daniel Nava Rodrigues^{k,l}, Massimo Loda^{b,m}, Anuradha Gopalanⁿ, Victor E. Reuterⁿ, Colin C. Pritchard^g, Joaquin Mateo^{k,l,5}, Diletta Bianchini^{k,l}, Susana Miranda^{k,l}, Suzanne Carreira^{k,l}, Pasquale Rescigno^{k,l}, Julie Filipenko^o, Jacob Vinson^o, Robert B. Montgomery^g, Himisha Beltran^{i,p}, Elisabeth I. Heath^{q,r}, Howard I. Scher^a, Philip W. Kantoff^a, Mary-Ellen Taplin^{i,6}, Nikolaus Schultz^{d,6}, Johann S. deBono^{k,l,6}, Francesca Demichelis^{e,6}, Peter S. Nelson^{g,6,7}, Mark A. Rubin^{b,c,6,7}, Arul M. Chinnaiyan^{h,s,6,7}, and Charles L. Sawyers^{f,t,6,7}

Contributed by Charles L. Sawyers, March 27, 2019 (sent for review February 19, 2019; reviewed by Samuel Aparicio, John T. Isaacs, and Nandita Mitra)

Heterogeneity in the genomic landscape of metastatic prostate cancer has become apparent through several comprehensive profiling efforts, but little is known about the impact of this heterogeneity on clinical outcome. Here, we report comprehensive genomic and transcriptomic analysis of 429 patients with metastatic castration-resistant prostate cancer (mCRPC) linked with longitudinal clinical outcomes, integrating findings from whole-exome, transcriptome, and histologic analysis. For 128 patients treated with a first-line next-generation androgen receptor signaling inhibitor (ARSI; abiraterone or enzalutamide), we examined the association of 18 recurrent DNA- and RNA-based genomic alterations, including androgen receptor (AR) variant expression, AR transcriptional output, and neuroendocrine expression signatures, with clinical outcomes. Of these, only *RB1* alteration was significantly associated with poor survival, whereas alterations in *RB1*, *AR*, and *TP53* were associated with shorter time on treatment with an ARSI. This large analysis integrating mCRPC genomics with histology and clinical outcomes identifies *RB1* genomic alteration as a potent predictor of poor

outcome, and is a community resource for further interrogation of clinical and molecular associations.

castration-resistant prostate cancer | integrative genomics | clinical outcomes | biomarkers

Several studies have described the genomic landscape of primary and metastatic castration-resistant prostate cancer (mCRPC), revealing distinct genomic subtypes in primary localized disease, including ETS fusion-positive and *SPOP*-mutated prostate cancer (1–5), and subsets of patients with advanced disease who harbor potentially clinically actionable alterations in their tumor or in the germline (4–6). Based on these findings, prospective trials are currently enrolling patients with defined genomic alterations, including PARP inhibitor studies for patients with alterations in *BRC42/1*, *ATM*, and other DNA repair genes (NCT02952534, NCT02987543, NCT02854436), and AKT inhibitor

Author contributions: W.A., G.H., F.D., P.S.N., M.A.R., A.M.C., and C.L.S. designed research; J.C., N.S., F.D., and M.A.R. performed research; J.M., D.B., S.M., S.C., P.R., R.B.M., H.B., E.I.H., H.I.S., P.W.K., M.-E.T., and J.S.d. contributed new reagents/analytic tools; W.A., G.H., D.P., J.A., I.C., M.C., M.B., D.R., E.M.V.A., A.S., T.F., J.M.M., B.D.R., N.D.S., L.P.K., S.T., Y.M.W., D.N.R., M.L., A.G., V.E.R., C.C.P., J.F., J.V., N.S., F.D., P.S.N., M.A.R., and A.M.C. analyzed data; and W.A., F.D., P.S.N., M.A.R., A.M.C., and C.L.S. wrote the paper.

Reviewers: S.A., BC Cancer Agency and University of British Columbia; J.T.I., Johns Hopkins Oncology Center; and N.M., University of Pennsylvania.

Conflict of interest statement: W.A. has consulted for Clovis Oncology, Janssen, ORIC Pharmaceuticals, and MORE Health; has received honorarium from Caret Healthcare; has received travel funds from Clovis Oncology, ORIC Pharmaceuticals, and GlaxoSmithKline; and has received research funding from AstraZeneca, Clovis Oncology, Zenith Epigenetics, and GlaxoSmithKline. G.H. received research funding from Janssen. J.A. is currently employed at AstraZeneca. E.M.V.A. has consulted for Tango Therapeutics, Genome Medical, Invitae, Illumina, Foresite Capital, and Dynamo; has received research support from Novartis and BMS; has equity in Tango Therapeutics, Genome Medical, Syapse, and Microsoft; has received travel reimbursement from Roche/Genentech; and is a coinventor of institutional patents filed on ERCC2 mutations and chemotherapy response, chromatin mutations and immunotherapy response, and methods for clinical interpretation. H.B. has received research funding from Millennium Pharmaceuticals, AbbVie/Stemcentrx, Astellas, Janssen, and Eli Lilly; and is a consultant/advisor for Sanofi Genzyme and Janssen. S.T. is a coinventor for University of Michigan patents on ETS fusion genes, and diagnostic field of use licensed to Hologic/Gen-Probe Inc., which has sublicensed rights to Roche/Ventana Medical Systems; is a consultant for and has received honoraria from Janssen, AbbVie, Sanofi, Almac Diagnostics, and Astellas/Medivation; has performed sponsored research from Astellas/Medivation and GenomeDx; and is an equity holder in and employee of Strata Oncology. J.M. has consulted for AstraZeneca and Janssen; and has received travel support or speaker fees from Astellas, AstraZeneca, Sanofi, and IPSEN. D.B. has received honoraria and travel support from Janssen. H.I.S. has consulted for Astellas, Ferring Pharmaceuticals, Janssen Biotech, Janssen Research and Development, Sanofi Aventis, Clovis Oncology, Merck, and WCG Oncology; is a member of the board of directors for Asterias Biotherapeutics; and has received research support from Illumina Inc., Innocrin, and Janssen. P.W.K. is a board member for Context Therapeutics LLC, has consulted for BIND Biosciences, BN Immunotherapeutics, DRGT, GE Healthcare, Janssen Pharmaceutica, Metamark, New England Research Institutes, OncoCellIMDX, Progenity, Sanofi-Aventis, Seer Biosciences Inc., Tarveda Therapeutics, and Thermo Fisher Scientific; serves on data safety monitoring boards for Genentech/Roche and Merck & Co.; and has investment interest in Context Therapeutics, DRGT, Seer Biosciences, Placon, and Tarveda Therapeutics. M.-E.T. has consulted for Janssen. J.S.d. is an advisory board member for AstraZeneca, Astellas, Bayer, Boehringer Ingelheim, Genentech/Roche, Genmab, GSK, Janssen, Merck Serono, Merck Sharp & Dohme, Menarini/Silicon Biosystems, Orion, Pfizer, Sanofi Aventis, and Taiho; and his institution has received funding or other support from AstraZeneca, Astellas, Bayer, Genentech, GSK, Janssen, Merck Serono, MSD, Menarini/Silicon Biosystems, Orion, Sanofi Aventis, and Taiho. P.S.N. is a compensated advisor to Janssen, Astellas, and Roche. M.A.R. is a coinventor on a patent for Gene Fusion Prostate Cancer (Harvard/Michigan) and EZH2 (Michigan) in the area of diagnostics and therapeutics; is a coinventor on a patent filed by Cornell University on AURKA and SPOP mutations in the field of prostate cancer diagnostics; receives royalties on licensing agreements for these inventions; and receives research support from Janssen, Eli Lilly, Millenium, and Sanofi-Aventis. C.L.S. serves on the board of directors of Novartis; is a cofounder of ORIC Pharmaceuticals and coinventor of enzalutamide and apalutamide; is a science advisor to Agios, Beigene, Blueprint, Column Group, Foghorn, Housey Pharma, Nextech, KSQ, Petra, and PMV; and is a cofounder of Seragon, purchased by Genentech/Roche in 2014.

This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Data deposition: The data reported in this paper are available in [Dataset S1](#) and at www.cbiportal.org, and have been deposited in GitHub, https://github.com/cBioPortal/datahub/tree/master/public/prad_su2c_2019.

See Commentary on page 11090.

¹W.A., J.C., and G.H. contributed equally to this work.

²Present address: Pathology Department, Institut Curie, 75005 Paris, France.

³Present address: Innovative Medicines and Early Development (IMED) Biotech Unit, AstraZeneca, CB4 0WG Cambridge, United Kingdom.

⁴Present address: Bioinformatics Unit, Hospital of Prato, 59100 Prato, Italy.

⁵Present address: Prostate Cancer Translational Research Group, Vall d'Hebron Institute of Oncology, 08035 Barcelona, Spain.

⁶M.-E.T., N.S., J.S.d., F.D., P.S.N., M.A.R., A.M.C., and C.L.S. contributed equally to this work.

⁷To whom correspondence may be addressed. Email: pnelson@fhccr.org, mark.rubin@dbmr.unibe.ch, arul@med.umich.edu, or sawyers@mskcc.org.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1902651116/-DCSupplemental.

Published online May 6, 2019.

Significance

The genomic landscape of metastatic castration-resistant prostate cancer (mCRPC) has been well-defined, but the association of genomic findings with patient clinical outcomes and with other characteristics including histology and transcriptional pathway activity remains poorly understood. Here, we describe comprehensive integrative analysis of genomic and transcriptomic profiles, histology, and clinical outcomes for 429 patients with mCRPC. Of all the molecular factors we examined, alterations in *RB1* had the strongest association with poor outcome. Our study identifies molecularly defined groups of patients who may benefit from a more aggressive treatment approach, with the genomic and outcome data made available to the research community for further interrogation.

studies for men with PI3K pathway alterations (NCT02525068, NCT03310541).

In addition, various genomic and histologic features of prostate cancer have been described as conferring a worse prognosis. Among these are the presence of neuroendocrine or small-cell characteristics in tumors, sometimes referred to as aggressive variant prostate cancer or neuroendocrine prostate cancer (7–9), the detection of androgen receptor (AR) splice variant 7 in circulating tumor cells (10, 11), and the presence of genomic alterations in TP53, RB1, DNA repair genes, AR, and PI3K pathway genes in circulating tumor DNA (12, 13). However, studies that comprehensively examine all of these characteristics—histology, genomics, and transcriptomics—and their association with outcomes in mCRPC are lacking.

Here, we expand a foundational genomic resource of mCRPC tumors (5) from 150 to 429 patients (444 tumors), and integrate the analysis of whole-exome sequencing, gene expression, and

histopathology with clinical outcomes, including survival and time on treatment with the next-generation androgen signaling inhibitors (ARSIs) enzalutamide and abiraterone acetate to identify the most important prognostic markers in mCRPC within a single large multiinstitutional genomic dataset, with tumor- and patient-level data made available for additional correlative analyses.

Results

Clinical and Histopathologic Parameters. A total of 429 patients were enrolled at one of seven international consortium centers, all of whom underwent biopsy for the collection of mCRPC tissue as well as collection of blood for matched normal DNA extraction. Whole-exome sequencing was successfully performed on 444 tumors (some patients underwent multiple biopsies), and RNA sequencing (RNA-seq) was successfully performed on a subset of these (332 tumors from 323 patients). Of the 444 biopsies, 37% were lymph node, 36% were bone, and 14% were liver (Fig. 1A). Samples underwent central histopathologic review (Fig. 1B), revealing neuroendocrine (NE) or small-cell features in 11.2% (41 of 366) of evaluable cases, including from patients who were enrolled on a trial of the Aurora kinase A inhibitor alisertib (14). Median age at diagnosis with prostate cancer was 61 y, median age at biopsy of the profiled sample was 67 y, and median overall survival from the time of biopsy was 16 mo (SI Appendix, Table S1). Samples were balanced for exposure status to ARSIs (47% ARSI naive, 46% previously exposed) (Fig. 1C). Sixty-three percent of samples were acquired before exposure to a taxane. Patients who were naive to both ARSI and taxane at the time of biopsy had the longest median overall survival from the date of biopsy, whereas patients previously treated with both an ARSI and a taxane had the shortest (Fig. 1D), consistent with their more advanced disease state at the time of tissue acquisition. Of note, time

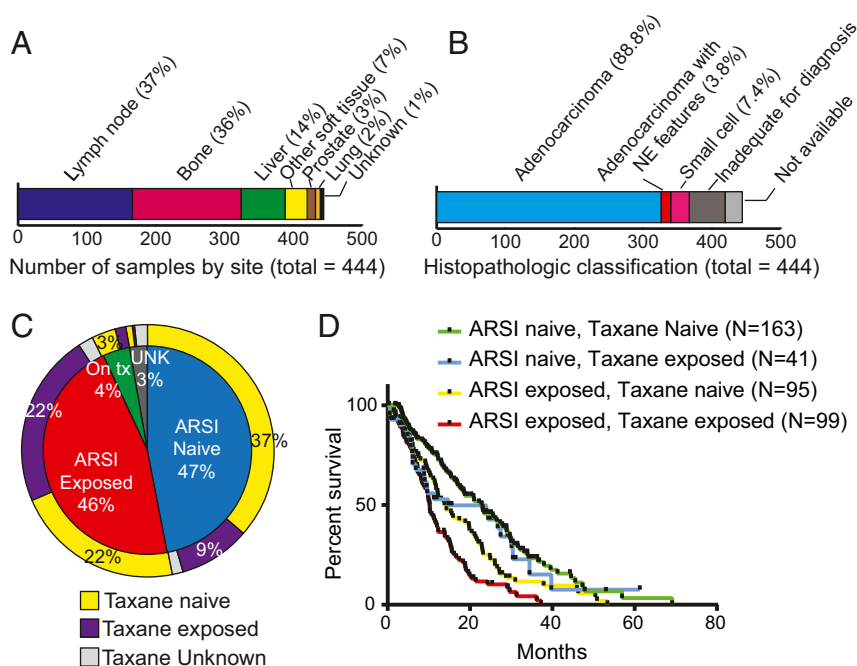


Fig. 1. Overview of sample and patient characteristics for 444 tumors from 429 patients with mCRPC. (A) Site of mCRPC tumors profiled. (B) Histopathologic classification of profiled tumors. Tumors were classified by central review as adenocarcinoma, pure small-cell/neuroendocrine cancer, adenocarcinoma with neuroendocrine features (also included mixed acinar/neuroendocrine carcinoma), or could not be classified due to scant material or no tumor visible on the slides that were available for review despite successful sequencing. (C) Patient exposure status to next-generation AR signaling inhibitors (abiraterone acetate, enzalutamide, or ARN509) and to taxanes at the time of biopsy for the 444 profiled tumors. (D) Overall survival (OS) from the date of biopsy of the profiled tumor. OS was longer for tumors from ARSI- and taxane-naïve patients compared with patients who had received an ARSI before the biopsy ($P < 0.01$, log-rank test). Survival was shortest when the patient had received both an ARSI and taxane chemotherapy at the time of biopsy.

on treatment with a first-line ARSI was highly associated with overall survival from the start of first-line therapy (Kendall's $\tau = 0.65$).

Landscape of Genomic Alterations. The frequency of genomic alterations was similar to that reported in prior cohorts (4, 5, 15), with AR, ETS family transcription factors including ERG and ETV1, TP53, and PTEN and RB1 emerging as the most commonly altered genes (Fig. 2A). Likewise, alterations in biological pathways (SI Appendix, Table S2) were also consistent with prior reports, with a significant (>20%) subset of patients harboring at least one alteration in a PI3K, cell-cycle, epigenetic, or DNA repair pathway gene. Single-nucleotide variants (SNVs) in the most frequently altered genes were found to be likely oncogenic (16) in the majority of cases (Fig. 2B), with a high fraction of oncogenic mutations in AR, TP53, PIK3CA, BRCA2, PTEN, APC, and CDK12. Mutations in ATM were predicted to be likely oncogenic in nearly 60% of cases, with the rest being missense mutations of unknown significance.

The large size of the dataset allowed for a comprehensive search for genomic alterations that are co-occurring or mutually exclusive (Fig. 2C). As expected, we found mutual exclusivity between alterations in genes of the ETS family (e.g., ERG and ETV1), and between alterations in ERG and SPOP or FOXA1, which represent distinct genomic subsets of prostate cancer (1, 3). Alterations in ERG and PTEN were co-occurring, in line with their synergistic role in promoting oncogenesis in mouse models of prostate cancer (17). We also confirmed co-occurrence between alterations in TP53 and RB1, known to occur at high frequency in neuroendocrine cancers (7, 9), and to confer aggressive behavior in prostate cancer models (18, 19). Interestingly, RB1 alteration had a tendency toward mutual exclusivity with alterations in AR. CHD1 alterations also tended to co-occur with SPOP mutations (3). We found strong co-occurrence of loss-of-function alterations in CDK12, a gene implicated in the control of genomic stability (20) whose inactivation in prostate cancer is associated with focal tandem duplications (21–23), with amplification of the cell-cycle genes CCND1 and CDK4, raising the possibility of vulnerability to CDK4/6 inhibitors for CDK12-mutated tumors. Conversely, while genomic alterations in RB1 and BRCA2, both

located on chromosome 13q, 16 Mb apart, had a tendency toward co-occurrence, this association did not reach statistical significance.

Association of Genomic Alterations with Clinical Outcomes. A key novelty of this dataset is the opportunity to correlate contemporaneously obtained comprehensive genomic profiles with clinical outcome. We focused our analysis on 18 of the most commonly altered genes and pathways. For clinical outcome, we restricted the analysis to those patients who were taxane-naïve and initiating therapy with a first-line ARSI for mCRPC ($n = 128$).

We examined the association of genomic alterations with overall survival from the start of a first-line ARSI ($n = 128$) and time on treatment with a first-line ARSI (subset of $n = 108$ patients who received the ARSI without another concurrent therapy) in univariate analysis (Table 1). In this analysis, genomic alterations in the PI3K pathway and its component genes (Fig. 3A) were not significantly associated with either time on therapy with ARSIs or with overall survival (Fig. 3B and C), unlike prior cell-free DNA (cfDNA)-based analysis (13). Furthermore, we explored the association of genomic alterations in the DNA repair genes BRCA2, BRCA1, and ATM (Fig. 3D) with clinical outcomes, given prior conflicting reports of prognostic significance of these alterations (13, 24–26). We again found no association between alterations in these genes and time on treatment or overall survival (Fig. 3E and Table 1). Notably, we found an association between SPOP mutation and longer time on treatment with a first-line ARSI (SI Appendix, Fig. S1), consistent with prior data showing enrichment of SPOP mutations in earlier disease relative to mCRPC (1, 4) and favorable prognosis of SPOP-mutated tumors (27, 28), though this did not translate into a survival benefit. Alterations in AR and TP53 were associated with a shorter time on an ARSI (Figs. 4D and 5C), though there was no association with overall survival (Table 1). Overall, genomic alterations in RB1 showed the strongest discrimination for a shorter time on an ARSI and survival, with concordance probability estimates (CPEs) of 0.82 and 0.77, respectively (Fig. 5A and B and Table 1). Aneuploid chromosomal

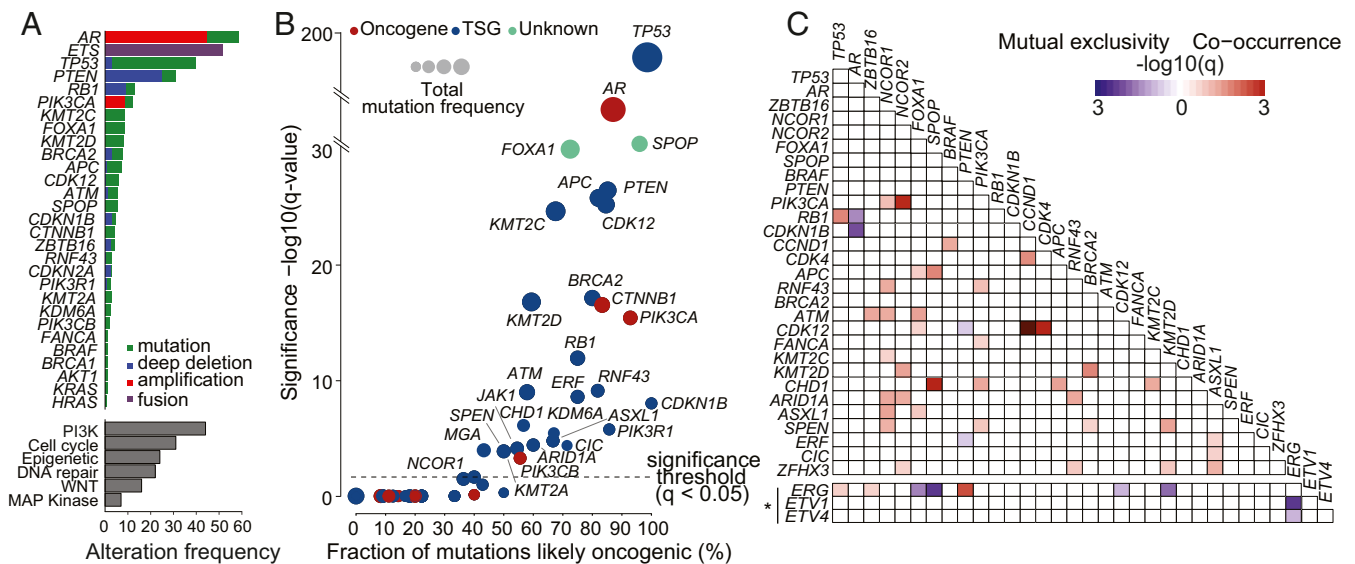


Fig. 2. Landscape of genomic alterations in 444 tumors from 429 patients with mCRPC. (A) Genomic alterations of potential biologic relevance by frequency. (A, Top) Frequency of alteration by gene. Frequency of ETS gene alterations applies to the subset of 323 patients who underwent tumor RNA sequencing, where fusion status could be determined. (A, Bottom) Frequency of alteration by pathway (SI Appendix, Table S2). (B) Fraction of SNVs considered to be likely oncogenic for genes harboring mutations. Genes colored in red are putative oncogenes, genes colored in blue are putative tumor suppressor genes (TSGs), and genes colored green are currently unknown. (C) Co-occurrence or mutual exclusivity between the most commonly altered genes. The significance of the relationship is represented by gradient. Relationships with P value < 0.05 with multiple hypothesis correction are shown. Associations involving ETS genes (*) apply only to cases where RNA-sequencing data are available.

Table 1. Association of common genomic alteration with overall survival and time on treatment with first-line ARSI

Gene/pathway alteration	Univariate <i>P</i> value for survival from first-line ARSI (CPE, <i>n</i> = 128 or as indicated)	Univariate <i>P</i> value for time on treatment with first-line ARSI (CPE, <i>n</i> = 108 or as indicated)
RB1	0.002 (CPE 0.768)	<0.001 (CPE 0.818)
TP53	0.072 (CPE 0.605)	0.046 (CPE 0.609)
WNT pathway	0.115	0.153
ETS fusion	0.159	0.206
APC	0.255	0.167
CTNNB1	0.274	0.448
ATM	0.331	0.850
BRCA2	0.327	0.418
BRCA2/BRCA1/ATM	0.495	0.611
AKT1	0.558	0.053
RNF43	0.614	0.844
AR	0.658	0.005 (CPE 0.651)
PTEN	0.676	0.412
PI3K pathway	0.699	0.138
PIK3CA	0.716	0.165
PIK3R1	0.752	0.892
PIK3CB	0.799	0.277
BRCA1	0.809	0.998
NEPC score	0.218 (<i>n</i> = 99)	0.930 (<i>n</i> = 80)
AR signaling score	0.847 (<i>n</i> = 99)	0.847 (<i>n</i> = 80)
RB1 loss score	<0.001 (<i>n</i> = 99)	0.014 (<i>n</i> = 80)
CCP score	0.002 (<i>n</i> = 99)	0.045 (<i>n</i> = 80)
AR-V7 SRPM	0.524 (<i>n</i> = 75)	0.329 (<i>n</i> = 56)
AR-V7/ARpromoter1-2	0.475 (<i>n</i> = 75)	0.378 (<i>n</i> = 56)
AR-V3 SRPM	0.444 (<i>n</i> = 75)	0.077 (<i>n</i> = 56)

Univariate log-rank analysis for association of common genomic alterations with survival from the start of a first-line ARSI for mCRPC (*n* = 128), and with time on treatment with a first-line ARSI for mCRPC (*n* = 108 patients who received a first-line ARSI as monotherapy). Where indicated, analysis was limited to a subset of patients who had RNA-sequencing data either from polyA libraries or both polyA and capture libraries. *P* < 0.05 are highlighted in bold.

status was associated with worse overall survival and time on treatment compared with diploid status (*SI Appendix, Fig. S2*).

Androgen Receptor Alterations. We confirmed a high frequency of genomic alterations in *AR*, namely amplifications and mutations, consistent with prior reports (5, 12) (*Fig. 24*). Using RNA-seq data, we also identified splice variants in *AR*, most commonly *AR* splice variant 7 (*AR-V7*) and variant 3 (*AR-V3*), both a product of splicing with cryptic exons, similar to our prior report (5) (*Fig. 44*). Genomic alterations (amplification and mutation) were associated with an increased *AR* expression score, consistent with increased *AR* output (*Fig. 4B*). Furthermore, genomic alterations in *AR* were detected at higher frequency postexposure to the ARSIs compared with ARSI-naïve tumors (*Fig. 4C*), suggesting an association with resistance to these next-generation *AR*-targeting agents. Consistent with this, we found an association of *AR* genomic alterations with a shorter time on treatment with a first-line ARSI (*Fig. 4D*) but not with overall survival from the start of a first-line ARSI (*Table 1*). Given prior data showing a strong association between *AR-V7* expression in circulating tumor cells (CTCs) and clinical outcomes (10, 11), we examined the association of *AR-V7* expression and outcomes in our tumor dataset. *AR-V7* levels were increased in tumors exposed to taxanes and to ARSI therapy (*SI Appendix, Fig. S3*). However, there was no association between *AR-V7* expression in tumors with either time on a first-line ARSI or overall survival (*Fig. 4E*

and *Table 1*). This was true even when deriving an *AR-V7* expression cutpoint that produced the maximum log-rank test (*AR-V7* cutpoint for survival 1.92, *P* = 0.62). Similar results were observed for time on treatment, *AR-V7/AR* promoter 1:2 ratio, and *AR-V3* (*Table 1*).

Integrative Analysis of Histopathology, Genomics, and Expression.

There is growing recognition that a subset of CRPC patients have a more fulminant clinical course—typically characterized by rapidly progressive visceral metastasis (versus bone), relatively low serum prostate-specific antigen (PSA) levels, and variable expression of neuroendocrine markers such as synaptophysin or chromogranin (29). However, there is a lack of consensus on how to precisely define this clinical state, particularly since the frequency of histologically defined neuroendocrine prostate cancer varies widely in different cohorts. Furthermore, there is increasing recognition that patients can develop *AR*-negative (*PSA*-low) disease that is histologically negative for neuroendocrine marker expression (30). RNA-based expression signatures have been proposed as a potential alternative diagnostic strategy to define this distinct clinical state (8, 30, 31). The availability of matched histology, whole-exome, and RNA-seq data from this cohort provides an opportunity to explore this question through an unbiased integrative approach.

As previously noted, 11.2% of tumors in our dataset had evidence of NE features on histopathologic review (*Fig. 1D*). Among patients who received treatment with an ARSI during their disease course, tumors with histopathologic NE features were enriched postexposure to an ARSI (10.5%) compared with ARSI-naïve tumors (2.3%) (*Fig. 5D*). Transcript-based NE score was not significantly different between the two groups, but a subset of tumors in the post-ARSI setting displayed a higher NE expression score (*Fig. 5E*). Of note, unsupervised gene expression clustering identified a distinct cluster of tumors with higher NE expression score, in line with prior reports and independent of site of metastasis (*SI Appendix, Fig. S4*) (31).

We performed an integrative analysis incorporating histology, expression-based *AR* signaling and NE scores, and *RB1/TP53* genomic status (7, 32) in all tumors where RNA-sequencing data were available (*Fig. 5F*). As expected, there was an inverse correlation between cases with a high *AR* signaling score and cases with a high NE score. We identified three groups based on the expression signatures. The first and largest group, characterized by high *AR* signaling and low NE score, consisted predominantly (86%) of adenocarcinomas without NE histologic features (*Fig. 5G*). The second group, demonstrating intermediate NE and *AR* scores (*n* = 17), included cases classified histologically as adenocarcinoma (59%) and cases that were “inadequate for diagnosis” (41%). The third group, demonstrating high NE score and low *AR* signaling, consisted predominantly (74%) of tumors harboring histologic NE features (*Fig. 5H*).

Although there was an association between NE expression score, histologic NE features, and *RB1/TP53* loss (*SI Appendix, Fig. S5*), concordance between these characteristics was imperfect (*Fig. 5F*). We posited that some cases with discrepancy between pathology and transcriptomic classification may demonstrate a distinct morphology. To address this, a second consensus review of those cases was performed by three study pathologists. We confirmed that all discrepant cases from group 3 (high NE score but showing adenocarcinoma histology) and all cases from group 2 (intermediate *AR* and NE scores) displayed adenocarcinoma features. However, we noted distinct nuclear features in about half of these cases, including various degrees of nuclear pleomorphism, irregular nuclear membrane contours, and/or high mitotic activity (*Fig. 5I* and *J*). Of note, within group 2, 10 cases (3% of total; *Fig. 5F*, box) had low *AR* and low NEPC expression scores. Of these, all histologically evaluable cases

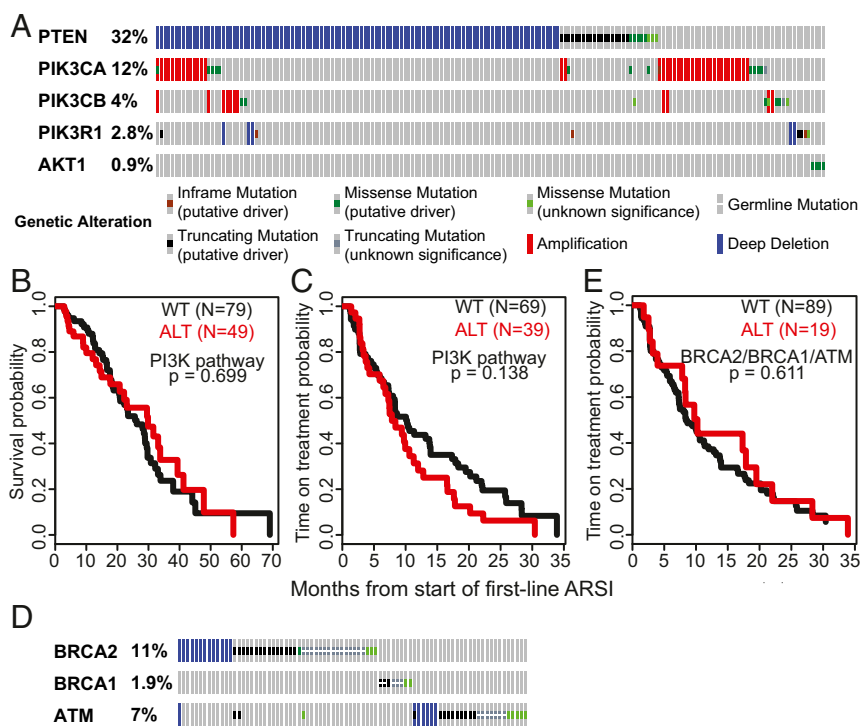


Fig. 3. Alteration in PI3K and homologous recombination repair genes and association with clinical outcomes. (A) Oncoprint of genomic alterations in PI3K pathway genes. (B and C) Kaplan–Meier analysis showing overall survival (B) and time on treatment with a first-line ARSI (C) in PI3K pathway altered (red) versus unaltered (black) tumors. (D) Oncoprint of genomic alterations in BRCA2, BRCA1, and ATM. (E) Kaplan–Meier analysis showing time on treatment with a first-line ARSI in BRCA2/1/ATM–altered (homozygous deletion or somatic or pathogenic germline mutation) (red) versus unaltered (black) tumors.

showed adenocarcinoma histology, and the majority harbored atypical nuclear features.

Given incomplete concordance between these histopathologic, genomic, and expression characteristics, we asked which of these features was most associated with clinical outcome. While there were insufficient histopathologic NE cases to make this determination, NE transcriptional score was not significantly associated with time on a first-line ARSI or overall survival (Table 1 and *SI Appendix*, Fig. S6). Of all molecular characteristics examined in a multivariate analysis, *RBI* alteration emerged as the only variable strongly associated with survival (relative risk 3.31) and with time on treatment with a first-line ARSI (relative risk 6.56) (Table 2). Consistent with this, expression scores for *RBI* loss and cell-cycle progression (CCP) were both associated with worse survival and time on treatment with a first-line ARSI (Table 1 and *SI Appendix*, Fig. S7).

Discussion

The landscape of genomic alterations in mCRPC has been established, with a subset of patients harboring potentially actionable alterations that are currently being explored in targeted prospective clinical trials. However, the majority of genomic alterations in prostate cancer do not yet have clear clinical applicability. Some studies have associated specific genomic or molecular features with clinical outcomes, though this has generally been performed in cell-free DNA or circulating tumor cells. Here, we present an integrative analysis of genomic alterations with expression and histologic assessment in tumors from patients with mCRPC, representing the clinical spectrum of advanced disease, with tissue collected pre- and posttreatment with ARSIs and taxanes.

Importantly, we find that *RBI* loss is the only genomic factor that is significantly associated with both survival and time on ARSI therapy in mCRPC. It is worth noting that the association

was strong despite likely underestimation of *RBI* loss by examining genomic homozygous loss alone, as *RBI* loss has also been shown to occur epigenetically, through structural genomic events like tandem duplication of partial exons, and focally, by immunohistochemistry (33). *TP53* and *AR* alteration were also associated with shorter duration of ARSI therapy, though the association was not as strong as for *RBI* loss and did not extend to survival. The *AR* findings suggest that AR targeting with the ARSIs abiraterone acetate and enzalutamide may be incomplete, and that further targeting of the protein may be clinically beneficial in patients who develop resistance to these agents.

We found no association between alterations in PI3K pathway genes or alterations in the DNA damage repair genes *BRCA2*, *BRCA1*, and *ATM* with overall survival and time on treatment with an ARSI. This is in contrast to a prior study that found an association between alterations in *BRCA2*, *BRCA1*, and *ATM* detected in cfDNA with both response to ARSIs and with survival (13). The differing conclusions may be related to differences between tumor and cfDNA profiling, potential difficulty in detecting homozygous loss in cfDNA relative to a tumor, and a possible bias introduced by requiring detectable cfDNA, though the authors accounted for cfDNA detection in a multivariate analysis. Furthermore, data for the prognostic role for *BRCA2*, *BRCA1*, and *ATM* alterations were previously conflicting, with another study showing better prognosis for tumors with alterations in these genes (25) and a third study showing no impact for the presence of germline DNA repair gene alterations on outcomes in mCRPC (26). Notably, we also found no association between AR-V7 in tumors and clinical outcomes, in contrast to prior CTC-based studies (10, 11). This finding requires further exploration of the concordance between AR-V7 expression in tissue versus CTCs, and the biological significance of AR-V7 detection in these contexts. Our findings suggest that AR-V7 RNA detection in tumors may have limited clinical utility. Overall, while profiling of cfDNA and CTCs offers advantages over tumor

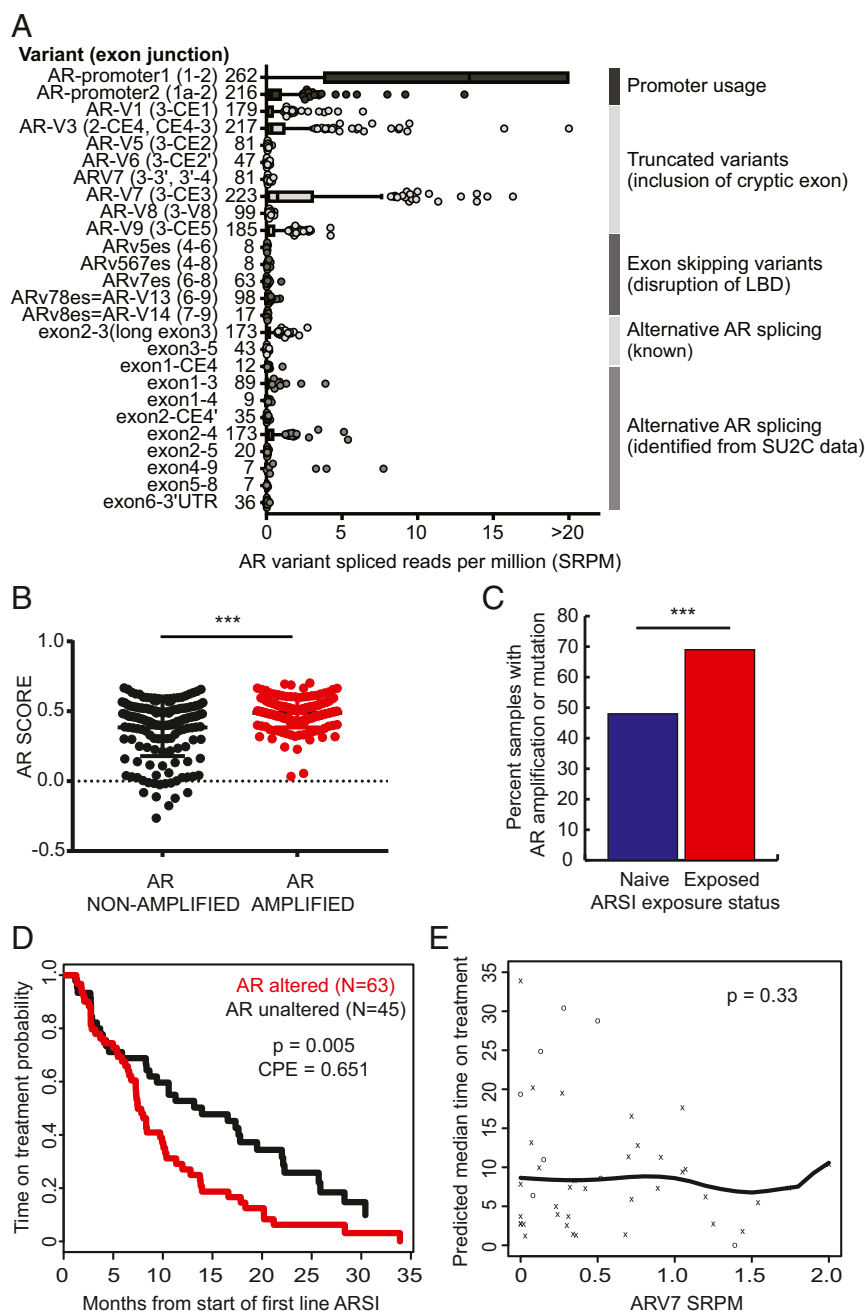


Fig. 4. Androgen receptor alterations and outcome. (A) AR splice variant landscape. LBD, ligand binding domain. (B) AR pathway expression score in AR-amplified ($n = 168$) versus nonamplified ($n = 159$) tumors. *** $P < 0.001$. (C) AR amplification frequency in ARSI-naive versus exposed tumors. (D) Kaplan–Meier analysis showing time on treatment with a first-line ARSI in AR-amplified versus nonamplified tumors. (E) Association between ARV7 expression and time on treatment with a first-line ARSI. o, censored event; x, off-treatment event.

profiling, including ease of access and the ability to capture genomic heterogeneity (34, 35), blood-based profiling may be limited by lower sensitivity of detection and may reflect the fact that CTCs and cfDNA are generally detectable in patients with more advanced or aggressive disease. We recognize that association with clinical outcomes was performed in a subset of 128 patients who were initiating first-line therapy for mCRPC with an ARSI, and that a larger cohort could reveal additional associations. Nonetheless, the size of our cohort meets or exceeds previously reported datasets, and it is unclear if smaller differences in outcomes would be considered clinically meaningful.

We found that neuroendocrine histology, generally viewed as conferring more aggressive clinical behavior, is more frequent postexposure to ARSIs (10.5%), though at a lower frequency than recently reported in another study (17%) (8), despite inclusion in our cohort of patients with neuroendocrine features from a clinical trial of alisertib (14). Prostate cancers with histologic neuroendocrine differentiation (36) typically have alterations in *TP53* and/or *RBI*, high neuroendocrine expression score, and low AR signaling score, though concordance is not complete. This is not surprising, given the complexity of defining histologic neuroendocrine differentiation, which relies on the identification of a variety of characteristic features (29). Some

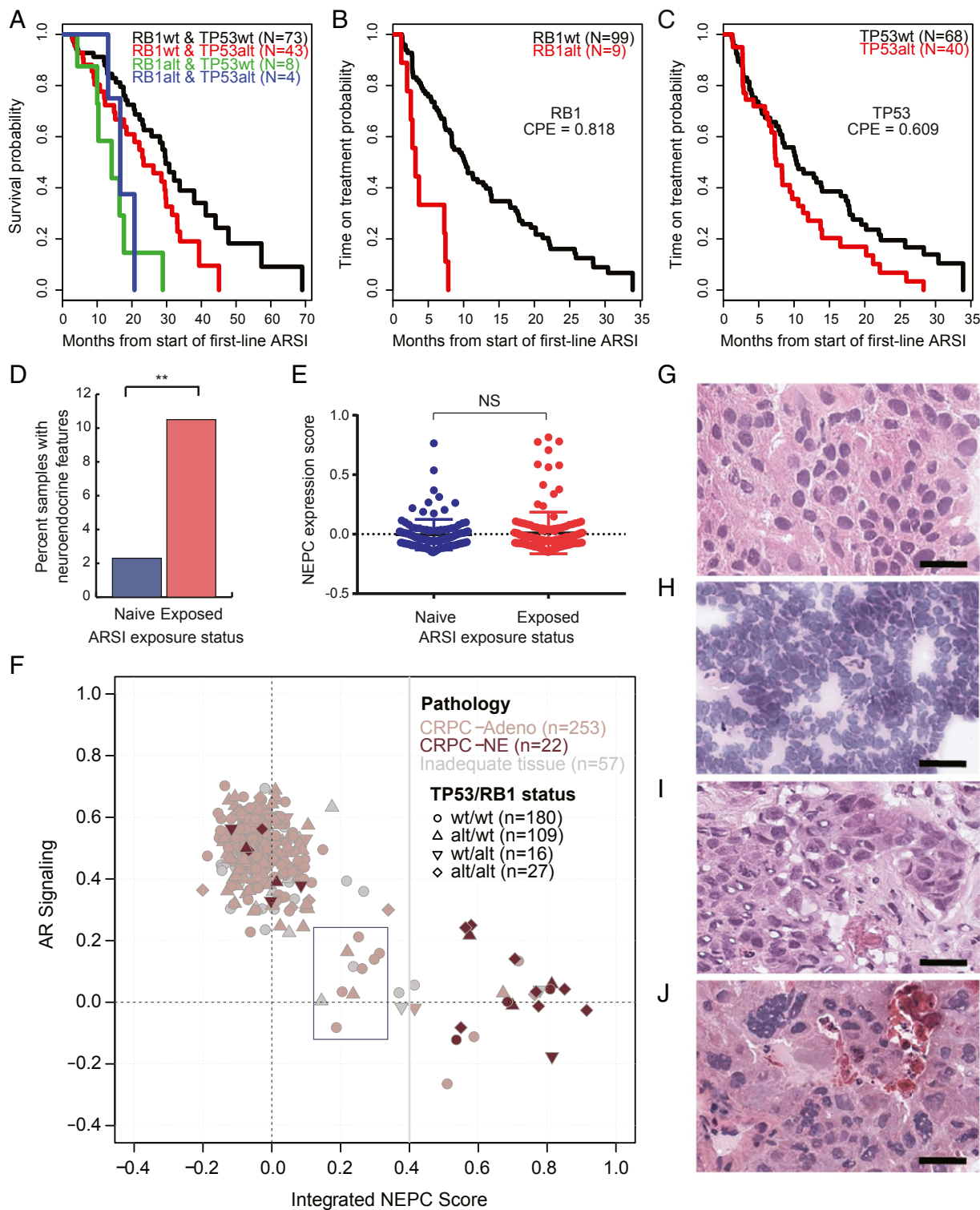


Fig. 5. Integrative analysis incorporating histopathology, transcript-based assessment of AR signaling and NEPC score, TP53 and RB1 genomic status, and clinical outcomes. (A) Kaplan–Meier analysis showing overall survival from the start of a first-line ARSI versus genomic status for TP53 and RB1 in $n = 128$ patients who received a first-line ARSI and underwent tissue profiling at baseline (before or within 90 d of therapy start). (B and C) Kaplan–Meier analysis showing time on treatment with a first-line ARSI by genomic status for RB1 and TP53. P values were generated from the log-rank statistic. (D) Frequency of histopathologic neuroendocrine features in pre- versus post-ARSI samples, among patients who received an ARSI at some point during their treatment history. Patients who were not reported to have received an ARSI at any point were excluded. $**P < 0.01$. (E) NEPC expression score in pre- ($n = 118$) versus post- ($n = 152$) ARSI samples, as in D. NS, not significant. (F) AR and NEPC expression scores, histopathology (CRPC-Adeno, no NE features; CRPC-NE, histopathologic NE features) and TP53/RB1 genomic status (circle, wild type for both; diamond, both altered) for the 332 tumors with RNA-sequencing data. Ten cases (3%, blue box) had low AR and low NEPC expression scores. (G–J) Representative cases of CRPC-Adeno (G), CRPC-NE, small-cell type (H), CRPC-Adeno showing intermediate transcriptomic scores (I), and CRPC-Adeno showing a high NEPC score/low AR signaling score (J). Tumors represented in I and J were noted to have distinct nuclear features, including various degrees of nuclear pleomorphism, irregular nuclear membrane contours, and/or high mitotic activity. (Scale bars, 25 μm .)

Table 2. Multivariate analysis evaluating the association of common genomic alterations with overall survival and time on treatment with first-line ARSI

Clinical outcome	Gene alteration(s)	Multivariate relative risk (95% CI for RR)
Overall survival from start of first-line ARSI	RB1–	1
	RB1+	3.31 (1.64, 6.67)
Time on treatment with first-line ARSI	RB1– and AR–	1
	RB1– and AR+	1.86 (1.18, 2.95)
	RB1+	6.56 (2.94, 14.62)

Common genomic alterations listed in Table 1 were included in this analysis. Only significant associations are shown. CI, confidence interval; RR, relative risk.

tumors with discordant histologic and molecular classification may be in transition from adenocarcinoma to NE differentiation, and displayed distinct nuclear features that may suggest such a transition. In such cases, paraffin-embedded tissue may aid in the classification by allowing for further examination of histomorphologic features and immunohistochemical staining of markers.

The size of our whole-exome sequencing dataset allows for genomic association analysis that was not previously possible. Through this analysis, we identified co-occurrence of alterations in *CDK12*, recently shown to confer immunogenic potential, and alterations in cell-cycle genes *CDK4* and *CCND1* (23), pointing toward a possible role for combination immune checkpoint blockade and CDK4/6 inhibition in clinical trials. Further laboratory studies will be needed to explore this and other potential biological interactions identified through genomic analysis.

In summary, we present an integrative analysis of genomic alterations, gene expression, histopathology, and clinical outcomes in the largest single mCRPC dataset to date, with the data made available to the research community for interrogation of genomic features in relation to outcomes. We find that *RB1* loss is the molecular factor most strongly associated with poor clinical outcomes in a contemporary cohort, highlighting the need for further investigation into mechanisms of resistance to AR therapies induced by loss of RB, and potential therapeutic strategies targeting this mechanism.

Methods

Patients and Samples. Subjects with mCRPC who were receiving standard-of-care therapy or treatment in a clinical trial [including trials combining AR therapies with other agents, a trial of the PARP inhibitor olaparib (37), and a trial of the Aurora kinase A inhibitor alisertib in patients with neuroendocrine features (14)] and who had disease amenable to biopsy under radiographic guidance were considered for inclusion at one of seven SU2C-PCF (Stand Up to Cancer/Prostate Cancer Foundation) International Prostate Cancer Dream Team consortium sites (Dana-Farber Cancer Institute, Karmanos Cancer Institute, Memorial Sloan Kettering Cancer Center, Royal Marsden, University of Michigan, University of Washington, and Weill Cornell Medicine) (5). All subjects included in this study provided written consent for research use of tumor tissue with institutional review board approvals or appropriate waivers (Office of Human Research Studies at the Dana-Farber Cancer Institute, Wayne State University Institutional Review Board, Memorial Sloan Kettering Cancer Center Institutional Review Board/Privacy Board, Royal Marsden Ethics Committee, University of Michigan Medical School Institutional Review Board, University of Washington Institutional Review Board, and Weill Cornell Medicine Institutional Review Board). Clinical data, including treatment history, duration of therapy, and survival, were collected using a web-based electronic data capture. All samples and clinical data were deidentified.

Histopathology. Pathology for all available cases was reviewed centrally by three board-certified pathologists with expertise in prostate cancer pathology, who were blinded to clinical and genomic data. Review was conducted on H&E-stained frozen sections, allowing for review of the exact material that was used for nucleic acid extraction. Each slide was assessed for the ability to make

a diagnosis based on the quality of the sample and presence of tumor cells; cases in which a specific pathology diagnosis could not be called were classified as inadequate for diagnosis. The remaining cases ($n = 366$) were classified according to a previously published system (36), by consensus in the event that all three pathologists did not agree. There was no additional material for paraffin embedding, immunohistochemistry, or other confirmatory studies.

Sequencing and Analysis. Flash-frozen needle biopsies and matched normal samples underwent nucleic acid extraction as previously described (5). Extracted DNA underwent whole-exome library construction and somatic mutation analysis as previously described. BAM files were aligned to the hg19 human genome build. Copy-number aberrations were quantified and reported for each gene as previously described (38, 39). Amplifications and homozygous deletions for a set of 20 genes previously implicated in prostate cancer (*SI Appendix, Table S3*) underwent further confirmatory review of segmentation files. Annotation of known or likely oncogenic SNVs was performed using the OncoKB platform (16).

Transcriptome libraries were prepared as previously described (5), using polyA+ RNA isolation, or captured using Agilent SureSelect Human All Exon V4 reagents, or in some cases using both polyA and capture methods. Library quality assessment and sequencing were performed as previously described. Paired-end transcriptome-sequencing reads were aligned to the human reference genome (GRCh38) using STAR (40). Gene expression as fragments per kilobase of exon per million fragments mapped (FPKMs) was determined using featureCounts against protein-coding genes from the Gencode v26 reference. Fusions in ETS genes (*ERG*, *ETV1*, *ETV4*, *ETV5*, *FLI1*) and *RAF1/BRAF* were detected using CODAC (41) and assessed manually in all cases where RNA-sequencing data were available. In addition, the presence of AR splice variants was quantified as the number of reads across specific splice junctions in splice reads per million (SRPMs) and as the ratio of reads across a specific splice junction to the sum of AR promoter 1 and promoter 2 reads (a surrogate of total AR expression), separately for polyA and capture libraries.

NEPC and AR signaling scores were computed by the Pearson's correlation coefficient between the log₂-transformed FPKM values of each score's gene list and a reference gene expression vector, as previously described (7, 32). CCP and RB loss scores were computed by the average (i.e., mean) Z score-transformed expression levels across each score's gene list, as previously described (42, 43). A high correlation ($R \geq 0.95$, $P < 0.001$, Pearson's correlation test) was noted between scores derived from polyA versus capture RNA-sequencing libraries (*SI Appendix, Fig. S8*), allowing for joint analysis of samples sequenced with either library construction method.

All data from SNV, copy-number, and expression analysis as well as clinical characteristics and outcomes measures (*Dataset S1*) have been made available in cBioPortal (44) (www.cbioportal.org), and have been deposited in GitHub, https://github.com/cBioPortal/datahub/tree/master/public/prad_su2c_2019.

Statistics and Genomic Association with Outcomes. Fisher's exact tests and unpaired *t* tests were performed in R (3.5.0) and GraphPad Prism software as indicated. For the analysis shown in Fig. 2B, enrichment analysis using a binomial distribution test was performed as previously described (45) to identify genes that had a significant fraction of known or likely oncogenic alterations (as defined by OncoKB) among all identified SNVs. Multiple hypothesis test correction was applied using the Benjamini-Hochberg method, with *q* values of < 0.05 considered significant for enrichment of oncogenic mutations among all SNVs identified for a gene. Kaplan-Meier analysis was performed from time of biopsy to death for all samples. Overall survival analysis was performed for the $n = 128$ subjects who received an ARSI (abiraterone, enzalutamide, or apalutamide) in the first-line setting before a taxane, either alone or in combination with another agent in a clinical trial, and where the profiled tissue was obtained before the start of therapy or within 90 d after starting first-line therapy. Time on treatment analysis was evaluated for a subset of $n = 108$ patients (of the 128 above), who received an ARSI in the first-line setting without another agent, so as not to confound the interpretation of response to the ARSI. *P* values for individual (univariate) association tests between genomic status and survival/time on treatment were generated from the log-rank statistic. In cases where a data-driven threshold value was used to determine the genomic status, the *P* value was computed from the maximum log-rank statistic. When a genomic class contained a small number of events, the *P* value was produced using a permutation log-rank test. A concordance probability estimate provided a metric to assess the level of separation between the Kaplan-Meier curves and is reported in relevant cases. Multivariate analyses were performed for the association of common genomic characteristics shown in Table 1 with overall survival and time on a first-line ARSI, with relative risk reported based on the Cox proportional hazards model. Kendall's tau, derived from

the Clayton copula, was used to evaluate the level of association between the time on therapy end point and overall survival.

ACKNOWLEDGMENTS. We thank the affected individuals who participated in this study. We thank the following for their support in the conduct of the study: Jyoti Athanikar, Karen Giles, Ritika Kundra, Brigit McLaughlin, and Liangzuo (Tony) Ren. This work was supported by the Prostate Cancer Foundation and by a Stand Up to Cancer Prostate Cancer Dream Team research grant. Stand Up to Cancer is a program of the Entertainment Industry Foundation administered by the American Association for Cancer Research Award SU2C-AACR-DT0712. Additional support was provided by Prostate Cancer Foundation Young Investigator Awards (to W.A., J.M., and

N.S.); Department of Defense Prostate Cancer Research Program Awards W81XWH-17-1-0124 (to W.A.), W81XWH-09-1-0147 (PCCTC), PC170510 and PC170503P2 (to C.C.P.), and W81XWH-17-1-0380 (to N.D.S.); NCI Prostate Cancer SPORE Awards P50CA186786 (University of Michigan), P50CA092629 (MSKCC), P50CA090381 (DFCI), P50CA211024 (WCMC), and P50CA097186 (University of Washington); NCI Cancer Center Award P30 CA008748 (MSKCC); NIH Award R01CA125612 (to M.A.R.); European Research Council Consolidator Grant 648670 (to F.D.); and the Nuovo-Soldati Foundation (J.C.). The RM and ICR team is supported by the Movember Foundation and Prostate Cancer UK, PCF, the ECMC network from Cancer Research UK, the Department of Health in the UK, and BRC grant funding.

^aDepartment of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY 10065; ^bDepartment of Pathology, Weill Medical College of Cornell University, New York, NY 10021; ^cDepartment for Biomedical Research, University of Bern, 3008 Bern, Switzerland; ^dDepartment of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY 10065; ^eDepartment of Cellular, Computational and Integrative Biology (CIBIO), University of Trento, 38123 Trento, Italy; ^fHuman Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY 10065; ^gFred Hutchinson Cancer Center, University of Washington, Seattle, WA 98109; ^hDepartment of Pathology, University of Michigan, Ann Arbor, MI 48109; ⁱDepartment of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02215; ^jBroad Institute, Cambridge, MA 02142; ^kInstitute of Cancer Research, London SW7 3RP, United Kingdom; ^lThe Royal Marsden National Health Service Foundation Trust, London SM2 5NG, United Kingdom; ^mDepartment of Oncologic Pathology, Dana-Farber Cancer Institute, Boston, MA 02215; ⁿDepartment of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY 10065; ^oProstate Cancer Clinical Trials Consortium, New York, NY 10065; ^pDepartment of Medicine, Weill Medical College of Cornell University, New York, NY 10021; ^qDepartment of Pathology, Wayne State University School of Medicine, Detroit, MI 48201; ^rDepartment of Oncology, Barbara Ann Karmanos Cancer Institute, Detroit, MI 48201; ^sHoward Hughes Medical Institute, University of Michigan, Ann Arbor, MI 48109; and ^tHoward Hughes Medical Institute, Memorial Sloan Kettering Cancer Center, New York, NY 10065

1. Cancer Genome Atlas Research Network (2015) The molecular taxonomy of primary prostate cancer. *Cell* 163:1011–1025.
2. Taylor BS, et al. (2010) Integrative genomic profiling of human prostate cancer. *Cancer Cell* 18:11–22.
3. Barbieri CE, et al. (2012) Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nat Genet* 44:685–689.
4. Abida W, et al. (2017) Prospective genomic profiling of prostate cancer across disease states reveals germline and somatic alterations that may affect clinical decision making. *JCO Precis Oncol*, 10.1200/PO.17.00029.
5. Robinson D, et al. (2015) Integrative clinical genomics of advanced prostate cancer. *Cell* 161:1215–1228.
6. Pritchard CC, et al. (2016) Inherited DNA-repair gene mutations in men with metastatic prostate cancer. *N Engl J Med* 375:443–453.
7. Beltran H, et al. (2016) Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. *Nat Med* 22:298–305.
8. Aggarwal R, et al. (2018) Clinical and genomic characterization of treatment-emergent small-cell neuroendocrine prostate cancer: A multi-institutional prospective study. *J Clin Oncol* 36:2492–2503.
9. Aparicio AM, et al. (2016) Combined tumor suppressor defects characterize clinically defined aggressive variant prostate cancers. *Clin Cancer Res* 22:1520–1530.
10. Antonarakis ES, et al. (2014) AR-V7 and resistance to enzalutamide and abiraterone in prostate cancer. *N Engl J Med* 371:1028–1038.
11. Scher HI, et al. (2017) Nuclear-specific AR-V7 protein localization is necessary to guide treatment selection in metastatic castration-resistant prostate cancer. *Eur Urol* 71:874–882.
12. Conteduca V, et al. (2017) Androgen receptor gene status in plasma DNA associates with worse outcome on enzalutamide or abiraterone for castration-resistant prostate cancer: A multi-institution correlative biomarker study. *Ann Oncol* 28:1508–1516.
13. Annala M, et al. (2018) Circulating tumor DNA genomics correlate with resistance to abiraterone and enzalutamide in prostate cancer. *Cancer Discov* 8:444–457.
14. Beltran H, et al. (2019) A phase II trial of the Aurora kinase A inhibitor alisertib for patients with castration-resistant and neuroendocrine prostate cancer: Efficacy and biomarkers. *Clin Cancer Res* 25:43–51.
15. Armenia J, et al.; PCF/SU2C International Prostate Cancer Dream Team (2018) The long tail of oncogenic drivers in prostate cancer. *Nat Genet* 50:645–651.
16. Chakravarty D, et al. (2017) OncoKB: A precision oncology knowledge base. *JCO Precis Oncol*, 10.1200/PO.17.00011.
17. Carver BS, et al. (2009) Aberrant ERG expression cooperates with loss of PTEN to promote cancer progression in the prostate. *Nat Genet* 41:619–624.
18. Ku SY, et al. (2017) Rb1 and Trp53 cooperate to suppress prostate cancer lineage plasticity, metastasis, and antiandrogen resistance. *Science* 355:78–83.
19. Mu P, et al. (2017) SOX2 promotes lineage plasticity and antiandrogen resistance in TP53- and RB1-deficient prostate cancer. *Science* 355:84–88.
20. Blazek D, et al. (2011) The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes. *Genes Dev* 25:2158–2172.
21. Quigley DA, et al. (2018) Genomic hallmarks and structural variation in metastatic prostate cancer. *Cell* 174:758–769.e9, and erratum (2018) 175:889.
22. Viswanathan SR, et al. (2018) Structural alterations driving castration-resistant prostate cancer revealed by linked-read genome sequencing. *Cell* 174:433–447.e19.
23. Wu YM, et al. (2018) Inactivation of CDK12 delineates a distinct immunogenic class of advanced prostate cancer. *Cell* 173:1770–1782.e14.
24. Abida W, Sawyers CL (2018) Targeting DNA repair in prostate cancer. *J Clin Oncol* 36:1017–1019.
25. Hussain M, et al. (2018) Targeting androgen receptor and DNA repair in metastatic castration-resistant prostate cancer: Results from NCI 9012. *J Clin Oncol* 36:991–999.
26. Mateo J, et al. (2018) Clinical outcome of prostate cancer patients with germline DNA repair mutations: Retrospective analysis from an international study. *Eur Urol* 73:687–693.
27. Boysen G, et al. (2018) SPOP-mutated/CHD1-deleted lethal prostate cancer and abiraterone sensitivity. *Clin Cancer Res* 24:5585–5593.
28. Blattner M, et al. (2017) SPOP mutation drives prostate tumorigenesis in vivo through coordinate regulation of PI3K/mTOR and AR signaling. *Cancer Cell* 31:436–451.
29. Beltran H, et al. (2014) Aggressive variants of castration-resistant prostate cancer. *Clin Cancer Res* 20:2846–2850.
30. Bluemn EG, et al. (2017) Androgen receptor pathway-independent prostate cancer is sustained through FGF signaling. *Cancer Cell* 32:474–489.e6.
31. Beltran H, et al. (2011) Molecular characterization of neuroendocrine prostate cancer and identification of new drug targets. *Cancer Discov* 1:487–495.
32. Hieronymus H, et al. (2006) Gene expression signature-based chemical genomic prediction identifies a novel class of HSP90 pathway modulators. *Cancer Cell* 10:321–330.
33. Nava Rodrigues D, et al. (2018) RB1 heterogeneity in advanced metastatic castration resistant prostate cancer. *Clin Cancer Res* 25:687–697.
34. Wyatt AW, et al. (2017) Concordance of circulating tumor DNA and matched metastatic tissue biopsy in prostate cancer. *J Natl Cancer Inst* 110:78–86.
35. Romanel A, et al. (2015) Plasma AR and abiraterone-resistant prostate cancer. *Sci Transl Med* 7:312re10.
36. Epstein JI, et al. (2014) Proposed morphologic classification of prostate cancer with neuroendocrine differentiation. *Am J Surg Pathol* 38:756–767.
37. Mateo J, et al. (2015) DNA-repair defects and olaparib in metastatic prostate cancer. *N Engl J Med* 373:1697–1708.
38. Brastianos PK, et al. (2015) Genomic characterization of brain metastases reveals branched evolution and potential therapeutic targets. *Cancer Discov* 5:1164–1177.
39. Shen R, Seshan VE (2016) FACETS: Allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic Acids Res* 44:e131.
40. Dobin A, et al. (2013) STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21.
41. Robinson DR, et al. (2017) Integrative clinical genomics of metastatic cancer. *Nature* 548:297–303.
42. Cuzick J, et al.; Transatlantic Prostate Group (2012) Prognostic value of a cell cycle progression signature for prostate cancer death in a conservatively managed needle biopsy cohort. *Br J Cancer* 106:1095–1099.
43. Ertel A, et al. (2010) RB-pathway disruption in breast cancer: Differential association with disease subtypes, disease-specific prognosis and therapeutic response. *Cell Cycle* 9:4153–4163.
44. Cerami E, et al. (2012) The cBio Cancer Genomics Portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov* 2:401–404.
45. Chang MT, et al. (2016) Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity. *Nat Biotechnol* 34:155–163.

Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer

Himisha Beltran^{1-3,13}, Davide Prandi^{4,13}, Juan Miguel Mosquera^{1,5}, Matteo Benelli⁴, Loredana Puca¹, Joanna Cyrta¹, Clarisse Marotz¹, Eugenia Giannopoulou⁶, Balabhadrapatruni V S K Chakravarthi⁷, Sooryanarayana Varambally⁷, Scott A Tomlins⁸, David M Nanus^{2,3}, Scott T Tagawa^{2,3}, Eliezer M Van Allen^{9,10}, Olivier Elemento^{1,6}, Andrea Sboner^{1,5,11}, Levi A Garraway^{9,10,12,14}, Mark A Rubin^{1,3,5,14} & Francesca Demichelis^{1,4,11,14}

An increasingly recognized resistance mechanism to androgen receptor (AR)-directed therapy in prostate cancer involves epithelial plasticity, in which tumor cells demonstrate low to absent AR expression and often have neuroendocrine features. The etiology and molecular basis for this ‘alternative’ treatment-resistant cell state remain incompletely understood. Here, by analyzing whole-exome sequencing data of metastatic biopsies from patients, we observed substantial genomic overlap between castration-resistant tumors that were histologically characterized as prostate adenocarcinomas (CRPC-Adeno) and neuroendocrine prostate cancer (CRPC-NE); analysis of biopsy samples from the same individuals over time points to a model most consistent with divergent clonal evolution. Genome-wide DNA methylation analysis revealed marked epigenetic differences between CRPC-NE tumors and CRPC-Adeno, and also designated samples of CRPC-Adeno with clinical features of AR independence as CRPC-NE, suggesting that epigenetic modifiers may play a role in the induction and/or maintenance of this treatment-resistant state. This study supports the emergence of an alternative, ‘AR-indifferent’ cell state through divergent clonal evolution as a mechanism of treatment resistance in advanced prostate cancer.

Prostate cancer remains a leading cause of cancer death in males worldwide¹. The mainstay of therapy for patients with metastatic spread, including castration-resistant disease, is hormonal therapy that targets the AR²⁻⁴. Enzalutamide and abiraterone are potent AR-targeted therapies approved for the treatment of men with castration-resistant prostate cancer (CRPC)^{5,6}. Although the use of these agents improves the survival and quality of life of individuals with CRPC, most patients ultimately develop resistance to them⁷. Predictive biomarkers

that help distinguish responders from nonresponders before starting the next line of hormonal therapy are needed. We and others have observed that a subset of resistant tumors show small-cell carcinoma or neuroendocrine features on metastatic biopsy (CRPC-NE)⁸⁻¹⁰. This phenomenon may therefore reflect an epithelial plasticity that enables tumor adaptation in response to AR-targeted therapies¹¹⁻¹⁴. Prognosis of CRPC-NE is poor owing to late recognition, heterogeneous clinical features and lack of effective systemic therapies.

One major hurdle in the diagnosis and treatment of androgen-independent prostate cancer, including CRPC-NE, is our lack of understanding of the genetic and epigenetic underpinnings of this aggressive subset. To address this we interrogated 114 metastatic tumor specimens from 81 individuals—including 51 with clinical and histologic features of CRPC-Adeno and 30 with features of CRPC-NE, as confirmed by pathologic consensus criteria⁸; we studied matched normal cells from all of the patients, multiple tumor biopsies from 17 patients and single tumor biopsies from 64 patients. We hypothesized that CRPC-NE could be distinguished from CRPC-Adeno on the basis of distinct molecular alterations and that this information could improve upon and supplement the current, often challenging diagnostic features that are reliant on morphological characterization. We also hypothesized that CRPC-NE that develops after therapy arises clonally from a CRPC-Adeno precursor, rather than from the selection of pre-existing neuroendocrine clones. Finally, we hypothesized that AR-independent prostate adenocarcinomas that share CRPC-NE-specific molecular alterations may represent tumors that are at high risk for either progression or transition to CRPC-NE.

We evaluated biopsies from a wide range of metastatic sites on the basis of the presence and accessibility of lesions, with more bone biopsies obtained from patients with CRPC-Adeno as compared to those with CRPC-NE (31% CRPC-Adeno versus 2% CRPC-NE;

¹Caryl and Israel Englander Institute for Precision Medicine, New York Presbyterian Hospital–Weill Cornell Medicine, New York, New York, USA. ²Department of Medicine, Division of Hematology and Medical Oncology, Weill Cornell Medicine, New York, New York, USA. ³Sandra and Edward Meyer Cancer Center at Weill Cornell Medicine, New York, New York, USA. ⁴Centre for Integrative Biology, University of Trento, Trento, Italy. ⁵Department of Pathology and Laboratory Medicine, Weill Cornell Medicine, New York, New York, USA. ⁶Department of Physiology and Biophysics, Weill Cornell Medicine, New York, New York, USA. ⁷Department of Pathology, University of Alabama, Birmingham, Alabama, USA. ⁸Department of Pathology, University of Michigan, Ann Arbor, Michigan, USA. ⁹Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA. ¹⁰Broad Institute of the Massachusetts Institute of Technology (MIT) and Harvard University, Cambridge, Massachusetts, USA. ¹¹Institute for Computational Biomedicine, Weill Cornell Medicine, New York, New York, USA. ¹²Center for Cancer Precision Medicine, Dana-Farber Cancer Institute and Brigham and Women’s Hospital, Boston, Massachusetts, USA. ¹³These authors contributed equally to this work. ¹⁴These authors jointly directed this work. Correspondence should be addressed to H.B. (hip9004@med.cornell.edu), M.A.R. (rubinma@med.cornell.edu) or F.D. (f.demichelis@unitn.it).

Received 9 October 2015; accepted 11 January 2016; published online 8 February 2016; doi:10.1038/nm.4045

$P < 0.05$, binomial test) (Fig. 1a). The clinical and pathologic features of the biopsy samples are summarized in **Supplementary Table 1** and **Supplementary Figure 1**. As expected, we found lower protein expression of the AR, on average, in CRPC-NE samples, using immunohistochemistry (Fig. 1b). We also quantified AR signaling status by measuring the expression of mRNAs that were included in a previously defined AR signature by Hieronymus *et al.*¹⁵ (**Supplementary Table 2**) and observed overall lower abundance of these mRNAs in CRPC-NE specimens as compared to CRPC-Adeno specimens (Fig. 1b); however, there was overlap with a wide range of values observed within each subtype, suggesting that there is a spectrum of AR signaling in advanced prostate cancer that spans pathologic subtypes.

To deepen our understanding of AR independence in general—and of the CRPC-NE phenotype in particular—we first performed whole-exome sequencing (WES) of 114 metastatic tumor–normal tissue pairs. The mutational landscape of CRPC-NE tumors was similar to that of CRPC-Adeno, but it was also consistent with those in published studies of CRPC-NE tumors (Fig. 1c)—including enrichment for the loss of *RBI*, which encodes the retinoblastoma tumor suppressor protein (in 70% of CRPC-NE versus 32% of CRPC-Adeno samples; $P = 0.003$, proportion test), and for the mutation or deletion of *TP53*, which encodes the p53 tumor suppressor protein (in 66.7% of CRPC-NE versus 31.4% of CRPC-Adeno samples; $P = 0.0043$, proportion test). Loss of *RBI* is common in primary small-cell prostate and lung carcinomas, and it promotes small-cell carcinoma pathogenesis in conjunction with a *TP53* mutation^{16–18}; in our series, concurrent loss of *RBI* and *TP53* was present in 53.3% of CRPC-NE versus 13.7% of CRPC-Adeno samples ($P < 0.0004$, proportion test).

Another feature distinguishing CRPC-NE tumors from CRPC-Adeno was a paucity of somatic alterations involving the *AR* gene in the former ($P < 0.0001$, Wilcoxon test; Fig. 1d,e). Genomic amplification, activating point mutations and splice variants involving *AR* are commonly observed in CRPC-Adeno and are associated with treatment resistance to AR-directed therapies^{19,20}. This observation was confirmed in our cohort; 29 samples showed a focal amplification or a point mutation in *AR*, and 21 samples had alterations in genes encoding known AR co-activators (*FOXA1*, *NCOR1*, *NCOR2* and *ZBTB16*). In contrast, *AR* point mutations were notably absent in CRPC-NE samples, and amplifications, when present, were of low level, broad and explained by tumor polyploidy. Although potentially affected by differences in prior therapies, we speculate that the absence of *AR* genomic alterations in CRPC-NE tumors may be due to the clonal selection of non-amplified CRPC-Adeno tumor subpopulations through selective pressure (in the context of AR-directed therapies). The *AR* splice variant 7 (AR-V7), which is associated with AR-driven resistance in CRPC^{21,22}, was observed in both CRPC-Adeno and CRPC-NE tumors, and although overall *AR* expression was lower in CRPC-NE tumors, the ratio of AR-V7 to wild-type *AR* mRNA was significantly lower in CRPC-NE samples than in CRPC-Adeno samples ($P = 0.0025$, Wilcoxon test) (**Supplementary Fig. 2**). These data suggested that AR signaling is attenuated in CRPC-NE. Together with the frequent loss of *RBI* and *TP53* in this subtype, these findings imply that distinct biological properties are associated with CRPC-NE.

The overall spectrum of genomic alterations, however, was similar in tumors of both CRPC subtypes (**Supplementary Fig. 2**) and was maintained after controlling for the site of biopsy and for prior exposure to chemotherapy (**Supplementary Fig. 3**). There were no significant differences in the distribution of nonsilent point mutations, polyploidy or copy-number genomic burden between the

CRPC-Adeno and CRPC-NE tumor subtypes (on average >30% of the genome was aberrant), and tumors of both subtypes showed a significantly higher number of clonal point mutations as compared with those of clinically localized prostate cancer ($P < 10^{-7}$, Wilcoxon test)^{23,24}. The median number of nonsilent single-nucleotide variants (SNVs) in metastatic samples was 41 (range: 2–729) (**Supplementary Tables 3 and 4**). Five of the six samples with the highest number of SNVs (115–663) showed genetic and/or protein expression alterations involving DNA mismatch repair genes (**Supplementary Fig. 4**), consistent with prior studies^{20,25}. Recurrent alterations are shown in **Figure 1c** and putative gene fusions (as determined by RNA sequencing (RNA-seq) analysis) are in **Supplementary Table 5**.

The substantial overlap in the overall somatic copy-number landscape between CRPC-Adeno and CRPC-NE tumors (Fig. 1f) was noteworthy in light of the marked genomic differences between adenocarcinomas and small-cell carcinomas observed in other tissue subtypes (for example, lung and gastrointestinal tract tumors). After correcting for admixture of non-tumor cells and for ploidy²⁶, we sought to identify regions of the genome that were differentially altered and noted copy-number alterations that are enriched in either CRPC-NE or CRPC-Adeno samples (**Supplementary Tables 6 and 7**). Putative cancer genes²⁷ within differentially deleted regions in CRPC-NE with concordant downregulation at the mRNA level (false-discovery rate (FDR) < 10% for both DNA and mRNA) are highlighted in **Figure 1f**. The top ranked gene was *CYLD*, which encodes cylindromatosis, a deubiquinating enzyme reported as a tumor suppressor involved in negative regulation of multiple signaling pathways including those involving nuclear factor (NF)- κ B²⁸, transforming growth factor (TGF)- β and Notch²⁹. *CYLD* was deleted in 51% of CRPC-NE samples and verified by fluorescence *in situ* hybridization analysis (**Supplementary Fig. 5**). We found that genomic loss of *CYLD* was associated with reduced mRNA expression and a modest decrease in the expression of genes encoding factors that are associated with AR signaling (in this study, in the SU2C/PCF 2015 cohort of CRPC²⁰, and in cell lines; **Supplementary Fig. 5**), suggesting that *CYLD* loss alone may be insufficient to promote AR indifference, and that it might cooperate with other molecular alterations to do so. By extending the computational framework of the algorithm CLONET^{23,26}—which first estimates the purity and ploidy of tumor samples and then calculates the clonality of somatic aberrations—we assessed for allele-specific copy-number clonality (**Supplementary Note**) and found both focal and broad copy-neutral or copy-aberrant loss of heterozygosity across our cohort (**Supplementary Fig. 5**). This included focal allelic imbalance of the *DEK* proto-oncogene in CRPC-NE tumors as compared to CRPC-Adeno ($P = 0.04$, binomial test). *DEK* belongs to a class of DNA topology modulators with a reported oncogenic role in prostate cancer, including in CRPC-NE^{30,31}.

In principle, several possible models could explain tumor evolution from a prostate adenocarcinoma to CRPC-NE (Fig. 2a). This process may rely on linear expansion of subclonal or clonal tumor cell populations originating from the primary tumor, with sequential acquisition of genomic alterations (linear)³². Alternatively, independent clones within the primary tumor or the metastasis could give rise to parallel and distinct resistant tumor populations (independent)³³. Finally, there may be genetic diversification in the primary tumor or in the metastatic lesion as a mechanism of adaptation, leading to selective pressure and divergent clonal evolution (divergent). Although prior studies have demonstrated concordance of the prostate cancer–specific genomic rearrangement *TMPRSS2-ERG*⁹ and other single-gene alterations^{18,34}

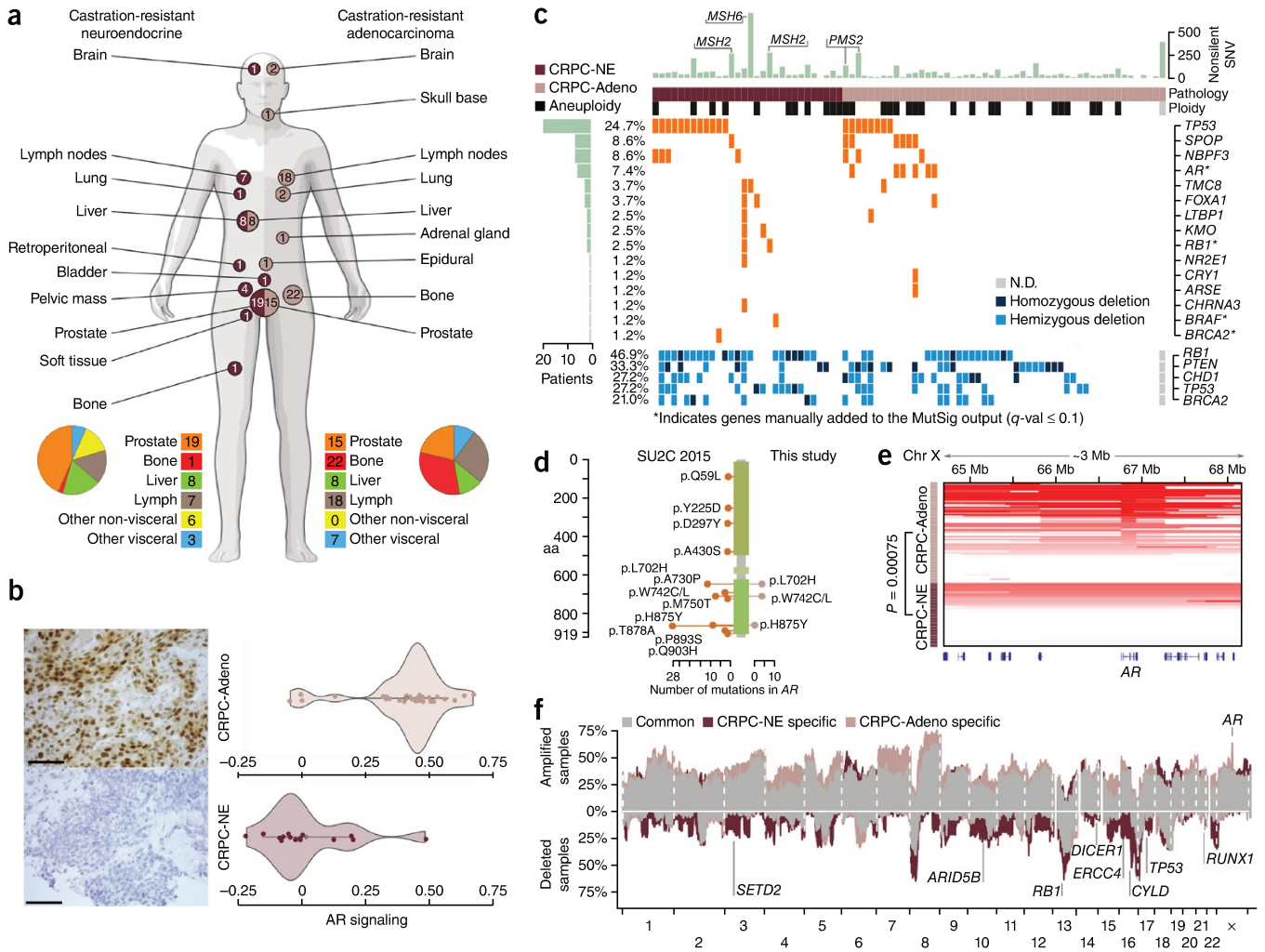


Figure 1 Clinical and mutational profile of the cohort. **(a)** Schematic illustrating the biopsy sites for CRPC-NE (dark pink) and CRPC-Adeno (light pink) patient subgroups. The numbers in the circles (and colored squares associated with the pie charts) indicate the numbers of samples from each site (or group of sites). The pie charts show the distribution of samples from each site for each subtype. **(b)** Left, immunohistochemical staining for AR protein expression in CRPC-Adeno (top) and CRPC-NE (bottom) samples. Scale bars, 50 μ m. Right, violin plots, based on the abundance of mRNA transcripts included in the AR signaling signature described previously by Hieronymus *et al.*¹⁵, showing the density of AR signaling in CRPC-Adeno (top) and CRPC-NE (bottom) samples. Each dot represents a sample; diamonds and solid lines represent the mean and 95% confidence interval, respectively. **(c)** Analysis for mutated genes. Each row represents a gene and each column an individual subject. Light green bars at the top correspond to the total number of nonsilent SNVs in an individual. Light green bars on the left indicate the number of subjects with nonsilent mutations in the genes indicated on the right. Bottom panel reports the copy number status of selected genes. N.D., not defined **(d)** Genomic location of AR mutations in samples from the SU2C/PCF 2015 cohort and this study. aa, amino acid. **(e)** Copy-number status of the AR locus. Color intensity is indicative of the level of amplification (darker color indicates higher number of AR copies) and the location shows the focality of AR amplification. **(f)** Frequency of copy-number aberrations that are concordant (gray color), CRPC-NE specific (dark pink) or CRPC-Adeno specific (light pink). Data are adjusted for tumor ploidy and purity; x axis denotes genomic position. Highlighted genes are significantly and preferentially aberrant in one class and demonstrate concordant differential mRNA levels (for DNA and mRNA: FDR \leq 10% for deletions and $P \leq$ 1% for amplifications); binomial test adjusted for multiple-hypotheses testing with Benjamini-Hochberg (FDR) or Mann-Whitney Wilcoxon (P value) correction

between adenocarcinoma and neuroendocrine foci in mixed tumors, suggesting a common cell of origin, these complex patterns of genetic evolution have not been more rigorously evaluated. To address this and to infer clonal-expansion dynamics, we studied serial tumor samples from individuals during the course of their disease. Patient WCMC7520 had a prostatectomy for a clinically localized Gleason 9 prostate adenocarcinoma with local lymph node involvement and was treated initially with adjuvant androgen-deprivation therapy (ADT) followed by chemotherapy at the time of metastatic disease and castration resistance (Fig. 2b). Twenty-nine months after starting ADT, he developed CRPC-NE, which was diagnosed by a metastatic biopsy. We evaluated

the primary tumor, the local lymph node and the CRPC-NE distant metastasis from two time points and observed a homozygous deletion of the BRCA2 tumor suppressor gene and a mutation in TP53 in biopsy samples from all of the sites, suggesting a common ancestor. Allele-specific DNA analysis highlighted diverse genomic states of other key genes such as MYCN, which encodes the N-myc oncoprotein (Fig. 2b); MYCN has been previously described as oncogenic in CRPC-NE⁹. After comparison of the samples from all of the sites, the patient's primary prostate specimen had lesions that suggested divergent, but clonal, paths to the lymph node and both distant metastases. However, from these data we cannot definitively rule out

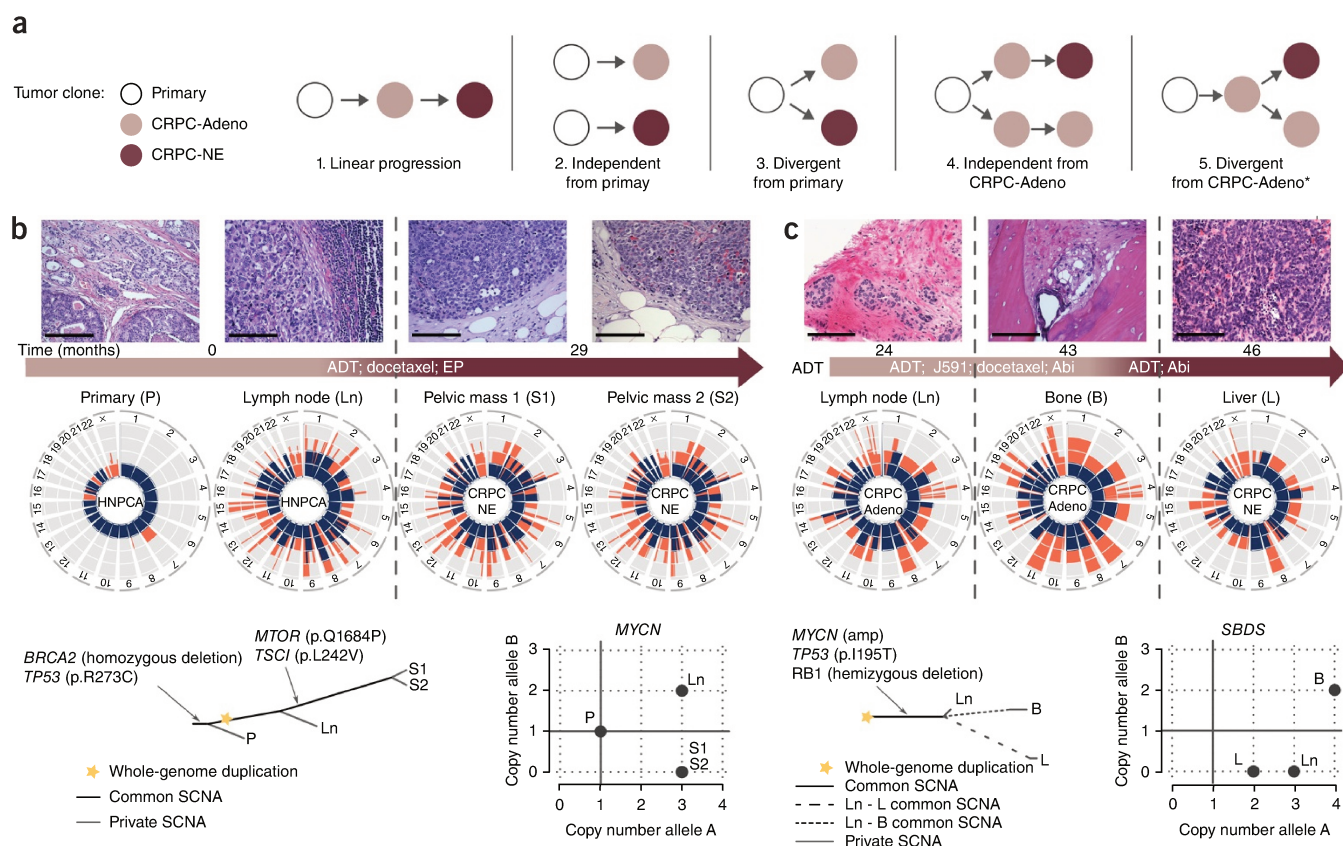


Figure 2 Tracing CRPC-NE emergence through allele-specific analysis. **(a)** Potential models of tumor evolution toward the neuroendocrine phenotype include linear progression from primary untreated adenocarcinoma to CRPC-Adeno to CRPC-NE (1), independent progression of two distinct subclonal or clonal populations within the primary (2) or castration-resistant (4) CRPC-Adeno to either CRPC-Adeno or CRPC-NE, and divergent clonal evolution of CRPC-NE from either primary adenocarcinoma (3) or CRPC-Adeno (5). Asterisk (*) indicates the favored model. **(b)** Allele-specific analysis of primary prostate adenocarcinoma and local lymph node metastasis removed at time of radical prostatectomy (RP) and of two metastatic CRPC-NE (treated) tumors (S1 and S2; at 3 years after RP from subject WCMC7520). Top, H&E-stained images for each sample (scale bars, 100 μ m) and intervening therapies are shown above and in the timeline, respectively. ADT, androgen-deprivation therapy; EP, etoposide and cisplatin chemotherapy. Middle, Circos plots summarizing genome-wide allele-specific DNA quantity in the tumor cells. Orange and blue bars denote the larger and the smaller copy-number values (respectively) of the two alleles along the genome. HNPCA, hormone-naive PCA. Bottom left, phylogenetic tree built from allele-specific copy-number data. Bottom right, allele-specific copy number of *MYCN*. SCNA, somatic copy-number aberration. **(c)** Allele-specific analysis of tumors at three time points from subject WCMC161: lymph node (CRPC-Adeno), bone biopsy (CRPC-Adeno) and liver biopsy (small-cell CRPC-NE). Top, H&E-stained images for each sample (scale bars, 100 μ m) and intervening therapies are shown above and in the timeline, respectively. Abi, abiraterone acetate with prednisone; J591, radiolabeled monoclonal antibody targeting prostate-specific membrane antigen. Middle, Circos plots summarizing genome-wide allele-specific DNA quantity in the tumor cells. Bottom left, phylogenetic tree built from allele-specific copy-number data. Bottom right, allele-specific copy number of *SBDS*.

metastasis-to-metastasis seeding, which was recently proposed as a mechanism of tumor progression³⁵. Patient WCMC161 showed progression of disease after multiple lines of therapy for CRPC—including the development of new visceral metastases at the time of progression on abiraterone with a liver biopsy showing small-cell carcinoma (Fig. 2c and Supplementary Fig. 6). Comparison of metastases from three different time points—CRPC-Adeno (adenocarcinoma, lymph node metastasis), CRPC-Adeno (adenocarcinoma, bone metastasis) and CRPC-NE (small-cell carcinoma, liver metastasis at progression on abiraterone therapy)—suggested divergent clonal evolution, as exemplified by the Shwachman-Bodian-Diamond syndrome (*SBDS*) gene (in which the allele-specific state of *SBDS* in metastatic samples was not compatible with linear evolution; Fig. 2c) and that the lymph node, bone and liver metastases arose from a common ancestor. Phylogenetic trees outlining disease evolution from the adenocarcinoma to a neuroendocrine phenotype in individual patients, constructed through analyses of additional multitumor cases (Supplementary Fig. 6), collectively rule against a parallel evolutionary model. On the basis of

these temporal case series, the high degree of clonality in CRPC and the overall similarity between the genomic profiles of CRPC-Adeno and CRPC-NE specimens, the most parsimonious model that explains the data is divergent clonal evolution of metastatic CRPC to either an AR-driven or an AR-independent state (Fig. 2a, model 5). In other words, CRPC-NE tumors appear clonal in origin with a clonal ancestry traceable back to a CRPC-Adeno precursor.

Although informative, the observed DNA changes did not appear to fully explain the clinical aggressiveness of CRPC-NE. We therefore posited that this phenotype may also be mediated by epigenetic changes. To this end, we evaluated CpG-rich methylation on a genome-wide scale by single-cytosine-resolution DNA methylation analysis (using enhanced reduced-representation bisulfite sequencing (eRRBS)). In contrast to the mostly similar genomic data, the CRPC-NE and CRPC-Adeno subtypes showed strong epigenetic segregation, using unsupervised analysis of unselected methylation sites (Fig. 3a, Supplementary Fig. 7 and Supplementary Table 8). In addition, methylation of first exons and gene promoters was associated with

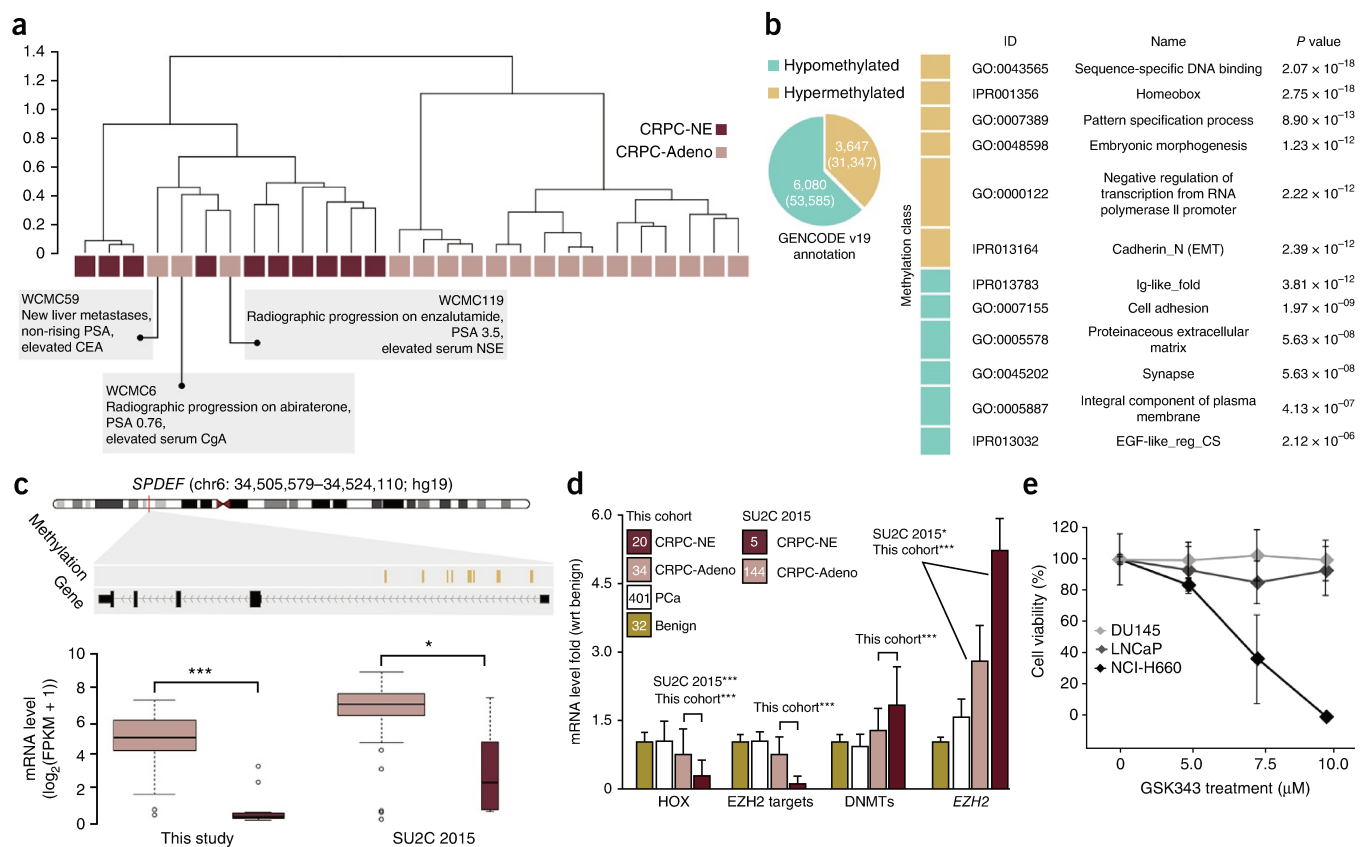


Figure 3 Methylation analysis of CRPC-NE and CRPC-Adeno tumors. **(a)** Hierarchical clustering of eRRBS data from 28 samples, using ‘1 – Pearson’s correlation’ as distance measure on unselected sites. Clinical features of outlier cases are described. CEA, carcinoembryonic antigen; CgA, chromogranin A; NSE, neuron-specific enolase. **(b)** Left, pie chart showing the number of differentially methylated genes, identified by annotating hyper- and hypomethylated loci (number is reported in parentheses) on GENCODE version 19. Right, selection of functional categories enriched after analysis of differentially methylated genes. **(c)** Top, genome track of *SPDEF* (chr6: 34,505,579–34,524,110; hg19). Bottom, box plot of *SPDEF* expression levels for samples in this study (left) and in the SU2C/PCF 2015 (SU2C 2015; right) cohort. * $P < 0.05$, *** $P < 0.001$; by Wilcoxon test. Lower and upper bars correspond to the minimum and maximum non-outlier values of the data distribution. Outliers are defined as values outside of the range ($Q1 - 1.5 \times (Q3 - Q1)$, $Q3 + 1.5 \times (Q3 - Q1)$), where $Q1$ and $Q3$ are the first and third quartile, respectively. **(d)** Bar plots showing the effect of *EZH2* transcription activity across 487 samples of differing pathology classifications. The bars are relative to the median values of changes in mRNA expression, with respect to benign prostate tissue samples, of homeobox genes with reduced expression in CRPC-NE versus CRPC-Adeno samples (FDR < 0.1); expression changes in a selection of mRNAs of *EZH2* target genes (*DKK1*, *NKD1*, *AMD1*, *HOXA13*, *HOXA11* and *NKX3-1*), DNA methyltransferase genes (indicated as DNMTs; *DNMT1*, *DNMT3B*, *DNMT3A* and *DNMT3L*) and *EZH2* are shown. Significance of differences between CRPC-NE and CRPC-Adeno subgroups are shown (maximum $P = 3 \times 10^{-5}$ for DNMTs). When significant, P values estimated in the SU2C/PCF 2015 cohort are shown. The number of samples for each pathology classification is reported inside the square symbols of the legend. Error bars are median absolute deviation. * $P < 0.05$, *** $P < 0.001$; by Wilcoxon test. **(e)** Cell viability in the prostate adenocarcinoma cell lines DU145 and LNCaP, as well as in the neuroendocrine prostate cell line NCI-H660, as assessed at 48 h after treatment with escalating doses of the *EZH2* inhibitor GSK343 (5.0, 7.5 and 10 μ M). Error bars are standard error of the mean.

significant changes in gene expression ($P < 0.001$, Wilcoxon test; **Supplementary Fig. 8**). Overall, a greater fraction of concordant epigenetic and transcriptomic events were observed with increasing significance by differential expression analysis (**Supplementary Fig. 8**), and 22% of the top dysregulated transcripts in CRPC-NE demonstrated concordant changes in DNA methylation ($P < 0.0002$). This raised the possibility that the transition to, or the advent of, the CRPC-NE subtype is associated with epigenetic dysregulation. Notably, the epigenetic signal identified three cases with clinical features of AR independence that were binned as adenocarcinoma on the basis of standard pathology but that segregated with CRPC-NE after unsupervised analysis (**Fig. 3a** and **Supplementary Fig. 8**). All three of these patients demonstrated radiographic progression in the setting of a stable or low serum amount of the androgen-regulated protein prostate-specific antigen (PSA). These data suggest that clustering predictions based on DNA methylation may provide additional

information associated with AR independence and CRPC-NE that potentially improves on those based on tumor morphology.

By using functional enrichment analysis of differentially methylated genes, we identified epigenetically dysregulated pathways, including those involving neuronal, cell-cell adhesion, developmental, epithelial-mesenchymal transition (EMT) and stem cell programs (**Fig. 3b**), which are thought to be relevant for CRPC-NE progression³⁶. Among our notable findings (**Supplementary Fig. 7**), we observed hypermethylation and reduced expression of the tumor suppressor gene *SPDEF* in CRPC-NE ($P < 10^{-9}$, Wilcoxon test) (**Fig. 3c**). *SPDEF* (which encodes a prostate-derived Ets factor) is a transcriptional activator and a regulator of cellular differentiation that has been reported to be involved in suppression of tumor metastasis through inhibition of EMT in prostate cancer³⁷. We confirmed promoter methylation of the *SPDEF* gene and downregulation of mRNA expression in the neuroendocrine prostate cancer cell line

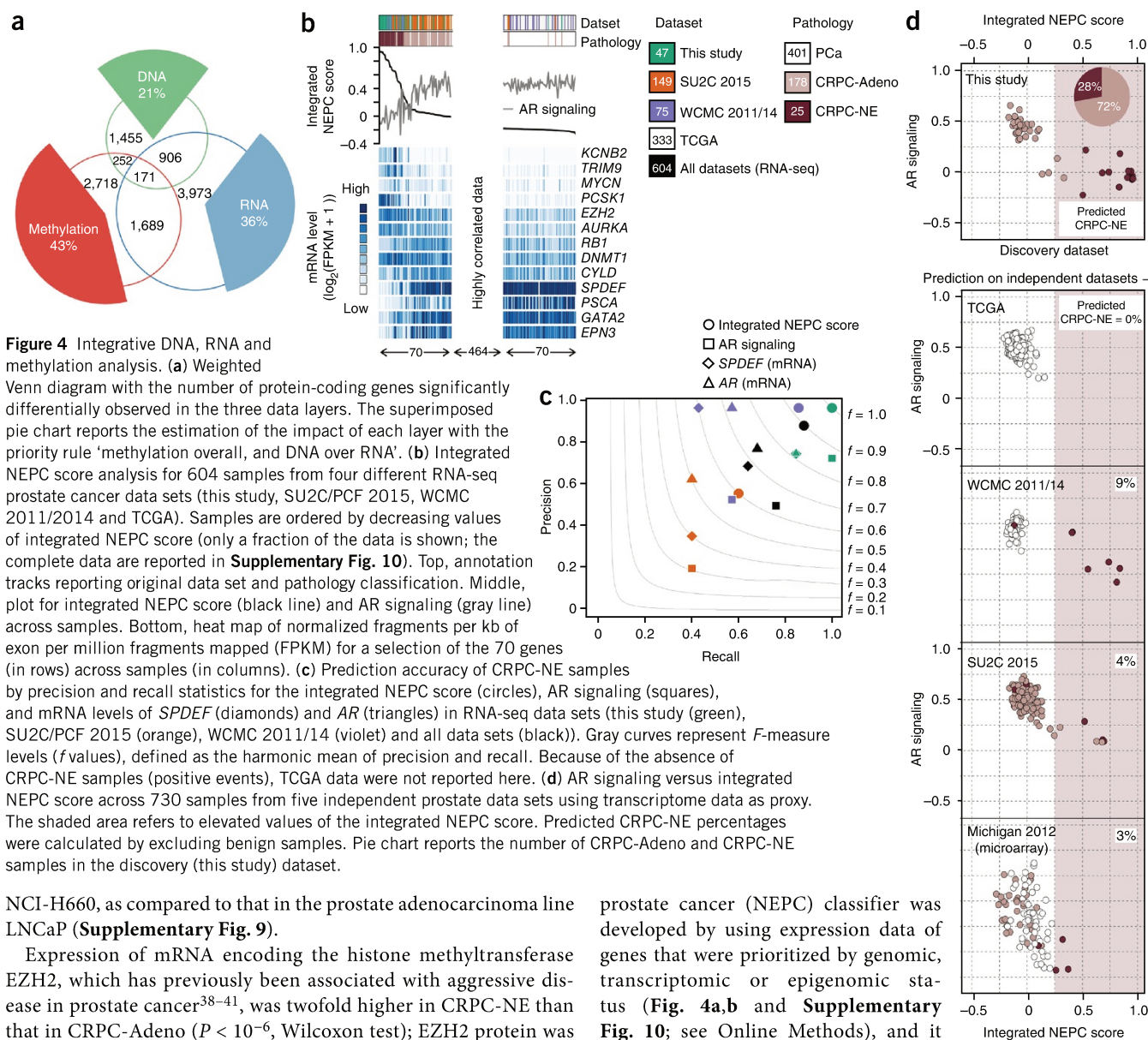


Figure 4 Integrative DNA, RNA and methylation analysis. **(a)** Weighted Venn diagram with the number of protein-coding genes significantly differentially observed in the three data layers. The superimposed pie chart reports the estimation of the impact of each layer with the priority rule 'methylation overall, and DNA over RNA'. **(b)** Integrated NEPC score analysis for 604 samples from four different RNA-seq prostate cancer data sets (this study, SU2C/PCF 2015, WCMC 2011/2014 and TCGA). Samples are ordered by decreasing values of integrated NEPC score (only a fraction of the data is shown; the complete data are reported in **Supplementary Fig. 10**). Top, annotation tracks reporting original data set and pathology classification. Middle, plot for integrated NEPC score (black line) and AR signaling (gray line) across samples. Bottom, heat map of normalized fragments per kb of exon per million fragments mapped (FPKM) for a selection of the 70 genes (in rows) across samples (in columns). **(c)** Prediction accuracy of CRPC-NE samples by precision and recall statistics for the integrated NEPC score (circles), AR signaling (squares), and mRNA levels of *SPDEF* (diamonds) and *AR* (triangles) in RNA-seq data sets (this study (green), SU2C/PCF 2015 (orange), WCMC 2011/14 (violet) and all data sets (black)). Gray curves represent *F*-measure levels (*f* values), defined as the harmonic mean of precision and recall. Because of the absence of CRPC-NE samples (positive events), TCGA data were not reported here. **(d)** AR signaling versus integrated NEPC score across 730 samples from five independent prostate data sets using transcriptome data as proxy. The shaded area refers to elevated values of the integrated NEPC score. Predicted CRPC-NE percentages were calculated by excluding benign samples. Pie chart reports the number of CRPC-Adeno and CRPC-NE samples in the discovery (this study) dataset.

NCI-H660, as compared to that in the prostate adenocarcinoma line LNCaP (**Supplementary Fig. 9**).

Expression of mRNA encoding the histone methyltransferase *EZH2*, which has previously been associated with aggressive disease in prostate cancer^{38–41}, was twofold higher in CRPC-NE than that in CRPC-Adeno ($P < 10^{-6}$, Wilcoxon test); *EZH2* protein was also more abundant in CRPC-NE (**Supplementary Fig. 9**). These findings add to its previously reported differential-expression status in CRPC-NE versus localized prostate cancer^{9,42}. Furthermore, *EZH2*-repressed target genes^{43,44} were significantly downregulated in CRPC-NE ($P < 10^{-7}$, Wilcoxon test) (**Fig. 3d**), including WNT signaling genes such as *DKK1* ($P = 0.0002$, Wilcoxon test), *NKD1* ($P = 0.0046$, Wilcoxon test) and homeobox factor (HOX)-encoding genes ($P = 0.001$, Wilcoxon test). Treatment with the *EZH2* inhibitor GSK343 resulted in a preferential decrease in the viability of NCI-H660 cells, as compared to that in other non-neuroendocrine prostate cancer cell lines (**Fig. 3e**), with a substantial downregulation of several CRPC-NE-associated genes after treatment, including *NCAM* (CD56), *MYCN*⁹ and *PEG10* (ref. 45) (**Supplementary Fig. 9**). Overall these data support a key role of the epigenome in the emergence and/or maintenance of CRPC-NE.

On the basis of the current gaps in the clinical and molecular assessment of CRPC-NE, we used these data to develop a molecular classifier to potentially improve the often-challenging diagnosis of CRPC-NE. This integrated 70-gene neuroendocrine

prostate cancer (NEPC) classifier was developed by using expression data of genes that were prioritized by genomic, transcriptomic or epigenomic status (**Fig. 4a,b** and **Supplementary Fig. 10**; see Online Methods), and it demonstrated both a precision and a recall of >0.99 in identifying CRPC-NE in our discovery cohort (**Fig. 4c,d** and **Supplementary Table 9**). Included within the classifier were also genes we had previously described as being over-expressed in CRPC-NE⁹, including *AURKA* ($P < 10^{-5}$, Wilcoxon test) and *MYCN* ($P < 10^{-4}$, Wilcoxon test) (**Supplementary Fig. 10**). Interrogation of transcriptomic data from 683 prostate samples (**Supplementary Table 10**) using data sets from The Cancer Genome Atlas (TCGA)⁴⁶, Grasso *et al.*¹⁹ (Michigan 2012 cohort), Robinson *et al.*²⁰ (SU2C/PCF 2015 cohort) and our group^{9,47} (WCMC 2011/14) revealed an elevated NEPC classifier score in up to 8% of metastatic tumors ($n = 191$) and none of treatment-naive prostate adenocarcinomas ($n = 460$) or of benign prostates ($n = 32$) (**Fig. 4d** and **Supplementary Figs. 10** and **11**). Of those with an elevated classifier score, we reviewed the pathology of those specimens and found that $>80\%$ of them had pathologic features of CRPC-NE (the remaining 20% were adenocarcinomas) (**Fig. 4d** and **Supplementary Fig. 12**). Although there were relatively small numbers of CRPC-NE cases in the validation cohorts, the integrated NEPC classifier demonstrated

superior precision and/or recall across data sets as compared to other variables, such as conventional neuroendocrine markers (i.e., *CGHA*, *SYP*, *NSE* and *CD56* transcripts) plus PSA (**Supplementary Fig. 12**), *AR* mRNA expression, *AR* signaling status and highly ranked differentially expressed genes (such as *SPDEF*) (**Fig. 4c**). Although we recognize the influence of other factors, including differences in prior therapies, on gene expression changes, we posit that castration-resistant tumors with a moderate or rising NEPC score may represent tumors with *AR*-independent features that are either in transition to or at a high risk for CRPC-NE progression during treatment with *AR* therapies. In fact, a subpopulation of prostate adenocarcinoma cells (LNCaP) that were treated long term with enzalutamide acquired molecular features of CRPC-NE (i.e., methylation of *SPDEF*) (**Supplementary Figs. 8 and 13**).

Although we identified cases with low *AR* signaling and NEPC classifier scores, and cannot exclude a less common alternatively distinct subset, our data favors a continuum of progression from an *AR*-driven state to an *AR*-independent state that is associated with neuroendocrine molecular features. These findings warrant the clinical evaluation of more specimens to further investigate the potential superiority and applicability of the classifier diagnostic criteria over more conventional diagnostic criteria and to verify whether this could be useful as a prognostic or predictive biomarker (associated with lack of response to *AR* therapies). Notably, this approach is amenable to metastatic prostate cancer biopsies in which tissue availability is limited and multiple immunohistochemical assays for current diagnostic methods are often impractical. The incorporation of different molecular layers helps apply the classifier to different data sets when only parts are available (DNA, RNA or methylation) (**Supplementary Table 9**) and paves the way for future studies that might apply the classifier to types of samples (such as circulating tumor DNA⁴⁸). If CRPC-NE alterations could be detected earlier during CRPC-Adeno disease progression, for instance, then such individuals could potentially be selected for CRPC-NE-directed (such as platinum chemotherapy) rather than *AR*-targeted systemic therapies or for potentially co-targeting therapeutic approaches. Furthermore, these data set the stage for dynamic testing of the reversibility of the CRPC-NE state with early intervention or epigenetic modifiers, possibly including EZH2 inhibitors.

In summary, our data support divergent evolution of CRPC-NE from one or more CRPC-Adeno cells (adaptation) rather than linear or independent clonal evolution, with selective pressure of subclonal populations with wild-type *AR* and the acquisition of new genomic and epigenomic drivers associated with decreased *AR* signaling and epithelial plasticity. However, there are also other possibilities that cannot be fully excluded, such as the de-differentiation of an adenocarcinoma to a more progenitor-like cell state—with some cells subsequently adopting neuroendocrine features due to local effects.

METHODS

Methods and any associated references are available in the [online version of the paper](#).

Accession codes. All BAM files and associated sample information are described in **Supplementary Table 11**; data are deposited in dbGap [phs000909.v.p1](#) and are accessible on the cBIO Portal for Cancer Genomics.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank our patients and their families for participation in this study. We would also like to acknowledge R. Montgomery, A. Armstrong and R. Szmulewitz for contributing samples and S.S. Chae for his technical assistance. H.B. is the Damon Runyon-Gordon Family Clinical Investigator and is supported (in part) by the Damon Runyon Cancer Research Foundation (award no. CI-67-13). This work was also supported by the Ann and William Bresnan Foundation (H.B. and D.M.N.), the Department of Defense (grant no. PC121341; H.B.), the US National Institutes of Health (NIH) (grant no. R01 CA116337 (H.B., F.D. and M.A.R.), R01CA157845 (S.V.), R01 CA183857 (S.A.T.), 1K08CA188615 (E.M.V.A.), U54 HG003067 (L.A.G.), U01CA162148 (L.A.G.), 5U01 CA11275-09 (J.M.M., M.A.R. and F.D.), the Starr Cancer Consortium (H.B., L.A.G. and M.A.R.), the American-Italian Cancer Foundation (L.P.), the Nuovo Soldati Foundation (J.C.), the A. Alfred Taubman Medical Institute (S.A.T.), a Prostate Cancer Foundation Young Investigator Award (E.M.V.A.), the Associazione Italiana per la Ricerca sul Cancro (AIRC; grant no. IG 13562; F.D.), the European Research Council (consolidator grant (CoG) SPICE (Synthetic lethal phenotype identification through cancer evolution analysis); F.D.) and the Prostate Cancer Foundation (H.B., S.A.T., M.A.R. and F.D.). H.B., J.M.M., S.A.T., D.M.N., S.T.T., E.M.V.A., O.E., A.S., L.A.G., M.A.R. and F.D. are supported by a Stand Up To Cancer-Prostate Cancer Foundation Prostate Dream Team translational cancer research grant. Stand Up To Cancer is a program of the Entertainment Industry Foundation that is administered by the American Association for Cancer Research (grant no. SU2C-AACR-DT0712).

AUTHOR CONTRIBUTIONS

H.B., M.A.R., L.A.G. and F.D. initiated and designed the study; H.B., S.A.T., D.M.N. and S.T.T. enrolled subjects and contributed samples and clinical data; J.M.M., L.P., J.C., C.M., B.V.S.K.C. and S.V. performed experiments; D.P., M.B., E.G., E.M.V.A., O.E., A.S. and F.D. did the statistical and bioinformatics analyses; H.B., M.A.R., L.A.G. and F.D. supervised the research; H.B., M.A.R., L.A.G. and F.D. wrote the first draft of the manuscript; and all authors contributed to the writing and editing of the revised manuscript, and approved the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

1. Siegel, R.L., Miller, K.D. & Jemal, A. Cancer statistics, 2015. *CA Cancer J. Clin.* **65**, 5–29 (2015).
2. Huggins, C. & Hodges, C.V. Studies on prostatic cancer: I. the effect of castration, of estrogen and of androgen injection on serum phosphatases in metastatic carcinoma of the prostate. 1941. *J. Urol.* **168**, 9–12 (2002).
3. Chen, C.D. *et al.* Molecular determinants of resistance to antiandrogen therapy. *Nat. Med.* **10**, 33–39 (2004).
4. Knudsen, K.E. & Scher, H.I. Starving the addiction: new opportunities for durable suppression of *AR* signaling in prostate cancer. *Clin. Cancer Res.* **15**, 4792–4798 (2009).
5. Scher, H.I. *et al.*; AFFIRM Investigators. Increased survival with enzalutamide in prostate cancer after chemotherapy. *N. Engl. J. Med.* **367**, 1187–1197 (2012).
6. de Bono, J.S. *et al.*; COU-AA-301 Investigators. Abiraterone and increased survival in metastatic prostate cancer. *N. Engl. J. Med.* **364**, 1995–2005 (2011).
7. Watson, P.A. *et al.* Emerging mechanisms of resistance to androgen receptor inhibitors in prostate cancer. *Nat. Rev. Cancer* **15**, 701–711 (2015).
8. Epstein, J.I. *et al.* Proposed morphologic classification of prostate cancer with neuroendocrine differentiation. *Am. J. Surg. Pathol.* **38**, 756–767 (2014).
9. Beltran, H. *et al.* Molecular characterization of neuroendocrine prostate cancer and identification of new drug targets. *Cancer Discov.* **1**, 487–495 (2011).
10. Yao, J.L. *et al.* Small-cell carcinoma of the prostate: an immunohistochemical study. *Am. J. Surg. Pathol.* **30**, 705–712 (2006).
11. Beltran, H. *et al.* Challenges in recognizing treatment-related neuroendocrine prostate cancer. *J. Clin. Oncol.* **30**, e386–e389 (2012).
12. Aparicio, A. & Tzelepi, V. Neuroendocrine (small-cell) carcinomas: why they teach us essential lessons about prostate cancer. *Oncology* **28**, 831–838 (2014).
13. Lin, D. *et al.* High-fidelity patient-derived xenografts for accelerating prostate cancer discovery and drug development. *Cancer Res.* **74**, 1272–1283 (2014).
14. Wang, H.T. *et al.* Neuroendocrine prostate cancer (NEPC) progressing from conventional prostatic adenocarcinoma: factors associated with time to development of NEPC and survival from NEPC diagnosis—a systematic review and pooled analysis. *J. Clin. Oncol.* **32**, 3383–3390 (2014).
15. Hieronymus, H. *et al.* Gene expression signature-based chemical genomic prediction identifies a novel class of HSP90 pathway modulators. *Cancer Cell* **10**, 321–330 (2006).
16. Chen, H. *et al.* Pathogenesis of prostatic small-cell carcinoma involves the inactivation of the p53 pathway. *Endocr. Relat. Cancer* **19**, 321–331 (2012).
17. Zhou, Z. *et al.* Synergy of p53 and Rb deficiency in a conditional mouse model for metastatic prostate cancer. *Cancer Res.* **66**, 7889–7898 (2006).

18. Tan, H.L. *et al.* Rb loss is characteristic of prostatic small-cell neuroendocrine carcinoma. *Clin. Cancer Res.* **20**, 890–903 (2014).
19. Grasso, C.S. *et al.* The mutational landscape of lethal castration-resistant prostate cancer. *Nature* **487**, 239–243 (2012).
20. Robinson, D. *et al.* Integrative clinical genomics of advanced prostate cancer. *Cell* **161**, 1215–1228 (2015).
21. Hu, R. *et al.* Ligand-independent androgen receptor variants derived from splicing of cryptic exons signify hormone-refractory prostate cancer. *Cancer Res.* **69**, 16–22 (2009).
22. Antonarakis, E. *et al.* AR-V7 and resistance to enzalutamide and abiraterone in prostate cancer. *N. Engl. J. Med.* **371**, 1028–1038 (2014).
23. Baca, S.C. *et al.* Punctuated evolution of prostate cancer genomes. *Cell* **153**, 666–677 (2013).
24. Barbieri, C.E. *et al.* Exome sequencing identifies recurrent *SPOP*, *FOXA1* and *MED12* mutations in prostate cancer. *Nat. Genet.* **44**, 685–689 (2012).
25. Pritchard, C.C. *et al.* Complex *MSH2* and *MSH6* mutations in hypermutated microsatellite-unstable advanced prostate cancer. *Nat. Commun.* **25**, 4988 (2014).
26. Prandi, D. *et al.* Unraveling the clonal hierarchy of somatic genomic aberrations. *Genome Biol.* **15**, 439 (2014).
27. Forbes, S.A. *et al.* COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res.* **43**, D805–D811 (2015).
28. Rajan, N. *et al.* The cylindromatosis gene product, CYLD, interacts with MIB2 to regulate Notch signaling. *Oncotarget* **5**, 12126–12140 (2014).
29. Lim, J.H. *et al.* CYLD negatively regulates transforming growth factor- β -signaling via deubiquitinating Akt. *Nat. Commun.* **3**, 771 (2012).
30. Theurillat, J.P. *et al.* Prostate cancer. Ubiquitylome analysis identifies dysregulation of effector substrates in *SPOP*-mutant prostate cancer. *Science* **346**, 85–89 (2014).
31. Lin, L. *et al.* Mechanisms underlying cancer growth and apoptosis by *DEK* overexpression in colorectal cancer. *PLoS One* **9**, e111260 (2014).
32. Nowell, P.C. The clonal evolution of tumor cell populations. *Science* **194**, 23–28 (1976).
33. Maley, C.C. *et al.* Genetic clonal diversity predicts progression to esophageal adenocarcinoma. *Nat. Genet.* **38**, 468–473 (2006).
34. Hansel, D.E. *et al.* Shared *TP53* gene mutation in morphologically and phenotypically distinct concurrent primary small-cell neuroendocrine carcinoma and adenocarcinoma of the prostate. *Prostate* **69**, 603–609 (2009).
35. Gudem, G. *et al.* The evolutionary history of lethal metastatic prostate cancer. *Nature* **520**, 353–357 (2015).
36. Smith, B.A. *et al.* A basal stem cell signature identifies aggressive prostate cancer phenotypes. *Proc. Natl. Acad. Sci. USA* **112**, E6544–E6552 (2015).
37. Steffan, J.J. *et al.* The transcription factor SPDEF suppresses prostate tumor metastasis. *J. Biol. Chem.* **287**, 29968–29978 (2012).
38. Börno, S.T. *et al.* Genome-wide DNA methylation events in *TPRSS2-ERG* fusion-negative prostate cancers implicate an EZH2-dependent mechanism with miR-26a hypermethylation. *Cancer Discov.* **2**, 1024–1035 (2012).
39. Varambally, S. *et al.* Genomic loss of microRNA-101 leads to overexpression of histone methyltransferase EZH2 in cancer. *Science* **322**, 1695–1699 (2008).
40. Yu, J. *et al.* A polycomb repression signature in metastatic prostate cancer predicts cancer outcome. *Cancer Res.* **67**, 10657–10663 (2007).
41. Karanikolas, B.D., Figueiredo, M.L. & Wu, L. Comprehensive evaluation of the role of EZH2 in the growth, invasion and aggression of a panel of prostate cancer cell lines. *Prostate* **70**, 675–688 (2010).
42. Clermont, P.L. *et al.* Polycomb-mediated silencing in neuroendocrine prostate cancer. *Clin. Epigenetics* **7**, 40 (2015).
43. Cheng, A.S. *et al.* EZH2-mediated concordant repression of Wnt antagonists promotes β -catenin-dependent hepatocarcinogenesis. *Cancer Res.* **71**, 4028–4039 (2011).
44. Kanduri, M. *et al.* A key role for EZH2 in epigenetic silencing of *HOX* genes in mantle-cell lymphoma. *Epigenetics* **8**, 1280–1288 (2013).
45. Akamatsu, S. *et al.* The placental gene *PEG10* promotes progression of neuroendocrine prostate cancer. *Cell Rep.* **12**, 922–936 (2015).
46. Cancer Genome Atlas Research Network. The molecular taxonomy of primary prostate cancer. *Cell* **163**, 1011–1025 (2015).
47. Chakravarty, D. *et al.* The estrogen receptor α -regulated lncRNA *NEAT1* is a critical modulator of prostate cancer. *Nat. Commun.* **5**, 5383 (2014).
48. Carreira, S. *et al.* Tumor clone dynamics in lethal prostate cancer. *Sci. Transl. Med.* **6**, 254ra125 (2014).

ONLINE METHODS

Code availability. The source code for the study clonality analysis is accessible from the Bitbucket version-controlled repository (<https://bitbucket.org/>; CLONET).

Cohort description and pathology classification. Tumor specimens were obtained prospectively through clinical protocols approved by the Weill Cornell Medicine (WCM) Institutional Review Board (IRB) with informed consent (IRB #1305013903 and IRB #1210013164) or retrospectively (IRB #0905010441), and germline (normal) DNA was obtained from either peripheral blood mononuclear cells (PBMCs) or benign tissue. The total number of subjects enrolled in this study was 81, all of male gender; no blinding or randomization was used. Tissues were processed as previously described^{20,49}. All hematoxylin and eosin-stained slides were reviewed by board-certified pathologists (J.M.M. and M.A.R.). Tumors were classified based on histomorphology as adenocarcinoma (A) or CRPC-NE (B–E) based on a published pathologic classification system⁸ (**Supplementary Fig. 1**). Category A represents usual prostate adenocarcinoma without neuroendocrine differentiation; category B represents usual prostate adenocarcinoma with neuroendocrine differentiation > 20%; category C represents small-cell carcinoma; category D represents large-cell neuroendocrine carcinoma; and category E represents mixed small-cell carcinoma–adenocarcinoma. Clinical and pathologic features of the cohort are summarized in **Supplementary Table 1** and **Supplementary Fig. 1**.

Immunohistochemistry. Immunohistochemistry (IHC) was performed using a Bond III automated immunostainer (Leica Microsystems, IL, USA) with the following antibodies and dilutions: anti-AR (MU256-UC, clone F39.4.1, BioGenex, CA, USA; dilution 1:800 with casein), anti-synaptophysin (RM-9111-S, clone SP11, Thermo Scientific; dilution 1:100), anti-chromogranin A (MU126-UC, clone LK2H10, BioGenex, CA, USA; dilution 1:400), anti-CD56 (NCL-SD56-504, clone CD564, Leica Biosystems, IL, USA; dilution 1:50), anti-PSA (MU014-UC, clone ErPr-8, BioGenex, CA, USA; dilution 1:400), anti-PSAP (MU013-UC, clone BGX013A, BioGenex, CA, USA; dilution 1:250), anti-PSMA (M3620, clone 3E6, Dako, CA, USA; dilution 1:100), anti-Ki67 (M7240, clone MIB-1, Dako, CA, USA; dilution 1:50), anti-ERG (ab92513, clone EPR3864, Abcam, MA, USA; dilution 1:100), anti-EZH2 (612667, clone 11/EZH2, BD Biosciences, CA, USA; dilution 1:20), anti-MLH1 (554073, clone G168-728, BD Biosciences, CA, USA; dilution 1:400), anti-PMS2 (556415, clone A16-4, BD Biosciences; dilution 1:100), anti-MSH2 (NA27, clone FE11, Calbiochem, CA, USA; dilution 1:100) and anti-MSH6 (610919, clone 44/MSH6, BD Biosciences, CA, USA; dilution 1:800), using the Bond Polymer Refine detection kit (Leica Microsystems, IL, USA). Antigen retrieval was performed using heat-mediated pH 6 retrieval for anti-ERG, anti-PSA and anti-PSMA; pressure cooker pH 6 retrieval for anti-EZH2; no retrieval for anti-chromogranin A and anti-PSAP; and heat-mediated pH 9 retrieval for all of the other antibodies. Study pathologists performed semi-quantitative evaluation for protein expression in nuclear (AR, ERG, MLH1, MSH2, MSH6 and PMS2), cytoplasmic (synaptophysin and chromogranin A), or both (EZH2) compartments, using a four-tier grading system: negative (0), weakly positive (1+), moderately positive (2+), and strongly positive (3+). The extent of positivity (percentage) was recorded. For evaluation of ERG and MMR protein expression, IHC was defined as either positive or negative. For the other antibodies, the following cut-offs were considered to determine positive expression: >20% of cells for synaptophysin and chromogranin A; >10% of nuclei for AR; and >10% of cells for EZH2.

DNA extraction, tumor purity, and exome sequencing. Slides were cut from frozen or fresh-frozen paraffin-embedded (FFPE) tissue blocks and examined by the study pathologists to select high-density cancer foci and ensure high purity of cancer DNA. We used previously developed protocols⁵⁰ that were successfully used in our earlier genomic studies^{20,23,24,49}. Following this protocol, each case was quantified for tumor purity and annotated for discrete areas of macrodissection, avoiding regions of necrosis or high stromal content. All cases were also quantified for tumor purity using an algorithm called CLONET^{23,26}. CLONET was developed and validated to specifically deal

with heterogeneous tumor samples, enabling optimal objective tumor purity and ploidy estimates by taking the germline heterozygous single-nucleotide polymorphism (SNP) genotype data from whole-exome sequence coverage to quantify the percentage of reads supporting the considered aberration. The resultant tumor purity values were used to adjust the genomic data for downstream processing and analysis.

Extraction and sequencing were performed as previously described^{49,51}. Briefly, DNA was extracted using Promega Maxwell 16 MDx. DNA was stored at –20 °C. Whole-exome capture libraries were constructed from tumor and normal tissue after sample-shearing, end repair, and phosphorylation and ligation to barcoded sequencing adaptors. Ligated DNA was size-selected for lengths between 200 and 350 bp and subjected to either exonic hybrid capture using SureSelect v2/v4 Exome bait (Agilent) or HaloPlex Exome (Agilent) (**Supplementary Tables 11** and **12**). The samples (70 CRPC-Adeno samples from 51 individuals and 44 CRPC-NE from 30 individuals) were multiplexed and sequenced using Illumina HiSeq for an intended mean-target exome coverage of 100× for the tumor and germline samples.

Sequence data processing and quality control. A fraction of study samples ($n = 78$) were preprocessed with the analytical pipeline of the Englander Institute for Precision Medicine at Weill Cornell–New York Presbyterian Hospital (IPM-Exome-pipeline v0.9)⁴⁹. FastQC was run on the raw reads to assess the quality of the raw reads as previously described⁴⁹. The remaining 36 samples were processed using the Broad Institute Firehose infrastructure (<http://www.broadinstitute.org/cancer/cga/Firehose>)²⁰. Cross-contamination between samples from other individual subjects sequenced in the same flow cell was monitored with the ContEst algorithm⁵². Normal-tumor pairs were checked for consistency using SNP panel identification assay (SPIA)⁵³. Processing pipelines returned segmented files for somatic copy-number aberrations (SCNAs) (**Supplementary Table 6**). No differences were observed in the results from the two pipelines ($P = 0.75$, Kolmogorov-Smirnov test).

Mutation-calling and identification of significantly mutated genes. To identify and characterize somatic single-nucleotide variants (SNVs) in targeted exons, we first applied MuTect from the Broad Institute Genome Analysis Toolkit⁵⁴, based on Bayesian statistical analysis to nominate putative SNVs upon coverage, allelic fraction, and base-quality extraction. Next, we used Oncotator⁵⁵ to annotate point mutations with variant- and gene-centric information relevant to cancer, and lastly we used MutSig⁵⁶ to identify genes harboring a greater number of mutations than expected by chance, given the background mutation rate, the sequence context and the genomic territory. For each gene identified (**Supplementary Table 3**), it returns the P value adjusted for multiple hypotheses testing with Benjamini-Hochberg procedure (q -value). Genes with q -values below 0.1 are considered significant. Finally, we searched for mutations enriched in one pathology class by means of binomial test adjusted with Benjamini-Hochberg procedure (q -value) (**Supplementary Table 4**).

Tumor ploidy, purity and copy-number estimates. Segmented data was used by CLONET to estimate ploidy and purity for each tumor sample as previously described²⁶. Each segment is represented by the \log_2 of the ratio between values proportional to the tumor and normal local coverage within the genomic segment. Briefly, the ploidy of an individual tumor, defined as the mean number of sets of chromosomes of a cell, is assessed using the mean coverage observed in an individual as a proxy of the number of observed alleles; CLONET recognizes shifts in the SCNAs \log_2 ratio distribution, which is reflective of an aneuploid genome, and is then used to estimate tumor ploidy. Next, a local optimization approach based on putative clonal mono-allelic deletions and germline heterozygous SNP loci (called ‘informative SNPs’ and identified by means of ASEQ⁵⁷) is applied to assess the purity (1-admixture) of each sample; the difference between observed and expected allelic fraction (AF) at informative SNPs (the latter being either 0 or 1) is proportional to tumor purity. Finally, CLONET computes a purified copy-number profile, adjusting each segment to account for both aneuploidy and tumor purity. Purified segments with mean \log_2 ratio less than –0.4 or greater than 0.4 were categorized as copy-number loss or gain, respectively.

Allele-specific copy-number analysis by CLONET. To determine the copy-number landscape of a tumor sample, allele-specific copy-number analysis was applied within the CLONET framework. In a 100% pure tumor sample, the empirical distribution of the allelic fractions (AFs) within a genomic segment reflects the aggregated signal from multiple cell populations. If a segment *S* is copy-number neutral and both alleles are represented, sequenced DNA fragments equally sample the reference and the alternative bases, i.e., the AF is around 0.5. If a segment *S* represents a 100% clonal mono-allelic deletion, the AF is either 0 or 1 depending on which allele is lost. Combinations of diverse representations of the two alleles of a diploid genome and subclonality lead to AFs in between. Low DNA purity further dilutes the signal from the expected values. CLONET uses a variable beta that represents the disproportion in the AF of informative SNPs within a segment *S*; beta values can be computed from the empirical distribution of the AFs. Importantly, the beta value of a genomic segment is independent from its \log_2 ratio, and the two measures can be used to infer the allele-specific copy number and the clonality state of each segment of a tumor genome. CLONET also provides a space transformation from the beta versus \log_2 ratio to the cnB versus cnA (**Supplementary Note and Supplementary Fig. 14**), where each segment is visualized at coordinates representing the number of copies of allele A (cnA) and allele B (cnB). In the absence of parental allelic information, we consider $cnA \geq cnB$. Noninteger values of cnA and/or cnB indicate that the copy-number signal results from the convolution of the copy-number states of multiple subclonal tumor cell populations. Finally, the cnB versus cnA space allows for direct comparison of allele-specific copy-number profiles of different tumor samples by mean of the Euclidean distance, irrespective of tumor ploidy and tumor purity.

RNA extraction, sequencing and processing. RNA was extracted from frozen material for RNA-sequencing (RNA-seq) using Promega Maxwell 16 MDx instrument, (Maxwell 16 LEV simplyRNA Tissue Kit (cat. # AS1280)). Specimens (34 CRPC-Adeno samples from 33 individuals and 15 CRPC-NE samples from 10 individuals) were prepared for RNA sequencing using TruSeq RNA Library Preparation Kit v2 as previously described^{9,47}. RNA integrity was verified using the Agilent Bioanalyzer 2100 (Agilent Technologies). cDNA was synthesized from total RNA using Superscript III (Invitrogen). Each sample was then sequenced with the HiSeq 2500 to generate 2×75 -bp paired-end reads. Details of the sequencing results are reported in **Supplementary Table 13**. Reads (FASTQ files) were mapped to the human genome reference sequence (hg19/GRC37; <http://hgdownload.cse.ucsc.edu/downloads.html#human>; major haplotypes only) using STAR v2.3.0e (ref. 58), and the resulting BAM files were subsequently converted into mapped-read format (MRF) using RSEQtools, a suite of tools for RNA-seq data processing and analysis⁵⁹. MRF files include only the primary alignments as determined by STAR and do not include reads mapped to the mitochondrial chromosome. Quantification of gene expression was performed via RSEQtools using GENCODE v19 (<http://www.genecodegenes.org/releases/19.html>) as reference gene-annotation set. A composite model of genes based on the union of all exonic regions from all transcripts of a gene was used, resulting in a set of 20,345 protein-coding genes. Expression levels (FPKM) were estimated by counting all nucleotides mapped to each gene and were normalized by the total number of mapped nucleotides (per million) and the gene length (per kb). Differential expression analysis was performed using the Mann-Whitney Wilcoxon test after transforming the FPKMs via $\log_2(\text{FPKM} + 1)$. Multiple-hypothesis testing was considered by using Benjamini-Hochberg (BH; FDR) correction. For downstream analyses (differential copy number and integrated NEPC score), genes with low expression in both CRPC-Adeno and CRPC-NE samples ((mean + s.d.) < 1 FPKM) were discarded.

NanoString assay. We employed a custom NanoString assay for cases without sufficient material for RNA-seq to evaluate the expression of AR-signaling genes, *TMPRSS2-ERG* fusion transcript, and neuroendocrine-associated genes. FFPE samples were cut into 10- μm -thick slides, annotated by the study pathologist, and RNA was extracted using the Ambion RecoverAll Total Nucleic Acid Isolation Kit. RNA quality control was performed on the Agilent 2100 Bioanalyzer system by annotating the total RNA concentration and the percentage of RNA >300 nucleotides (nt) in length. For samples with more

than 50% of total RNA >300 nt, 100 ng input RNA was used; for samples with less than 50% of total RNA >300 nt, the input RNA was proportionally increased according to the level of degradation. Samples were run on the NanoString nCounter Analysis System according to the manufacturer's directions. Briefly, total RNA was hybridized overnight at 65 °C, then run on the Prep Station at maximum sensitivity. Cartridges were then scanned on the Digital Analyzer at 555 fields of view. Raw count data was normalized using the nSolver analysis software version 2.0, which normalizes samples according to positive- and negative-control probes and the geometric mean of the six housekeeping primers.

Differential copy-number analysis. For each gene in a panel of more than 19,000 RefSeq genes, we computed the \log_2 ratio adjusted by ploidy and tumor purity, and then we aggregated the mutation frequencies by specimen (**Supplementary Table 7**). Deletions and amplifications were defined using the thresholds on the \log_2 ratio as described above. The relative frequencies of deletions and amplifications in CRPC-NE and CRPC-Adeno specimens are compared by means of binomial test adjusted for multiple-hypotheses testing with the Benjamini-Hochberg procedure (FDR). For each gene, **Supplementary Table 7** also reports the mean expression level in CRPC-NE and CRPC-Adeno samples together with the Mann-Whitney Wilcoxon *P* value and FDR correction for the likelihood that the expression in the classes is different.

Androgen-receptor signaling. We assessed AR signaling using the expression levels of 30 genes (**Supplementary Table 2**) that were previously reported as defining the pathway¹⁵. For each specimen with expression levels available, either from RNA-seq or Nanostring analysis, we considered its correlation to a reference sample known to have active AR signaling. This reference sample was based on LNCaP cells and was generated by taking the average values of the 30 AR-regulated genes across three replicates on the Nanostring assay. We then computed the Pearson's correlation coefficient for each specimen to this reference sample and considered this as the 'AR signaling'. To validate the approach, we tested prostate cancer cell lines with known AR activity (**Supplementary Fig. 15**).

Fusion detection and ERG rearrangement status. To detect *ERG* rearrangement status we used several assays, including fluorescence *in situ* hybridization (FISH) break-apart assay and/or immunohistochemistry (IHC) as previously described⁶⁰. Whole-exome sequencing copy-number analysis (WES) and Nanostring-RNA-sequencing (RNA-seq) were also used to assess *ERG* deletion and fusion transcript levels, respectively. If *ERG* fusion was detected at the DNA, mRNA or protein level, the sample was considered positive. For other noncanonical and canonical fusion transcript detection, FusionSeq was used as previously described⁶¹.

CYLD fluorescence *in situ* hybridization (FISH). To assess *CYLD* deletion in tissues, we developed a dual-color FISH assay consisting of a locus-specific probe (RP11-327F22) plus a reference probe spanning a stable region of the chromosome (RP11-488I20). All clones were tested on metaphase spreads^{9,62}. *CYLD* deletion was defined by the absence at least one copy, on average, per nucleus, as compared to two reference signals. At least 100 nuclei were evaluated per tissue section using a fluorescence microscope (Olympus BX51; Olympus Optical, Tokyo, Japan).

Methylation profiling and data processing. Sample preparation for enhanced reduced representation bisulfite sequencing (eRRBS) was performed at the Weill Cornell Medicine Epigenomics Core Facility as previously described⁶³ and included 18 CRPC-Adeno samples from 10 individuals and 10 CRPC-NE samples from 8 individuals. In brief, the preparation steps included: 1) MspI enzyme digestion; 2) end repair of digested DNA; 3) adenylation; 4) adaptor ligation, with pre-annealed 5-methylcytosine-containing Illumina adapters; 5) isolation of library fragments of 150 to 400 bp from a 1.5% agarose gel; 6) bisulfite conversion using the EZ DNA Methylation Kit (Zymo Research, Irvine, CA); 7) polymerase chain reaction (PCR) amplification; each library was prepared with FastStart High Fidelity DNA Polymerase (Roche,

Indianapolis, IN) and Illumina PCR primers PE1.0 and 2.0. PCR products were isolated using Agencourt AMPure XP (Beckman Coulter, Brea, CA) beads per the manufacturer's protocol (Agencourt). Amplified libraries were evaluated using a Qubit 1.0 fluorometer and Quant-iT dsDNA HS Assay Kit (Invitrogen, Grand Island, NY) for quantitation and bioanalyzer visualization (Agilent 2100 Bioanalyzer; Agilent, Santa Clara, CA). After filtering out genomic regions of somatic homozygous deletions (\log_2 ratio < -1.1), as inferred from the corresponding WES data (Supplementary Table 6), differentially methylated sites (CRPC-NE versus CRPC-Adeno) were identified by methylKit⁶⁴. Differentially methylated genes were identified by annotating, using BedTools⁶⁵, differentially methylated regions on gene promoters (defined as 2-kb genomic regions upstream of the set of 5' gene coordinates), first exons, gene bodies, CpG islands (using the University of California Santa Cruz (UCSC) table browser) and CpG shores (2-kb genomic regions upstream and downstream of CpG islands). GENCODE v19 was used as a gene set. ToppFun⁶⁵ was used to perform functional enrichment analysis. Supplementary Table 8 lists differentially methylated genes and includes genomic feature annotations.

Site-directed CpG methylation. Targeted CpG methylation analysis for *SPDEF* was performed using OneStep qMethyl PCR Kit per the manufacturer's protocol (Zymo Research, Irvine, CA). 20 ng of genomic DNA from cell lines was used for qMethyl PCR. Primers used for qMethyl PCR were: Primer1 forward: 5'-CCGGTGACATCCGTGTGTTTC-3', Primer1 reverse: 5'-AATCGCCGGTACTACTCTTG-3', Primer2 forward: 5'-GATTCTGCTCTCCACCTTC-3', Primer2 reverse: 5'-CCAGCAGCCCTCAAA GCAAC-3'. Amplification parameters were: 45 cycles (denaturation: 95 °C, 30 s; annealing: 64 °C, 30 s; extension: 72 °C, 30 s).

Integrated NEPC score. The integrated neuroendocrine prostate cancer (NEPC) score estimates the likelihood of a test sample to be CRPC-NE. It is calculated as the Pearson's correlation coefficient between the test vector and a reference CRPC-NE vector based on a set of 70 genes (Supplementary Table 9 and Supplementary Figs. 10 and 15), using normalized FPKM values of the test sample. The gene set stems from the integration of differentially deleted or amplified and/or expressed and/or methylated genes in CRPC-NE and CRPC-Adeno. Specifically, 16 differentially deleted genes were selected from among putative cancer genes²⁷ (see section 'Differential copy-number analysis'). The following strategy was used to identify both differentially expressed genes that better distinguish CRPC-NE and CRPC-Adeno samples. We selected differentially expressed protein-coding genes with $FDR \leq 10^{-2}$, resulting in a total of 2,425 genes, which correspond to 1,301 overexpressed and 1,124 underexpressed genes. For each gene, we performed a 'receiver-operator curve' (ROC) analysis using the normalized FPKMs as a threshold parameter and calculated the 'area under the curve' (AUC). ROCs were built by considering only samples sequenced, excluding two samples (7,520 and 4,240) that were previously published⁹, leaving 34 CRPC-Adeno and 13 CRPC-NE samples. Only those differentially expressed genes with an $AUC \geq 0.95$ and with a 2 < fold change < 0.5 were included in the classifier, resulting in a list of 49 genes (25 overexpressed and 24 underexpressed in CRPC-NE versus CRPC-Adeno), 21 of which were found as differentially methylated between the CRPC-NE and CRPC-Adeno samples. Concordant information between RNA and methylation was found for 11 genes (see Supplementary Table 9). In addition, we considered two genes (*MYCN* and *AURKA*) that we previously described as associated with a CRPC-NE phenotype⁹, *EZH2* ($FDR = 7.9 \times 10^{-4}$) and *DNMT1* ($FDR = 6.9 \times 10^{-5}$) for their role in controlling DNA methylation⁶⁷, and *RB1* ($FDR = 0.056$), which was reported as a key driver in the pathogenesis of CRPC-NE¹⁶⁻¹⁸. For each of the resulting 70 genes, we calculated the mean of the normalized FPKM across the 13 CRPC-NE samples with RNA-seq data and defined the resulting set of averages as reference CRPC-NE vector. The integrated NEPC score was tested across 719 prostate samples with available transcriptome data from multiple data sets (Supplementary Table 10). RNA-seq data were processed as described above. Processed SU2C/PCF 2015 (ref. 20) and Grasso *et al.*¹⁹ (Michigan 2012) data were downloaded from cBioPortal⁶⁸. Because data for four genes (*ARHGAP8*, *BRINP1*, *C7orf76* and *MAP10*) were not available from cBioPortal, for Michigan 2012 we used a reduced version of the integrated NEPC score (indicated as 'integrated NEPC

score'). Samples with integrated NEPC scores ≥ 0.40 ('elevated integrated NEPC score' in the main text) were nominated as putative CRPC-NE tumors (Fig. 4c and Supplementary Table 14). In order to take into account the lower signal-to-noise ratio and the reduced version of the integrated NEPC score in the Michigan 2012 microarray data, in Figure 4d we consider those samples with an integrated NEPC Score ≥ 0.25 ('significant integrated NEPC score' in Fig. 4 legend) to be 'CRPC-NE-like'. AR signaling and integrated NEPC score values for each sample are reported in Supplementary Table 15.

Cell line studies. LNCaP clone FGC, NCI-H660 cells and medium were purchased from the American Type Culture Collection (ATCC), and the cells were cultured at 37 °C in 5% CO₂. Cell line authentication was performed (DDC Medical, Fairfield, OH), and cells were tested for mycoplasma contamination. To create a resistant cell line, the AR-dependent LNCaP line was grown in medium containing 1 μ m enzalutamide for approximately 6 months. A dose-response curve was made by plating 2.5×10^3 cells in 10 μ l Matrigel (Corning) onto an Ibidi 96-well microplate and covering them with 40 μ l of medium containing increasing doses of enzalutamide for 10 d (medium changed every 4 d). Viability was analyzed using the CellTiter-Blue assay (Promega); western blot analysis was performed with 50 μ g protein, and staining was done with 1:1,000 anti-NCAM (Abcam, ab137086), 1:500 anti-SPDEF (Biorbyt orb13642), and 1:10,000 anti-GAPDH (Millipore, AB2302) antibodies. Site-directed CpG methylation of *SPDEF* was performed as described above. Cells were blocked in 5% BSA in PBS for 1 hour on ice, then stained with 5 μ g/ml Brilliant Violet 421-conjugated anti-human CD56 (Biolegend) for 1 hour on ice in the dark. Cells were washed twice with 8 ml PBS (8 min \times 500g) and resuspended in FACS sorting buffer (1 \times PBS, 1 mM EDTA, 25 mM HEPES pH 7.0, 1% FCS (heat inactivated), 1% pluronic). 0.5 μ g/ml DAPI was added immediately before sorting to stain dead cells. Cells were sorted on a BD FACSCalibur under the guidance of the Weill Cornell Flow Cytometry core and analyzed with FloJo vX.0.7.

CYLD expression was silenced using ON-TARGET plus *CYLD*-specific short interfering RNA (siRNA) (Dharmacon L-004609-00-0005). ON-TARGET plus non-targeting pool (Dharmacon, D-001810-10-05) was used as a control. RNAs from LNCaP cells were purified using Maxwell 16 LEV simplyRNA Cells Kit (Promega) according to the manufacturer's instructions. cDNA was generated using the qScript cDNA SuperMix (Quanta BioSciences). SYBR Green-based qRT-PCR experiments were performed on a Roche LightCycler 480 II sequence detection system using Roche SYBR. The following oligonucleotides were used: Human *CYLD*: 5'-TTTTCGTGTGTTGAAAGTACAAT-3' (forward), 5'-TTCTGCGTACACTCTCTG-3' (reverse). Human *ACTB* (encoding β -actin): 5'-TCCCTGGAGAAGAGCTACG-3' (forward) 5'-GTAGTTTCGTGG ATGCCACA-3' (reverse). *CYLD* relative expression was normalized to expression of *ACTB*.





LNCaP, NCI-H660, and DU145 cells and medium were purchased from ATCC and used for GSK343 drug-treatment studies using escalating doses (5, 7.5, 10 μ M). Cells were plated in Ibidi 96-well microplates embedded in 5 μ l Matrigel. Matrigel droplets were then covered with 40 μ l medium. After 48 h, medium was replaced with fresh medium containing increasing doses of GSK343 for 14 d (medium changed every 4 d) in three replicates. Viability was analyzed using the CellTiter-Blue assay (Promega) according to the manufacturer's protocol. Total RNA was isolated from the cells using the RNeasy Mini Kit (Qiagen, Valencia, CA) and NanoString analysis was performed for quantitative analysis of mRNA expression of the reported genes.

Statistical analysis. For statistical tests, we used the two-sided Mann-Whitney-Wilcoxon test (also referred to as Wilcoxon test in the main text) to check for significant shifts between two distributions. When appropriate, we used Kolmogorov-Smirnov test to check for discrepancies in the compared distributions. The proportion test has been used to determine whether the deviations between the observed and the expected counts are significant. Finally, Supplementary Figure 5e uses a Student's *t*-test. The statistical test used is indicated in the respective figure legend or in the corresponding main text. All the tests are two sided. When appropriate, *P* values were adjusted for multiple-hypotheses testing with the Benjamin-Hochberg procedure. No statistical methods were used to predetermine sample size.

49. Beltran, H. *et al.* Whole-exome sequencing of metastatic cancer and biomarkers of treatment response. *JAMA Oncol.* **1**, 466–474 (2015).
50. Esgueva, R. *et al.* Next-generation prostate cancer biobanking: toward a processing protocol amenable for the International Cancer Genome Consortium. *Diagn. Mol. Pathol.* **21**, 61–68 (2012).
51. Van Allen, E.M. *et al.* Whole-exome sequencing and clinical interpretation of formalin-fixed, paraffin-embedded tumor samples to guide precision cancer medicine. *Nat. Med.* **20**, 682–688 (2014).
52. Cibulskis, K. *et al.* ContEst: estimating cross-contamination of human samples in next-generation sequencing data. *Bioinformatics* **27**, 2601–2602 (2011).
53. Demichelis, F. *et al.* SNP panel identification assay (SPIA): a genetic-based assay for the identification of cell lines. *Nucleic Acids Res.* **36**, 2446–2456 (2008).
54. Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–219 (2013).
55. Ramos, A.H. *et al.* Oncotator: cancer variant annotation tool. *Hum. Mutat.* **36**, E2423–E2429 (2015).
56. Lawrence, M.S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).
57. Romanel, A., Lago, S., Prandi, D., Sboner, A. & Demichelis, F. ASEQ: fast allele-specific studies from next-generation sequencing data. *BMC Med. Genomics* **8**, 9 (2015).
58. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
59. Habegger, L. *et al.* RSEQtools: a modular framework to analyze RNA-seq data using compact, anonymized data summaries. *Bioinformatics* **27**, 281–283 (2011).
60. Park, K. *et al.* Antibody-based detection of *ERG* rearrangement-positive prostate cancer. *Neoplasia* **12**, 590–598 (2010).
61. Sboner, A. *et al.* FusionSeq: a modular framework for finding gene fusions by analyzing paired-end RNA-sequencing data. *Genome Biol.* **11**, R104 (2010).
62. Mosquera, J.M. *et al.* Concurrent *AURKA* and *MYCN* gene amplifications are harbingers of lethal treatment-related neuroendocrine prostate cancer. *Neoplasia* **15**, 1–10 (2013).
63. Lin, P.C. *et al.* Epigenomic alterations in localized and advanced prostate cancer. *Neoplasia* **15**, 373–383 (2013).
64. Akalin, A. *et al.* methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* **13**, R87 (2012).
65. Dale, R.K., Pedersen, B.S. & Quinlan, A.R. Pybedtools: a flexible Python library for manipulating genomic datasets and annotations. *Bioinformatics* **27**, 3423–3424 (2011).
66. Chen, J., Bardes, E.E., Aronow, B.J. & Jegga, A.G. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res.* **37**, W305–W311 (2009).
67. Viré, E. *et al.* The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* **439**, 871–874 (2006).
68. Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).

Article

Small Cell Carcinoma of the Ovary, Hypercalcemic Type (SCCOHT) beyond *SMARCA4* Mutations: A Comprehensive Genomic Analysis

Aurélie Auguste ¹, Félix Blanc-Durand ², Marc Deloger ³, Audrey Le Formal ¹, Rohan Bareja ^{4,5}, David C. Wilkes ⁴, Catherine Richon ⁶, Béatrice Brunn ², Olivier Caron ⁶, Mojgan Devouassoux-Shisheboran ⁷, Sébastien Gouy ², Philippe Morice ², Enrica Bentivegna ², Andrea Sboner ^{4,5,8}, Olivier Elemento ^{4,8}, Mark A. Rubin ⁹, Patricia Pautier ², Catherine Genestie ¹⁰, Joanna Cyrta ^{4,9,11} and Alexandra Leary ^{1,2,*}

¹ Medical Oncologist, Gynecology Unit, Lead Translational Research Team, INSERM U981, Gustave Roussy, 94805 Villejuif, France; Aurelie.auguste@gustaveroussy.fr (A.A.); Audrey.leformal@gustaveroussy.fr (A.L.F.)

² Gynecological Unit, Department of Medicine, Gustave Roussy, 94805 Villejuif, France; felix.blancdurand@gmail.com (F.B.-D.); beatrice.brunn@gustaveroussy.fr (B.B.); sebastien.gouy@gustaveroussy.fr (S.G.); philippe.morice@gustaveroussy.fr (P.M.); enrica.bentivegna@gustaveroussy.fr (E.B.); patricia.pautier@gustaveroussy.fr (P.P.)

³ Bioinformatics Core Facility, Gustave Roussy Cancer Center, UMS CNRS 3655/INSERM 23 AMMICA, 94805 Villejuif, France; marc.deloger@gustaveroussy.fr

⁴ Caryl and Israel Englander Institute for Precision Medicine, Weill Cornell Medicine, New York, NY 10001, USA; rob2056@med.cornell.edu (R.B.); dcw2001@med.cornell.edu (D.C.W.); ans2077@med.cornell.edu (A.S.); ole2001@med.cornell.edu (O.E.); joanna.cyrta@gmail.com (J.C.)

⁵ Institute for Computational Biomedicine, Weill Cornell Medicine, New York, NY 10001, USA

⁶ Genomic Platform Gustave Roussy Cancer Institute, 94805 Villejuif, France; Catherine.RICHON@gustaveroussy.fr (C.R.); olivier.caron@gustaveroussy.fr (O.C.)

⁷ Department of Pathology, Hospital de la Croix Rousse, 69000 Lyon, France; mojgan.devouassoux@chu-lyon.fr

⁸ Department of Pathology and Laboratory Medicine, Weill Cornell Medicine, New York, NY 10001, USA

⁹ Department for BioMedical Research, University of Bern, 3001 Bern, Switzerland; mark.rubin@dbmr.unibe.ch

¹⁰ Department of Pathology, Gustave Roussy, 94805 Villejuif, France; catherine.genestie@gustaveroussy.fr

¹¹ Department of Pathology, Institut Curie, Université Paris Sciences et Lettres, 6 rue d'Ulm, 75005 Paris, France

* Correspondence: Alexandra.leary@gustaveroussy.fr; Tel.: +33-1-42-11-45-71; Fax: +33-1-42-11-52-14

Received: 29 March 2020; Accepted: 11 June 2020; Published: 19 June 2020



Abstract: Small cell carcinoma of the ovary, hypercalcemic type (SCCOHT) is an aggressive malignancy that occurs in young women, is characterized by recurrent loss-of-function mutations in the *SMARCA4* gene, and for which effective treatments options are lacking. The aim of this study was to broaden the knowledge on this rare malignancy by reporting a comprehensive molecular analysis of an independent cohort of SCCOHT cases. We conducted Whole Exome Sequencing in six SCCOHT, and RNA-sequencing and array comparative genomic hybridization in eight SCCOHT. Additional immunohistochemical, Sanger sequencing and functional data are also provided. SCCOHTs showed remarkable genomic stability, with diploid profiles and low mutation load (mean, 5.43 mutations/Mb), including in the three chemotherapy-exposed tumors. All but one SCCOHT cases exhibited 19p13.2-3 copy-neutral LOH. *SMARCA4* deleterious mutations were recurrent and accompanied by loss of expression of the *SMARCA2* paralog. Variants in a few other genes located in 19p13.2-3 (e.g., *PLK5*) were detected. Putative therapeutic targets, including *MAGEA4*, *AURKB* and *CLDN6*, were found to be overexpressed in SCCOHT by RNA-seq as compared to benign ovarian tissue. Lastly, we provide additional evidence for sensitivity of SCCOHT to HDAC, DNMT and EZH2 inhibitors. Despite their aggressive clinical course, SCCOHT show remarkable inter-tumor homogeneity and display genomic

stability, low mutation burden and few somatic copy number alterations. These findings and preliminary functional data support further exploration of epigenetic therapies in this lethal disease.

Keywords: ovary; small cell carcinoma; hypercalcemic; SMARCA4; SWI/SNF

1. Introduction

Small cell carcinoma of the ovary of the hypercalcemic type (SCCOHT) is a rare, highly aggressive tumor that affects mainly young women (median age: 24 years). Prognosis is poor, as most patients die within two years of diagnosis [1]. The histogenesis of SCCOHT remains unclear, although there is increasing evidence in favor of a germ cell origin [2,3]. In addition, it has been proposed that SCCOHT may represent the ovarian counterpart of malignant rhabdoid tumors [4].

While there is no international consensus regarding the optimal treatment of SCCOHT, it usually involves multimodal chemotherapy, radical surgery and possibly, radiotherapy [5]. However, no randomized studies have been conducted to date and the available data consist of case reports or small retrospective series with heterogeneous management strategies. The only prospective clinical study in SCCOHT, a multicenter phase II trial conducted at Institut Gustave Roussy, tested combination chemotherapy (PAVEP: cisplatin, adriamycin, vepeside and cyclophosphamide) followed by radical surgery and high dose chemotherapy with autologous stem cell transplant, and demonstrated a three year survival rate of 49% among 27 SCCOHT patients [6]. This shows that even with intensive regimens, prognosis remains dismal, and that despite frequent initial response to chemotherapy, relapses are almost inevitable and tend to be refractory to second line chemotherapy.

The literature describing the genomic features of SCCOHT was scarce until 2014, when four groups identified loss-of-function mutations in *SMARCA4* (Brahma-related gene 1, BRG1) as a highly recurrent event in SCCOHT [3,4,7,8]. *SMARCA4* encodes one of the two possible catalytic subunits of the Switch/Sucrose Non-Fermentable (SWI/SNF) chromatin-remodeling complex. Others have since confirmed this finding, with *SMARCA4* mutations being found in over 90% of cases [9]. Nevertheless, effective treatment options to target this rare and lethal disease are still lacking.

We aimed to conduct an integrated genomic analysis of an independent cohort of SCCOHT by WES, RNA-Seq and aCGH to check for the presence of additional recurrent genomic alterations, which could allow the proposal of alternative treatment strategies.

2. Materials and Methods

2.1. Patients and Samples

Fresh-frozen tumor samples from 8 patients with SCCOHT were identified from the tumor banks of Institut Gustave Roussy, Cochin University Hospital, Grenoble University Hospital, Longjumeau University Hospital and Hôpital de la Croix Rousse. Central review for histological diagnosis was conducted by an expert pathologist. Matched blood samples were available for 6 patients. All patients provided written informed consent allowing the use of their tumor and non-tumor tissues for research. Approval from the hospital's institutional review board was obtained for the study and funding was obtained via an educational grant awarded by the Foundation Gustave Roussy (local IRB approval RT12014). In addition, a further 33 formalin-fixed paraffin-embedded (FFPE) SCCOHT samples were available for validation studies. Tumors were obtained with patient consent and all data were anonymized.

2.2. DNA Extraction

DNA was extracted from fresh-frozen tumors and matched blood using the AllPrep DNA Mini Kit (Qiagen, Valencia, CA, USA) according to the manufacturer's instructions, and quantified using

Qubit (Thermo Fisher Scientific, Waltham, MA, USA). DNA integrity was measured using an Agilent BioAnalyzer (Agilent, Santa Clara, CA, USA).

2.3. Whole Exome Sequencing (WES)

Exome capture and library preparation were performed using the Sure Select Human All Exome v5 and SureSelectXT kits, respectively (Illumina, Agilent Technologies, CA, USA). Sequencing was done on matched tumor and normal samples using HiSeq2000 (Illumina, San Diego, CA, USA) in paired-end mode with a mean target depth of 100X. Reads were mapped using BWA-MEM (V0.7.5a-r405) [10] against reference genome hg19. Analysis of coverage was done using GATK (2.7.4-g6f46d11) [11] Depth of Coverage. Local realignment was performed around indels using GATK Realigner Target Creator and GATK Indel Realigner.

Variants were called with VarScan 2 [12], using hg19 as the reference genome and requiring a minimum tumor read depth of 6, a minimum somatic read depth of 8 and a minimum tumor allelic frequency of 0.10. Results were then annotated using SnpEff (4.3t) [13] and SnpSift (4.3t) [14] with dbSNP (v150_hg19) (<http://www.ncbi.nlm.nih.gov/SNP/>) and dbNSFP (v2.9) [15].

Tumor mutation burden (TMB) was calculated based on the number of non-synonymous, somatic-only mutations (single nucleotide variants and small insertions/deletions) with a somatic p-value threshold at <0.05 per megabase in coding regions considered as having sufficient coverage (6× in tumors and 8× in matched normal samples) by the variant caller [16].

Specific germline mutation analysis could not be performed, because a clause pertaining to germline testing was not included in the consent form at the time when each patient's consent was obtained.

2.4. Oligonucleotide CGH Microarrays

DNA was labeled and hybridized, and CGH microarray analysis was performed as detailed in Supplementary Methods. Resulting log₂ (ratio) values were segmented using the CBS [17] algorithm implementation from the DNA copy package for R. Aberration status calling was automatically performed for each profile according to its internal noise (absolute variation of log₂ (ratio) values across consecutive probes on the genome). All genomic coordinates were established on the UCSC Homo sapiens genome build hg19 [18].

2.5. SMARCA2 Promoter Sequencing

Sanger sequencing of SMARCA2 promoter polymorphism sites was performed on DNA from 8 fresh-frozen SCCOHT tumor samples and 2 cell lines (BIN67 and SCCOHT-1). The following primers were used: for the -741 site, Forward—TTTGAAGCTTGCAGTCCTT, Reverse—CCGGCTGAAACTTTTTCTCC; for the -1321 site, Forward—CCCAGTTGCTCAAATGGAGT, Reverse—AGGTCGGTGTGGTGAGAC. After PCR, 10 uL from a 50 uL reaction were run on a 2% agarose gel to confirm amplification. The remaining PCR reaction was purified using the Qiagen QIAquick PCR Purification Kit, quantified and 10 ng together with 25 pmol of either the Forward or the Reverse primer were submitted to Genewiz (USA) for Sanger sequencing.

2.6. RNA Sequencing (RNA-Seq), Real-Time RT-PCR and Differential Expression Analysis

RNA-seq was performed on RNA from 8 fresh-frozen SCCOHT tumors, on a HiSeq2000 sequencer, using paired-end 2 × 76 bp stranded mode. Raw reads were mapped against human genome (hg19) with the STAR (v2.3.0) 2-pass method [19] and potential duplicates were marked using Picard tools (<http://picard.sourceforge.net/>). Remaining reads were split into exon segments and STAR mapping qualities were reassigned in order to fit GATK (v3.2-2) Indel Realignment requirements [11,20,21]. After local realignment around indels, a base quality score recalibration (BQSR) process was applied, and the variant calling step was done with HaplotypeCaller in RNA-seq mode. Finally, the raw variants

list obtained above was filtered on a Phred-scaled p-value using Fisher's exact test to detect strand bias ($FS > 30.0$) and Variant Confidence/Quality by Depth ($QD < 2.0$) values. RNA-seq data were also used to identify variants following Broad Institute Best Practices.

Differential RNA-seq gene expression analysis between SCCOHT samples and benign ovarian tissue from the GTEx dataset was performed using rank-normalized expression values and is detailed in Supplementary Methods.

Quantitative real-time polymerase chain reaction (real-time RT-PCR) analyses to assess the expression levels of *SMARCA2* were performed on RNA from 8 fresh-frozen SCCOHT tumors and from the BIN-67 cell line, as detailed in Supplementary Methods.

2.7. Immunohistochemistry

SMARCA4 (BRG1) and *SMARCA2* (BRM) protein expression was assessed by immunohistochemistry (IHC) using the anti-BRG1 (Santa Cruz, sc-10768) and anti-BRM (Abcam, ab15597) antibodies at dilutions of 1/200 and 1/50, respectively. After paraffin removal and hydration, slides were immersed in 10 mM citrate buffer pH 6 for 30 min for antigen retrieval, incubated with primary antibody for one hour at room temperature, washed and incubated with biotinylated secondary antibody for 30 min at room temperature. Streptavidin-biotin amplification (VECTASTAIN Elite ABC Kit) was then performed for 30 min, followed by peroxidase/diaminobenzidine substrate chromogenic reaction. IHC for SOX2 was performed using a Bond III automated immunostainer and the Bond Polymer Refine detection system (Leica Microsystems, IL, USA). Slides were deparaffinized and heat-mediated antigen retrieval was performed using the Bond Epitope Retrieval 2 solution at pH 9 (H2). The anti-SOX2 antibody clone D6D9 (Cell Signaling Technology) was used at 1/100 dilution.

2.8. Cell Culture and Viability Assays

These methods are available in the Supplementary Methods section.

3. Results

3.1. Clinical Data and Mutational Profiles of SCCOHT: A General Overview

The available clinical data are summarized in Supplementary Table S1. The cohort comprised eight patients with a mean age at diagnosis of 31 years (range, 14–40), all of whom were diagnosed with stage III–IV tumors (Figure 1A). Three patients had received chemotherapy prior to sample collection. After a median follow-up of 10 months (range, 3–36 months), seven patients died of disease, while one patient achieved remission and remained disease-free at 36 months follow-up. Histomorphology was reviewed by an expert pathologist (C.G.) and was confirmed to be consistent with SCCOHT for all cases (Figure 1B, Supplementary Figure S1).

WES was conducted on six tumor-normal pairs. The mean depth of coverage was 109X, with at least 98% of the targeted exome covered by at least 10 reads and 95% showing a read quality score (QC) ≥ 30 (Supplementary Table S2).

The tumor mutation burden (TMB), calculated as specified in the Methods section, was low, with a median of 5.60 mutations/Mb (mean, 5.43 mutations/Mb; range: 3.56–6.42). Very few genes showed somatic-only mutations in more than one sample (Supplementary Table S3). These included: *SMARCA4* (three cases, variants in coding regions were predicted to be deleterious and detailed hereafter); *HMCN2* (three cases, missense variants were predicted as benign by Polyphen-2); *ADGRV1* (two cases—one stop gain and one missense variant—both heterozygous), *FANCD2* (two cases, splice region variant predicted to be of low functional impact by the SnpEff tool), and *LRRK2* (two cases, intronic variants).

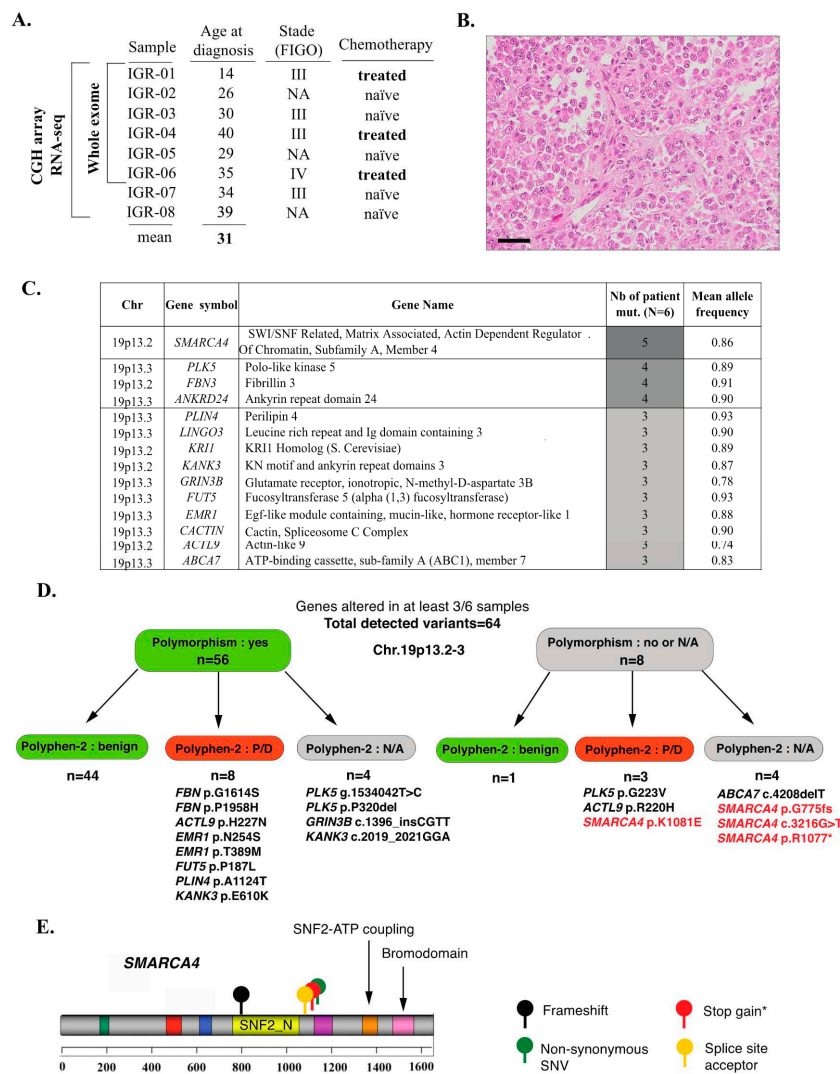


Figure 1. An overview of mutational profiles of SCCOHT. (A) Clinical characteristics of the cohort and tests performed. (B) Representative histopathology of a SCCOHT case from this cohort (IGR-04), including rhabdoid features; hematoxylin-eosin-saffron, scale bar: 50 μ m. (C) Combined analysis of somatic-only and LOH-related alterations: an overview of the 14 genes altered in at least 50% of samples. (D) Breakdown of variants detected in the 14 recurrently altered genes, including classification as known polymorphisms (Genome Aggregation Database v.2.1.1) and Polyphen-2 functional prediction scores. N/A: not available. P/D: possibly or probably damaging. The variants for which functional impact cannot be ruled out are explicitly listed. (E) Type and localization of the mutations found by WES in the *SMARCA4* gene; * indicates that this identical mutation was found in two independent patients.

In order to also account for alterations that may be related to an LOH event, we performed a second analysis using the following criteria: mutant allele frequency higher in tumor than in normal tissue; somatic *p*-value < 0.001 (Fisher’s exact test); and location in a coding region. Using these cut-offs, 500 variants in 335 genes were retained. Among those, fourteen genes were altered in at least 50% of cases (“recurrently altered” genes, Figure 1D,E, Supplementary Tables S4 and S5); importantly, all of these genes were located in 19p13.2-3 and all variants were detected at high allelic frequencies (mean variant allele fraction: 0.88, range: 0.74–0.93), suggesting a recurrent loss-of-heterozygosity (LOH) event in 19p13.2-3. The majority (56/64) of these variants were known polymorphisms (variant frequency in the general population $\geq 1\%$) (Figure 1D, Supplementary Tables S4 and S5) and 45/64 were classified as benign by the Polyphen-2 classifier. Variants that were not polymorphisms and that were classified as potentially or probably damaging, or for which functional prediction scores were not

available (Figure 1D) included: four variants in *SMARCA4* in five patients (described in detail hereafter; Figure 1E), the p.G223V variant in *PLK5* (one patient), the p.R220H variant in *ACTL9* (one patient) and the c.4208delT frameshift in *ABCA7* (one patient); all of these variants were Sanger-verified (Supplementary Table S4).

Lastly, no mutations in the following cancer-related genes were observed in any of the SCCOHT tumors, even at low allelic fractions: *TP53*, *KRAS*, *PIK3CA*, *PTEN*, *BRAF*, *EGFR*, *AKT1*, *CDKN1A* (p21) or *ERBB2*.

3.2. Inactivating *SMARCA4* Mutations and Related Findings

In line with previous studies, *SMARCA4* (Brahma-related gene 1, BRG1) was mutated by WES in 5/6 (83%) SCCOHT samples in our series (Figure 1C,D). The mean allelic fraction was high (0.86) and consistent with homozygous alterations. The encountered *SMARCA4* mutations were p.N774 frameshift, c.3216-1G>T (splice), p.R1077* stop gain (this mutation was identical in two patients) and p.K1081E (predicted as deleterious by the Polyphen-2 classifier). These variants have been previously reported in Le Loarer et al. [22], as part of the control cohort. All mutations occurred upstream of the SNF2-ATP coupling domain and the bromodomain, suggesting that the functional impact would be a loss of protein expression or function (Figure 1C). In line with those genomic findings, we observed loss of *SMARCA4* (BRG1) protein expression by IHC in all cases that displayed *SMARCA4* mutations (Figure 2A).

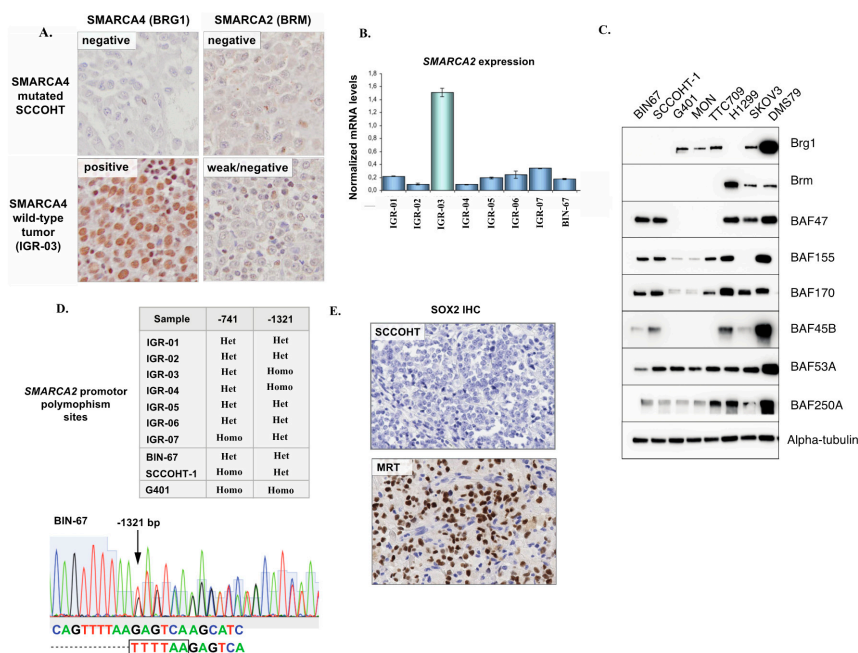


Figure 2. *SMARCA4* and *SMARCA2* expression in SCCOHT. (A) Representative *SMARCA4* and *SMARCA2* immunohistochemistry in a *SMARCA4* mutated SCCOHT and in the one *SMARCA4* wild-type. Tumor harboring concomitant *ARID1A* and *ARID1B* mutations (IGR-03). (B) Real-time RT-PCR for *SMARCA2* in patient tumor samples from this study and in a SCCOHT cell line (BIN-67); expression levels are normalized to three housekeeping genes (*YWHAZ/GUSB/HPRT1*). (C) Western blot showing expression of several SWI/SNF subunits in SCCOHT cell lines (BIN-67, SCCOHT-1) compared to MRT (G401, MON, TTC709), *SMARCA4*-mutated lung cancer (H1299), high-grade endometrioid adenocarcinoma of the ovary (SKOV3) and neuroendocrine small cell lung cancer (DMS79) cell lines. (D) Results of Sanger sequencing of the *SMARCA2* promoter insertional polymorphism sites, and an example of a heterozygous polymorphism status (−1321 site) in BIN-67 cells. (E) Representative IHC for SOX2 in SCCOHT and a positive control (SOX2-positive MRT) in patient FFPE tumor samples.

One tumor (IGR-03) diagnosed as SCCOHT did not exhibit a *SMARCA4* mutation but instead, harbored concomitant and potentially biallelic loss-of-function alterations in two other SWI/SNF genes: *ARID1A* (two frameshifts—p.Q555fs and p.T1004fs) and *ARID1B* (stop gained R1944*). Consequently, *SMARCA4* (BRG1) protein expression in this tumor was retained by IHC (Figure 2A).

Of note, one case (IGR-01) also showed a p.Arg635* stop gain in the *SMARCA1* gene, in addition to a deleterious *SMARCA4* mutation.

3.3. *SMARCA2* Loss of Expression in SCCOHT

Recent studies have shown that in addition to *SMARCA4* inactivation, SCCOHT exhibit a loss of expression of the *SMARCA2* paralog [23]. In our cohort, all *SMARCA4*-mutated SCCOHT (n = 5) showed low/absent *SMARCA2* transcript levels by real-time RT-PCR and complete absence of the *SMARCA2* (BRM) protein by IHC (Figure 2A,B), in keeping with previous studies. Combined loss of *SMARCA4* and *SMARCA2* expression in SCCOHT was also confirmed in our extended series of 33 FFPE SCCOHT, as we recently reported [24]. No *SMARCA2* loss-of-function mutations or deletions were found by WES. In the one *SMARCA4* wild-type tumor (IGR-03) which showed concomitant *ARID1A* and *ARID1B* mutations, *SMARCA2* expression was higher at the mRNA level (real-time RT-PCR) than in *SMARCA4* mutated samples, and interpreted as ambiguous/low at the protein level (IHC).

We also validated loss of *SMARCA4* and *SMARCA2* protein expression in two SCCOHT cell lines, BIN-67, and SCCOHT-1, by immunoblotting (Figure 2C). Notably, we observed that expression of other SWI/SNF subunits was retained, in line with a recent study by Pan et al. characterizing the presence of a residual SWI/SNF complex with altered functions in SCCOHT tumor cells [25].

The existence of homozygous insertional polymorphisms of the *SMARCA2* promoter, located −741 bp and −1321 bp from the transcription start site, has previously been linked to loss of *SMARCA2* expression in lung cancer [26]. Thus, we performed Sanger sequencing of the −741 and −1321 promoter sites in eight SCCOHT tumor samples and in two SCCOHT cell lines (BIN-67 and SCCOHT-1). One tumor and one cell line (SCCOHT-1) were homozygous for the −741 polymorphism, another tumor was homozygous for the −1321-promoter site polymorphism, and all other cases displayed a heterozygous −741 and −1321 polymorphism site status (Figure 2D). Overall, we concluded that a homozygous polymorphism site status in the *SMARCA2* promoter, previously described in lung cancer, was not a feature of SCCOHT.

Concomitant loss of *SMARCA4* and *SMARCA2* expression is also a feature of *SMARCA4*-deficient thoracic sarcomas (*SMARCA4*-DTS) and of a subset of malignant rhabdoid tumors (MRTs) [22,27,28]. *SMARCA4*-DTS have been reported to consistently show strong expression of the neural stem cell transcription factor SOX2 [27]. To verify whether the same was true for SCCOHT, 10 FFPE SCCOHT tumors were tested by IHC. Six showed no SOX2 staining (Figure 2E), two showed scattered positive cells and only two showed focal staining (<10% of tumor surface).

3.4. Validation of the p.G223V *PLK5* Variant in a Larger Series of SCCOHT Samples

PLK5 is the most recently described member of the Polo-Like Kinase family (PLK) family and has been implicated in involved in DNA damage response and cell cycle checkpoint control [29]. Given the presence of *PLK5* variants in a subset of SCCOHT detected by WES, the potentially damaging p.G223V variant was chosen for further Sanger validation in an extended cohort of 33 FFPE SCCOHT tumors (Supplementary Table S6). Overall, this mutation was detected in 3/33 (9%) of SCCOHT, suggesting that although it may be present in SCCOHT, it is not a highly recurrent event.

3.5. Somatic Copy Number Alterations (SCNAs) in SCCOHT

Eight fresh-frozen tumors were available for aCGH analysis. As shown in Figure 3B, the aCGH profiles of seven of the eight tumors showed remarkable genomic stability, with few SCNAs. The one tumor exhibiting genomic instability (IGR-07) harbored a loss of *BRCA2*, interpreted as heterozygous

[log₂ (ratio) = -0.4]. No SCNA was common to ≥50% of tumors, however, 16 recurrent gains were shared by at least three of the eight tumors (Table 1). Four of these genes showed a log₂ (ratio) > 2.3, which represents a five-fold increase in copy number, suggestive of amplification: *SHMT2*, *NDUFA4L2*, *LRP1* and *NXPH4*.

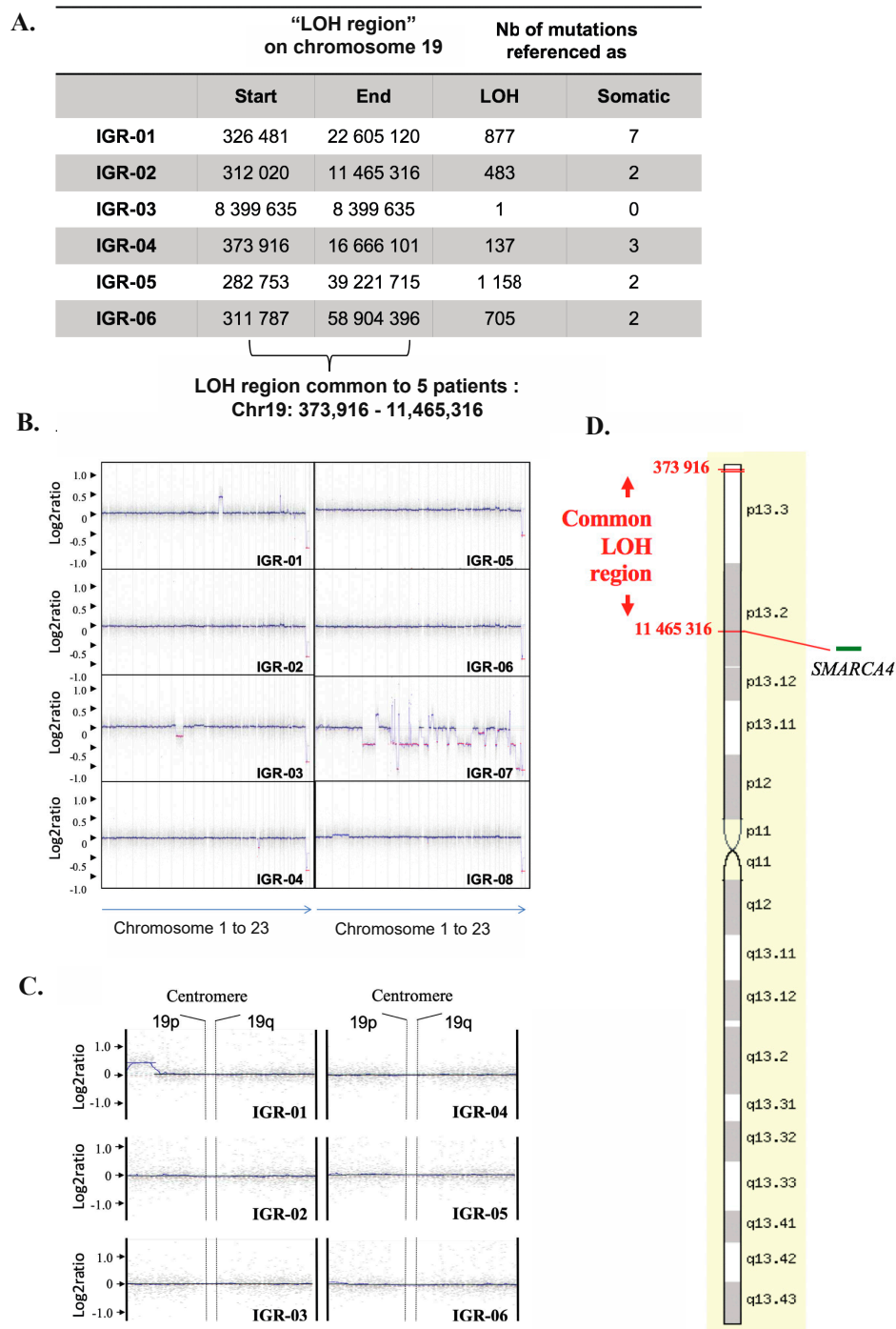


Figure 3. SCCOHT demonstrate remarkable genomic stability and recurrent 19p CN-LOH. (A) LOH regions obtained by WES in each tumor identifies a common "LOH region" on chromosome 19 for all SCCOHTs except IGR03: Chr19:373916-11465316. (B) CGH array profiles for each patient. (C) Zoom on 19p in all tumors fails to show a heterozygous copy number loss, thus suggestive of copy neutral LOH. (D) Artificial representation of the "common LOH region" on chromosome 19 in tumors (source: <http://www.genecards.org>).

Table 1. Recurrent Gains Shared by at Least three of eight Tumors. Aberrant SCNAs Were Defined as $\log_2(\text{ratio}) < -1$ or > 1 .

Localization	Gene Symbol	Description	Mean Log2Ratio
12q12-q14	SHMT2	Serine hydroxymethyltransferase 2 (mitochondrial)	2.33
12q13.3	NDUFA4L2	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 4-like 2	2.33
12q13.3	NXPH4	Neurexophilin 4	2.33
12q13.3	LRP1	Low density lipoprotein receptor-related protein 1	2.33
14q32.2	BEGAIN	Brain-enriched guanylate kinase-associated	2.23
14q32.2	LINC00523	Long intergenic non-protein coding RNA 523	2.23
16q24	CBFA2T3	Core-binding factor, runt domain, alpha subunit 2; translocated to, 3	1.40
16q24	APRT	Adenine phosphoribosyltransferase	1.40
16q24.3	ACSF3	Acyl-CoA Synthetase Family Member 3	1.40
16q24.3	CTU2	Cytosolic thioridylase subunit 2 homolog (S. pombe)	1.40
16q24.3	GALNS	Galactosamine (N-acetyl)-6-sulfate sulfatase	1.40
16q24.3	MIR4722	MicroRNA 4722	1.40
16q24.3	PABPN1L	Poly(A) binding protein, nuclear 1-like (cytoplasmic)	1.40
16q24.3	CDT1	Chromatin licensing and DNA replication factor 1	1.40
16q24.3	PIEZO1	Piezo-type mechanosensitive ion channel component 1	1.40
16q24.3	TRAPPC2L	Trafficking protein particle complex 2-like	1.40

3.6. Copy-Neutral Loss-of-Heterozygosity (CN-LOH) at the 19p13.2-3 Locus

WES revealed recurrent loss-of-heterozygosity (LOH) at the 19p13.2-3 region in five of six tumors and nominated the smallest common LOH region as Chr19:373.916-11.465.316 (Figure 3A). We mapped this “common LOH-region” by aCGH and, as shown in Figure 3B,C, confirmed that no copy number losses were present in this region, thus supporting the presence of a recurrent copy neutral-LOH (CN-LOH) event (Figure 3D).

3.7. Gene Expression Profiles of SCCOHT

To compare gene expression data in SCCOHT with our genomic findings, we conducted RNAseq-based differential expression analysis comparing six *SMARCA4*-mutated SCCOHT cases from our cohort and five samples of benign ovarian tissue from premenopausal women from the GTEx dataset. The analysis was conducted on rank-normalized gene expression values to reduce batch effect (Supplementary Table S7). As expected, *SMARCA2* expression was significantly lower in SCCOHT than in benign ovarian tissue, with mean rank-normalized expression values of 0.50 vs. 0.95, \log_2 fold change = -0.92 and $\text{padj} = 0.015$. No statistically significant difference in *SMARCA4* expression was detected, possibly due to low *SMARCA4* expression in two of the benign ovarian samples.

None of the four genes that showed recurrent amplifications in SCCOHT (*NDUFA4L2*, *SHMT2*, *NXPH4*, *LRP*) was significantly overexpressed in this analysis. *PLK5* showed very low expression values in both groups (mean rank-normalized expression values of 0.18 and 0.17, respectively). SCCOHT also did not show significant overexpression of *SOX2* (consistently with our IHC data) or of *PTHLH* (which encodes Parathyroid Hormone-Related Protein, previously postulated to cause hypercalcemia in a subset of SCCOHT).

Differential expression analysis also allowed the nomination of some genes potentially overexpressed in SCCOHT. Overall, ~1900 significantly differentially expressed genes showed \log_2 fold change > 1 or < -1 and $\text{padj} < 0.05$ (Supplementary Table S8). The most significantly overexpressed genes (top 100) included: cancer-testis antigens (e.g., *MAGEA4*, which was also the most significantly overexpressed gene, *MAGEA9*, *DSCR8*, *SYCE3*); the *AURKB* gene, encoding an Aurora kinase involved in mitotic progression; the tyrosine kinase receptor gene *ERBB4* (HER4); genes encoding metalloproteinases (e.g., *MMP10*, *MMP9*, *MMP1*); genes related to neural development (e.g., *NCAM2*, *NTS*, *ATCAY*, *CBLN2*); embryonic stem cell genes (*CLDN6*, which encodes an embryonic cell junction protein); and germ cell markers (*SALL4*) (Figure 4A, Supplementary Table S8). Conversely, the expression of some genes known to be highly expressed in benign ovarian tissue was significantly lower in SCCOHT (e.g., *INHA*, *FOXL2*, *AMHR2*). Gene Set Enrichment Analysis (GSEA) (Figure 4B, Supplementary Table S9) showed that gene sets significantly enriched in SCCOHT included those

related to E2F targets and cell cycle progression, DNA repair, activation of oncogenic pathways (KRAS, MYC, mTORC1), as well as gene sets related to *SMARCB1* (SNF5) knockdown, consistent with a deregulated SWI/SNF complex.

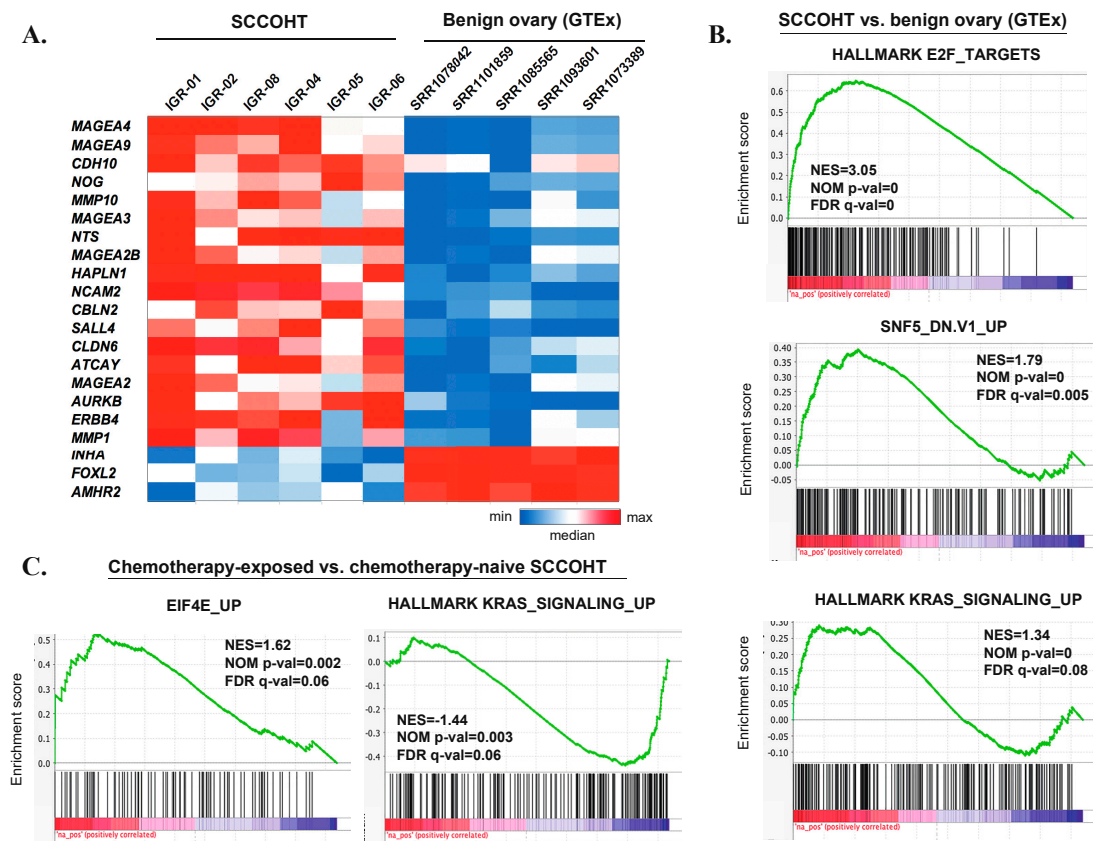


Figure 4. An overview of transcriptomic profiles of SCCOHT. (A) Graphic heatmap representation of rank-normalized expression values for selected, most significantly deregulated genes in the differential expression analysis between SCCOHT and benign ovarian tissue (GTEx). (B) Selected GSEA results for the differential expression analysis between SCCOHT and benign ovarian tissue (GTEx). (C) Selected GSEA results for the differential expression analysis between chemotherapy-exposed SCCOHT samples (IGR-01, IGR-04, IGR-06) and chemotherapy-naïve samples (IGR-02, IGR-05, IGR-08).

3.8. Genomic and Transcriptomic Profiles of Chemotherapy-Naïve Versus Chemotherapy-Exposed SCCOHT

SCCOHT are characterized by initial chemosensitivity, but almost invariably relapse. Thus, we compared mutation profiles in the three treatment-naïve and three chemotherapy-exposed tumors, to determine whether some alterations were enriched in post-treatment samples. The chemotherapy regimens received by patients IGR01, IGR04 and IGR06 prior to sample collection/surgery were VIP/Doxorubicin, BEP/PAVEP and EP/PAVEP, respectively (Supplementary Table S1).

Tumor mutation burden was not significantly higher in the three post-chemotherapy samples than in the three chemo-naïve ones: mean, 4.98 non-synonymous mutations/Mb (range: 3.56–6.42) and mean, 5.87 mutations/Mb (range, 5.21–6.42), respectively ($p = 0.38$, unpaired t-test). Among somatic-only mutations (i.e., mutations not imputable to an LOH event), two genes were altered in at least two post-chemotherapy samples, but not in chemo-naïve samples: *ADGRV1* (IGR-01 and IGR-06—one stop gain and one missense alteration) and *FANCD2* (IGR-04 and IGR-06—splice region variant predicted to be of low functional impact by the SnpEff tool). No SCNAs were differentially detected in the chemotherapy-exposed versus chemotherapy-naïve tumors (data not shown).

RNAseq-based differential expression analysis between the chemotherapy-exposed and the chemotherapy-naïve samples did not identify any significantly differentially expressed genes (data not shown). However, when analyzing ranked genes in their totality, GSEA analysis revealed several gene sets with significant positive or negative enrichment (Supplementary Table S10). Among these, there was a positive enrichment of the gene set reflecting genes upregulated upon overexpression of Eukaryotic Translation Initiation Factor 4E (eEIF4E), a positive enrichment of gene sets downregulated upon KRAS overexpression, and a negative enrichment of genes upregulated upon KRAS overexpression (Figure 4C).

3.9. Epigenetic Vulnerabilities of SCCOHT Associated with SWI/SNF Deregulation

In a recent study, Pan et al. have shown that the loss of catalytic SWI/SNF activity in SCCOHT largely alters SWI/SNF functions as an epigenetic regulator [25]. To assess the putative sensitivity of SCCOHT to currently available epigenetic treatments, cell lines with differing *SMARCA4* genotypes were treated with the histone deacetylase inhibitor trichostatin A (TSA) and the DNA methyltransferase inhibitor 5'-dAZAC. The SCCOHT cell line BIN-67, which harbors an inactivating *SMARCA4* mutation and shows complete loss of *SMARCA2* expression (Figure 2C), was exquisitely sensitive to 5'-dAZAC and TSA at sub-nanomolar concentrations (Figure 5A,B). Conversely, the H1299 lung cancer cell line, which carries a *SMARCA4* mutation, but shows retained *SMARCA2* expression (*SMARCA4*+/SMARCA2+), and the ovarian high-grade endometrioid adenocarcinoma cell line SKOV3 (*SMARCA4*+/*SMARCA2*+) were completely resistant to 5'-dAZAC and 100-fold less sensitive to TSA than BIN-67 (Figure 5A,B).

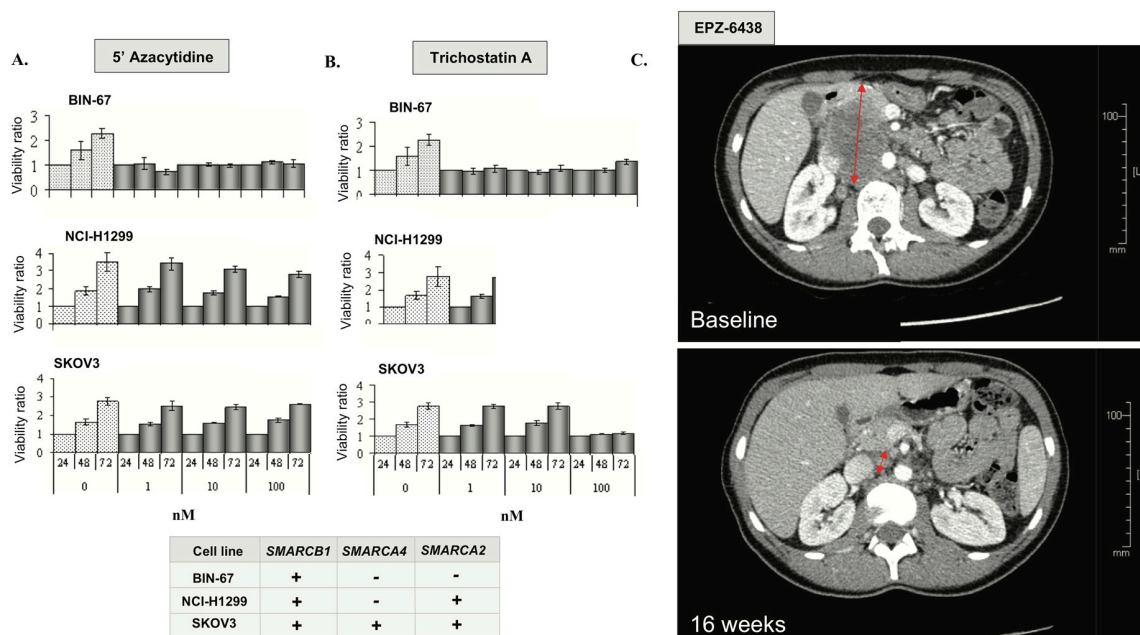


Figure 5. Epigenetic vulnerabilities in SCCOHT A, B. Anti-proliferative effects of 5'-AZAC (A) and TSA (B); – designates protein loss or loss-of-function mutation and/or loss of expression; + designates absence of mutation (wild-type status) and retained expression. (C) Rapid clinical response in *SMARCA4*-mutated SCCOHT treated with the EZH2 inhibitor EPZ-6438. A CT scan of the tumor at baseline and after four months of EPZ-6438 treatment with 70% decrease in tumor volume (RECIST 1.1).

Many studies suggest an antagonistic relationship between the SWI/SNF complex and Polycomb proteins, such as Enhancer of Zest 2 (EZH2) [30]. Combined loss of *SMARCA4* and *SMARCA2* in SCCOHT cell lines may induce an oncogenic dependency on EZH2 activation [31] and confer extreme sensitivity to EZH2 inhibitors in vitro and in vivo [32]. Based on this rationale, we enrolled a patient

with *SMARCA4*-mutated SCCOHT in a phase I trial of tazemetostat (EPZ-6438), a highly selective EZH2 inhibitor [33]. This 25-year-old patient initially presented with stage IV SCCOHT treated with surgery, combination platinum-based chemotherapy followed by high dose consolidation and autologous stem cell rescue (Figure 5C). Unfortunately, she relapsed within eight weeks and was, therefore, enrolled in the EZH2 inhibitor clinical trial. She presented partial response (RECIST 1.1) after four months of treatment and remained progression-free for eight months. Although the clinical benefit was relatively short, the degree of response in this patient with highly chemo-resistant disease supports further investigation of epigenetic strategies in SCCOHT.

4. Discussion

SCCOHT are rare tumors that occur in young women and their prognosis remains poor, despite aggressive multimodal therapy. We present an integrated molecular characterization of additional cases of SCCOHT from an independent cohort.

Intriguingly, our findings and the previously published data show that these aggressive tumors carry a diploid DNA content, which is a rare phenomenon in a highly lethal malignancy [34,35]. In addition, we show that SCCOHT have a very low mutation load (mean, 5.43 mut/Mb) and lack mutations in genes most altered across various cancer types. Collectively, these observations support the hypothesis that SCCOHT are largely driven by epigenetic deregulation and not by genomic instability.

Importantly, our results underscore marked inter- and intra-tumor homogeneity of SCCOHTs. Combined WES and aCGH analysis revealed a recurrent copy-neutral LOH (CN-LOH) at the 19p13.2-3 locus. CN-LOH can account for inactivation of tumor suppressor genes and likely implicates the loss of the normal allele and duplication of the mutated copy. Notably, 19p LOH has previously been detected in SCCOHT by WES [4,36], but our results provide additional evidence for a copy-neutral nature of this event. Of note, telomeric CN-LOH has been linked to meiotic errors occurring during cross-over [37], which could be in line with the postulated germ cell origin of SCCOHT tumors [2,3], although further studies are needed to support this hypothesis. The 19p CN-LOH associated with inactivating *SMARCA4* mutations has also been reported in non-small cell lung cancer [38].

In line with previous studies, we found that *SMARCA4* mutations were present in all but one SCCOHT (5/6) in our series. *SMARCA4* encodes one of the enzymatic (ATP-ase) subunits of mammalian SWI/SNF, a chromatin remodeling complex which directs nucleosomes and modulates gene expression. The importance of SWI/SNF alterations in oncogenesis or tumor progression is being increasingly acknowledged, as alterations in SWI/SNF subunits are found in over 20% of human cancers [36]. Early preclinical studies suggested that *SMARCA4*-mutated tumors (such as non-small cell lung cancers) were critically reliant on the *SMARCA2* paralog [29,39]. Conversely, SCCOHT do not seem amenable to this synthetic lethality strategy, given the complete loss of *SMARCA2* expression demonstrated in our series and in previous studies [22]. This loss of expression is not explained by mutations in the coding sequence of *SMARCA2*. In an effort to explore the underpinnings of *SMARCA2* silencing, we report for the first time that homozygous polymorphisms at the two *SMARCA2* promoter polymorphism sites (−741 bp and −1321 bp), previously linked to *SMARCA2* silencing in cancer [15,40], do not seem to be a recurrent event responsible for *SMARCA2* silencing in SCCOHT. Nevertheless, since most tumors in our study showed a heterozygous promoter polymorphism site status, further explorations are warranted to elucidate whether heterozygous polymorphisms can contribute to *SMARCA2* silencing. In particular, in malignant rhabdoid tumor cell lines, increased binding of epigenetic silencers HDAC9 and MEF2D at *SMARCA2* promoter sites has been associated with such heterozygous polymorphisms [41].

One case (IGR-03) from our series exhibited concomitant inactivating mutations in *ARID1A* and *ARID1B*, two paralog DNA-binding subunits of SWI/SNF, but did not show *SMARCA4* mutations. This tumor was also the only case in which 19p CN-LOH was not present. Concomitant *ARID1A/B* alterations occur in ~25% of dedifferentiated endometrial and ovarian carcinomas [42]. While case IGR-03 could illustrate the challenges of differential diagnosis between SCCOHT and dedifferentiated

ovarian carcinoma, another possibility is the existence of a molecular and morphologic overlap between those two entities, both of which are characterized by a poorly differentiated, aggressive tumor and a critically deregulated SWI/SNF complex.

Of note, one case (IGR-01) showed a p.Arg635* stop gain in the *SMARCA1* gene in addition to a deleterious *SMARCA4* mutation. *SMARCA1* encodes the ATP-ase of another chromatin remodeling complex, ISWI, and is located on the X chromosome, suggesting that this alteration, which was seen at an allele frequency of 0.37, could potentially carry a deleterious impact.

As a complement to the genomic findings, we also show for the first time that SCCOHT are not characterized by SOX2 overexpression, contrary to another aggressive *SMARCA4/SMARCA2* double-negative malignancy—*SMARCA4*-deficient thoracic sarcoma (*SMARCA4*-DTS). This emphasizes the existence of biological differences between *SMARCA4*-DTS and SCCOHT, in addition to previously described discrepancies, such as higher genomic instability in *SMARCA4*-DTS [26], and could have potential implications in diagnostic pathology.

In addition to *SMARCA4* loss-of-function alterations, a few variants were seen in genes other than *SMARCA4*, all localized in the 19p13.2-3 locus and subject to the CN-LOH event, for which a functional impact could not be ruled out. In particular, the *PLK5* p.G223V variant, predicted as potentially damaging by the Polyphen-2 classifier, was found in 3/33 SCCOHT samples of the extended cohort. The protein kinase domain of *PLK5* is truncated in humans compared to mice, but the residual protein containing the polo-box binding domain may act as a stress inducible tumor suppressor regulating G1 arrest [29]. Nevertheless, the relevance of these variants remains to be validated functionally.

Differential expression analysis comparing SCCOHT and benign ovarian tissue allowed to nominate several genes potentially overexpressed in SCCOHT. Nevertheless, it must be kept in mind that the bulk benign ovarian tissue used as the control in this analysis does not represent the exact cell origin of SCCOHT, which remains unknown. Among other findings, we observed significant overexpression of some putative therapeutic targets. Cancer-testis antigens have been proposed as targets for immunotherapy approaches and Melanoma-associated antigen 4 (*MAGEA4*), which was the most highly overexpressed gene in SCCOHT, is currently being investigated as a TCR-engineered T-cell target (NCT03247309). The *AURKB* gene encodes Aurora B kinase, implicated in mitotic progression, and may be targeted by pharmaceutical inhibitors (e.g., GSK1070916). The overexpression of receptor tyrosine kinase genes, such as *ERBB4* (HER4), could potentially be in line with a recent study showing marked vulnerability of SCCOHT cells to multi-kinase inhibition [43]. Intriguingly, SCCOHT also showed expression of neural differentiation markers and embryonic stem cell markers, in keeping with what has previously been reported in malignant rhabdoid tumors [44]. Some of these markers could represent treatment opportunities, such as the embryonic cell junction protein Claudin-6 (*CLDN6*), against which monoclonal antibodies were recently part of a clinical trial in ovarian cancers (NCT02054351). Nevertheless, further studies are needed to confirm overexpression of these putative treatment targets at the protein level and to validate their functional relevance in SCCOHT.

To further explore putative therapeutic approaches, we found that the SCCOHT cell line BIN67 was exquisitely sensitive to TSA and 5'-dAZAC, while cells with a *SMARCA4* mutation and retained *SMARCA2* expression were resistant to these epigenetic therapies. In addition, we describe a clinically meaningful response to single agent EZH2 inhibitor in a patient with SCCOHT, in keeping with what has previously been demonstrated in vitro and in vivo [32]. Collectively, data from our series and from previous studies suggest that SCCOHT tumors characterized by the loss of both SWI/SNF catalytic subunits may be sensitive to treatment with HDAC, DNA methyltransferase and/or EZH2 inhibitors, and that such strategies merit further investigation in this lethal disease.

Since SCCOHT often display initial chemosensitivity, but subsequently show rapid progression, we compared the genomics of treatment-naïve (n = 3) versus chemotherapy-treated tumors (n = 3), to uncover candidate resistance genes. Critically, neither the tumor mutation burden nor the somatic copy number alterations were significantly increased in post-chemotherapy samples. The only alterations seen in more than one post-chemotherapy sample and not in chemotherapy-naïve

samples were *ADGRV1* (one stop gain and one missense alteration) and *FANCD2* (splice region variants). Variants in *ADGRV1*, which encodes adhesion G protein-coupled receptor 1, were potentially heterozygous, and the implication of this protein in cancer progression is unclear. Variants in *FANCD2* were predicted to be of low functional impact by the SnpEff tool. Although differential expression analysis did not reveal any specific genes significantly deregulated between the two groups, GSEA performed on the totality of ranked genes nominated several deregulated gene expression programs, including an enrichment of genes related to eIF4E upregulation and a putative downregulation of the KRAS pathway. Further studies comparing paired samples from the same patient before and after treatment are needed to elucidate molecular underpinnings of treatment resistance in SCCOHT. Nevertheless, our data suggest that it may rely on other mechanisms than acquiring drug resistance mutations, contrary to what has been described in other tumor types.

5. Conclusions

SCCOHT are unique tumors. Despite their aggressive clinical course, they display genomic stability, low mutation load, few SCNAs, and a remarkably homogeneous genomic profile. They are almost universally characterized by a 19p CN-LOH, loss-of-function mutations in *SMARCA4* and concomitant loss of *SMARCA2* expression. SCCOHT do not seem to acquire additional mutations after exposure to chemotherapy. Some additional molecular alterations reported herein could be further explored as therapeutic opportunities, such as the overexpression of putative therapeutic targets (e.g., *MAGEA4*, *AURKB* or *CLDN6*). Our preliminary in vitro data and the reported patient case also support the view that SCCOHT are sensitive to epigenetic modulators, such as HDAC, DNMT and EZH2 inhibitors, in line with other studies and with early results of phase I trials, and warrant further exploration of epigenetic treatment strategies in this lethal disease.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2073-4409/9/6/1496/s1>, Table S1: Clinical details on 8 SCCOHT cases studied by exome sequencing and/or CGH arrays, Table S2: Summary of Whole-Exome Sequencing (WES) statistics. Table S3: Filters used to screen all single nucleotide variants (SNVs) identified using WES, with numbers of variants left at each filtering step, Tables S4 & S5: Complete genomic description and validation status of recurrent mutations, Table S6: Targeted mutational analysis for PLK5 mutations in 33 FFPE SCCOHT samples, Table S7: Complete description of recurrent mutations in the GR series of SCCOHTs and comparison with other SCCOHT series and with the rate of mutation in HGSOc. Table S8: Filters used to screen significant differences with other SCCOHT series and benign ovarian tissue, Table S9: GSEA results for SCCOHT vs benign ovarian tissue (GTEx); results with NOM p-val < 0.05 and FDR q-val < 0.25 are shown, Table S10: GSEA results for chemotherapy-exposed vs chemotherapy-naïve SCCOHT tumors, C6 Oncogenic gene sets ensemble; results with NOM p-val < 0.05 and FDR q-val < 0.25 are shown, Table S11: List of primers used in Sanger method to validate mutations of interest identified by WES. Figure S1: Morphological aspect of 4 SCCOHT samples.

Author Contributions: Conceptualization, A.A., J.C. and A.L.; Data curation, A.A., M.D., J.C. and A.L.; Formal analysis, A.A., M.D., J.C. and A.L.F.; Investigation, C.G.; Methodology, A.A., R.B. and C.G.; Project administration, B.B.; Supervision, R.B., D.C.W., C.R., O.C., M.D.-S., S.G., P.M., E.B., A.S., O.E., M.A.R., P.P., C.G. and A.L.; Validation, J.C. and A.L.; Writing—original draft, A.A., F.B.-D. and J.C.; Writing—review & editing, A.A., F.B.-D., J.C. and A.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by an educational grant from Fondation Gustave Roussy.

Acknowledgments: Bio-Informatics Platform, Gustave Roussy (R. Tien-Tao, Y. Boursin), Genomics Platform Gustave Roussy (O. Bernard and K. Diop for WES), Translational Research Platform Gustave Roussy: Histocytology module (F. Drusch), Molecular Biopath Module (P. Saulnier, Targeted NGS), Gustave Roussy Research Biobank (Centre de Ressources Biologiques), Pathology SIRIC SOCRATE Module, Gustave Roussy (grant INCa-DGOS-INSERM 6043), National Rare Ovarian Tumor Observatory (http://www.ovaire-rare.org/TMRG/public/accueil_public.aspx). A special thanks to Barbara Vanderhyden, Centre for Cancer Therapeutics, Ottawa Hospital Research Institute, Ontario K1H 8L6 for kindly providing the BIN-67 cell line and Bernard Weissman (University of North Carolina School of Medicine, Chapel Hill, NC, USA) for kindly providing the TTC-709 cell line; Cochin Hospital, Dept of Medical Oncology and Pathology Dpts, University Hospital Grenoble, Longjumeau University Hospital and Hôpital de la Croix Rousse for contributing SCCOHT tumor samples; and to the urology team INSERM U981 for providing inhibitors (A. Chauchereau).

Conflicts of Interest: The authors declare that they have no competing interest.

References

1. Young, R.H.; Oliva, E.; Scully, R.E. Small cell carcinoma of the ovary, hypercalcemic type. A clinicopathological analysis of 150 cases. *Am. J. Surg. Pathol.* **1994**, *18*, 1102–1116. [[CrossRef](#)] [[PubMed](#)]
2. McCluggage, W.G. Ovarian neoplasms composed of small round cells: A review. *Adv. Anat. Pathol.* **2004**, *11*, 288–296. [[CrossRef](#)] [[PubMed](#)]
3. Kupryjańczyk, J.; Dansonka-Mieszkowska, A.; Moes-Sosnowska, J.; Plisiecka-Hałas, J.; Szafron, Ł.; Podgórska, A.; Rzepecka, I.K.; Konopka, B.; Budziłowska, A.; Rembiszewska, A.; et al. Ovarian small cell carcinoma of hypercalcemic type—Evidence of germline origin and smarca4 gene inactivation. A pilot study. *Pol. J. Pathol.* **2013**, *4*, 238–246. [[CrossRef](#)] [[PubMed](#)]
4. Witkowski, L.; Carrot-Zhang, J.; Albrecht, S. Germline and somatic SMARCA4 mutations characterize small cell carcinoma of the ovary, hypercalcemic type. *Nat. Genet.* **2014**, *46*, 438–443. [[CrossRef](#)] [[PubMed](#)]
5. Barondeau, J.; Rodgers, M.; Braun, L.; Azarow, K.; Forouhar, M.; Faucette, K. Small cell ovarian carcinoma: A rare, aggressive tumor masquerading as constipation in a teenager with a fatal outcome. *J. Pediatr. Hematol. Oncol.* **2010**, *32*, e139–e141. [[CrossRef](#)]
6. Pautier, P.; Ribrag, V.; Duvillard, P. Results of a prospective dose-intensive regimen in 27 patients with small cell carcinoma of the ovary of the hypercalcemic type. *Ann. Oncol.* **2007**, *18*, 1985–1989. [[CrossRef](#)]
7. Jelinic, P.; Mueller, J.J.; Olvera, N. Recurrent SMARCA4 mutations in small cell carcinoma of the ovary. *Nat. Genet.* **2014**, *46*, 424–426. [[CrossRef](#)]
8. Ramos, P.; Karnezis, A.N.; Craig, D.W.; Sekulic, A.; Russell, M.L.; Hendricks, W.P.D.; Corneveaux, J.J.; Barrett, M.T.; Shumansky, K.; Yang, Y.; et al. Small cell carcinoma of the ovary, hypercalcemic type, displays frequent inactivating germline and somatic mutations in SMARCA4. *Nat. Genet.* **2014**, *46*, 427–429. [[CrossRef](#)]
9. Lin, D.I.; Chudnovsky, Y.; Duggan, B.; Zajchowski, D.; Greenbowe, J.; Ross, J.S.; Gay, L.M.; Ali, S.M.; Elvin, J.A. Comprehensive genomic profiling reveals inactivating SMARCA4 mutations and low tumor mutational burden in small cell carcinoma of the ovary, hypercalcemic-type. *Gynecol. Oncol.* **2017**, *147*, 626–633. [[CrossRef](#)]
10. Li, H.; Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **2010**, *26*, 589–595. [[CrossRef](#)]
11. McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernytzky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M.; et al. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **2010**, *20*, 1297–1303. [[CrossRef](#)] [[PubMed](#)]
12. Koboldt, D.C.; Zhang, Q.; Larson, D.E. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **2012**, *22*, 568–576. [[CrossRef](#)] [[PubMed](#)]
13. Cingolani, P.; Platts, A.; Wang, L.L.; Coon, M.; Nguyen, T.; Wang, L.; Land, S.J.; Lu, X.; Ruden, D.M. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **2012**, *6*, 80–92. [[CrossRef](#)] [[PubMed](#)]
14. Cingolani, P.; Patel, V.M.; Coon, M. Using *drosophila melanogaster* as a model for genotoxic chemical mutational studies with a new program, SnpSift. *Front. Genet.* **2012**, *3*, 35. [[CrossRef](#)]
15. Liu, X.; Jian, X.; Boerwinkle, E. dbNSFP: A lightweight database of human nonsynonymous SNPs and their functional predictions. *Hum. Mutat.* **2011**, *32*, 894–899. [[CrossRef](#)] [[PubMed](#)]
16. Xu, Z.; Dai, J.; Wang, D.; Lu, H.; Dai, H.; Ye, H.; Gu, J.; Chen, S.; Huang, B. Assessment of tumor mutation burden calculation from gene panel sequencing data. *Onco Targets* **2019**, *12*, 3401–3409. [[CrossRef](#)]
17. Olshen, A.B.; Venkatraman, E.S.; Lucito, R.; Wigler, M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **2004**, *5*, 557–572. [[CrossRef](#)]
18. Karolchik, D.M.; Baertsch, R. The UCSC genome browser database. *Nucleic Acids Res.* **2003**, *31*, 51–54. [[CrossRef](#)]
19. Engström, P.G.; Steijger, T.; Sipos, B.; Grant, G.R.; Kahles, A.; Rättsch, G.; Rättsch, G.; Goldman, N.; Hubbard, T.J.; Harrow, J.; et al. Systematic evaluation of spliced alignment programs for RNA-seq data. *Nat. Methods* **2013**, *10*, 1185–1191. [[CrossRef](#)]
20. DePristo, M.A.; Banks, E.; Poplin, R.; Garimella, K.V.; Maguire, J.R.; Hartl, C.; Philippakis, A.; del Angel, G.; Rivas, M.; Hanna, M.; et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **2011**, *43*, 491–498. [[CrossRef](#)]

21. Van der Auwera, G.A.; Carneiro, M.O.; Hartl, C.; Poplin, R.; Del Angel, G.; Levy-Moonshine, A.; Jordan, T.; Shakir, K.; Roazen, D.; Thibault, J.; et al. From FastQ data to high confidence variant calls: The Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinform.* **2013**, *43*, 11.10.1–11.10.33. [[CrossRef](#)]
22. Le Loarer, F.; Watson, S.; Pierron, G.; de Montpreville, V.T.; Ballet, S.; Firmin, N.; Auguste, A.; Pissaloux, D.; Boyault, S.; Paindavoine, S.; et al. SMARCA4 inactivation defines a group of undifferentiated thoracic malignancies transcriptionally related to BAF-deficient sarcomas. *Nat. Genet.* **2015**, *47*, 1200–1205. [[CrossRef](#)] [[PubMed](#)]
23. Karnezis, A.N.; Wang, Y. Dual loss of the SWI/SNF complex ATPases SMARCA4/BRG1 and SMARCA2/BRM is highly sensitive and specific for small cell carcinoma of the ovary, hypercalcaemic type. *J. Pathol.* **2016**, *238*, 389–400. [[CrossRef](#)]
24. Genestie, C.; Blanc-Durand, F.; Auguste, A.; Pautier, P.; Dunant, A.; Scoazec, J.-Y.; Gouy, S.; Morice, P.; Bentivegna, E.; Maulard, A.; et al. Clinical utility of SMARCA4 testing by immunohistochemistry in rare ovarian tumours. *Br. J. Cancer* **2019**, *122*, 564–568. [[CrossRef](#)] [[PubMed](#)]
25. Pan, J.; McKenzie, Z.M.; D’Avino, A.R.; Mashtalir, N.; Lareau, C.A.; St Pierre, R.; Wang, L.; Shilatfard, A.; Kadoch, C. The ATPase module of mammalian SWI/SNF family complexes mediates subcomplex identity and catalytic activity-independent genomic targeting. *Nat. Genet.* **2019**, *51*, 618–626. [[CrossRef](#)]
26. Liu, G.; Gramling, S.; Munoz, D.; Cheng, D.; Azad, A.K.; Mirshams, M.; Chen, Z.; Xu, W.; Roberts, H.; Shepherd, F.A.; et al. Two novel BRM insertion promoter sequence variants are associated with loss of BRM expression and lung cancer risk. *Oncogene* **2011**, *30*, 3295–3304. [[CrossRef](#)]
27. Hasselblatt, M.; Nagel, I.; Oyen, F.; Bartelheim, K.; Russell, R.B.; Schüller, U.; Junckerstorff, R.; Rosenblum, M.; Alassiri, A.H.; Rossi, S.; et al. SMARCA4-mutated atypical teratoid/rhabdoid tumors are associated with inherited germline alterations and poor prognosis. *Acta Neuropathol.* **2014**, *128*, 453–456. [[CrossRef](#)]
28. Venneti, S.; Le, P.; Martinez, D.; Xie, S.X.; Sullivan, L.M.; Rorke-Adams, L.B.; Bruce, P.; Alexander, J. Malignant rhabdoid tumors express stem cell factors, which relate to the expression of EZH2 and Id proteins. *Am. J. Surg. Pathol.* **2011**, *35*, 1463–1472. [[CrossRef](#)]
29. Andrysik, Z.; Bernstein, W.Z.; Deng, L.; Myer, D.L.; Li, Y.-Q.; Tischfield, J.A.; Stambrook, P.J. The novel mouse Polo-like kinase 5 responds to DNA damage and localizes in the nucleolus. *Nucleic Acids Res.* **2010**, *38*, 2931–2943. [[CrossRef](#)]
30. Wilson, B.G.; Wang, X.; Shen, X.; McKenna, E.S.; Lemieux, M.E.; Cho, Y.-J.; Koellhoffer, E.C.; Pomeroy, S.L.; Orkin, S.H.; Roberts, C. Epigenetic antagonism between polycomb and SWI/SNF complexes during oncogenic transformation. *Cancer Cell* **2010**, *18*, 316–328. [[CrossRef](#)]
31. Kim, K.H.; Kim, W.; Howard, T.P.; Vazquez, F.; Tsherniak, A.; Wu, J.N. SWI/SNF-mutant cancers depend on catalytic and non-catalytic activity of EZH2. *Nat. Med.* **2015**, *21*, 1491–1496. [[CrossRef](#)] [[PubMed](#)]
32. Wang, Y.; Chen, S.Y.; Karnezis, A.N.; Colborne, S.; Santos, N.D.; Lang, J.D. The histone methyltransferase EZH2 is a therapeutic target in small cell carcinoma of the ovary, hypercalcaemic type. *J. Pathol.* **2017**, *242*, 371–383. [[CrossRef](#)] [[PubMed](#)]
33. Italiano, A.; Soria, J.-C.; Toulmonde, M.; Michot, J.-M.; Lucchesi, C.; Varga, A. Tazemetostat, an EZH2 inhibitor, in relapsed or refractory B-cell non-Hodgkin lymphoma and advanced solid tumours: A first-in-human, open-label, phase 1 study. *Lancet Oncol.* **2018**, *19*, 649–659. [[CrossRef](#)]
34. Eichhorn, J.H.; Young, R.H.; Scully, R.E. Primary ovarian small cell carcinoma of pulmonary type. A clinicopathologic, immunohistologic, and flow cytometric analysis of 11 cases. *Am. J. Surg. Pathol.* **1992**, *16*, 926–938. [[CrossRef](#)] [[PubMed](#)]
35. Förster, C.; Ostertag, H.; Schmitt, J.; Roessner, A. Small cell carcinoma of the ovary, hypercalcemic type. A case report with immunohistochemical, ultrastructural and cytophotometric analysis and review of the literature. *Gen. Diagn. Pathol.* **1997**, *142*, 365–370.
36. O’Keefe, C.; McDevitt, M.A.; Maciejewski, J.P. Copy neutral loss of heterozygosity: A novel chromosomal lesion in myeloid malignancies. *Blood* **2010**, *115*, 2731–2739. [[CrossRef](#)]
37. Kadoch, C.; Hargreaves, D.C.; Hodges, C.; Elias, L.; Ho, L.; Ranish, J. Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. *Nat. Genet.* **2013**, *45*, 592–601. [[CrossRef](#)]
38. David, M.P.; Venkatramani, R.; Lopez-Terrada, D.H.; Roy, A.; Patil, N.; Fisher, K.E. Multimodal molecular analysis of an atypical small cell carcinoma of the ovary, hypercalcemic type. *Cold Spring Harb. Mol. Case Stud.* **2018**, *4*, a002956. [[CrossRef](#)]

39. Medina, P.P.; Carretero, J.; Fraga, M.F.; Esteller, M.; Sidransky, D.; Sanchez-Cespedes, M. Genetic and epigenetic screening for gene alterations of the chromatin-remodeling factor, SMARCA4/BRG1, in lung tumors. *Genes Chromosomes Cancer* **2004**, *41*, 170–177. [[CrossRef](#)]
40. Hoffman, G.R.; Rahal, R.; Buxton, F.; Xiang, K.; McAllister, G.; Frias, E. Functional epigenetics approach identifies BRM/SMARCA2 as a critical synthetic lethal target in BRG1-deficient cancers. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 3128–3133. [[CrossRef](#)]
41. Kahali, B.; Yu, J.; Marquez, S.B.; Thompson, K.e.n.e.t.h.W.; Liang, S.Y.; Lu, L. The silencing of the SWI/SNF subunit and anticancer gene BRM in Rhabdoid tumors. *Oncotarget* **2014**, *5*, 3316–3332. [[CrossRef](#)] [[PubMed](#)]
42. Coatham, M.; Li, X.; Karnezis, A.N.; Hoang, L.N.; Tessier-Cloutier, B.; Meng, B. Concurrent ARID1A and ARID1B inactivation in endometrial and ovarian dedifferentiated carcinomas. *Mod. Pathol.* **2016**, *29*, 1586–1593. [[CrossRef](#)] [[PubMed](#)]
43. Lang, J.D.; Hendricks, W.P.D.; Orlando, K.A.; Yin, H.; Kiefer, J.; Ramos, P. Ponatinib shows potent antitumor activity in small cell carcinoma of the ovary hypercalcemic type (SCCOHT) through multikinase inhibition. *Clin. Cancer Res.* **2018**, *24*, 1932–1943. [[CrossRef](#)] [[PubMed](#)]
44. Muscat, A.; Popovski, D.; Jayasekara, W.S.N.; Rossello, F.J.; Ferguson, M.; Marini, K.D. Low-dose histone deacetylase inhibitor treatment leads to tumor growth arrest and multi-lineage differentiation of malignant rhabdoid tumors. *Clin. Cancer Res.* **2016**, *22*, 3560–3570. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Aberrant Activation of a Gastrointestinal Transcriptional Circuit in Prostate Cancer Mediates Castration Resistance

Highlights

- A GI-lineage transcriptome is prevalent in castration-resistant prostate cancer
- HNF4G and HNF1A regulate each other and the GI-lineage transcriptome
- HNF4G is a pioneer factor that generates and maintains enhancers at GI-lineage genes
- Exogenous HNF4G expression in prostate cancer leads to castration resistance

Authors

Shipra Shukla, Joanna Cyrta, Devan A. Murphy, ..., Howard I. Scher, Ping Chi, Yu Chen

Correspondence

chip@mskcc.org (P.C.),
cheny1@mskcc.org (Y.C.)

In Brief

Shukla et al. identify an aberrantly expressed gastrointestinal-lineage transcriptome governed by HNF4G and HNF1A in ~30% of castration-resistant prostate cancer. HNF4G is a pioneer factor for this transcriptional program and its ectopic expression at physiologic levels reduces sensitivity to hormone deprivation.



Aberrant Activation of a Gastrointestinal Transcriptional Circuit in Prostate Cancer Mediates Castration Resistance

Shipra Shukla,¹ Joanna Cyrta,^{2,3} Devan A. Murphy,¹ Edward G. Walczak,¹ Leili Ran,¹ Praveen Agrawal,¹² Yuanyuan Xie,¹ Yuedan Chen,¹ Shangqian Wang,¹ Yu Zhan,¹ Dan Li,¹³ Elissa W.P. Wong,¹ Andrea Sboner,^{2,3,4} Himisha Beltran,^{3,5} Juan Miguel Mosquera,^{2,3} Jessica Sher,¹ Zhen Cao,¹ John Wongvipat,¹ Richard P. Koche,⁶ Anuradha Gopalan,⁷ Deyou Zheng,^{8,9,10} Mark A. Rubin,^{2,3} Howard I. Scher,^{5,11} Ping Chi,^{1,5,11,*} and Yu Chen^{1,5,11,14,*}

¹Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA

²Department of Pathology and Laboratory Medicine, Weill Cornell Medicine, New York, NY 10065, USA

³Englander Institute for Precision Medicine of Weill Cornell Medicine and New York-Presbyterian, New York, NY 10065, USA

⁴Institute for Computational Biomedicine, Weill Medical College, New York, NY 10065, USA

⁵Department of Medicine, Division of Hematology and Medical Oncology, Weill Cornell Medicine, New York, NY 10065, USA

⁶Center of Epigenetics Research, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA

⁷Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA

⁸Departments of Neurology, Albert Einstein College of Medicine, Bronx, NY 10461, USA

⁹Departments of Genetics, Albert Einstein College of Medicine, Bronx, NY 10461, USA

¹⁰Departments of Neuroscience, Albert Einstein College of Medicine, Bronx, NY 10461, USA

¹¹Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA

¹²Department of Pathology, New York University School of Medicine, New York, NY 10016, USA

¹³Yale School of Medicine, New Haven, CT 06511, USA

¹⁴Lead Contact

*Correspondence: chip@mskcc.org (P.C.), cheny1@mskcc.org (Y.C.)

<https://doi.org/10.1016/j.ccell.2017.10.008>

SUMMARY

Prostate cancer exhibits a lineage-specific dependence on androgen signaling. Castration resistance involves reactivation of androgen signaling or activation of alternative lineage programs to bypass androgen requirement. We describe an aberrant gastrointestinal-lineage transcriptome expressed in ~5% of primary prostate cancer that is characterized by abbreviated response to androgen-deprivation therapy and in ~30% of castration-resistant prostate cancer. This program is governed by a transcriptional circuit consisting of HNF4G and HNF1A. Cistrome and chromatin analyses revealed that HNF4G is a pioneer factor that generates and maintains enhancer landscape at gastrointestinal-lineage genes, independent of androgen-receptor signaling. In HNF4G/HNF1A-double-negative prostate cancer, exogenous expression of HNF4G at physiologic levels recapitulates the gastrointestinal transcriptome, chromatin landscape, and leads to relative castration resistance.

INTRODUCTION

The prostate gland is an androgen-dependent male reproductive organ. Upon oncogenic transformation, prostate cancer retains a remarkable lineage-specific dependence on androgen-receptor (AR) signaling, and is characterized by an almost universal

initial response to androgen-deprivation therapy (ADT) (Huggins and Hodges, 1941). However, the depth and duration of response is highly variable with eventual progression to the lethal castration-resistant prostate cancer (CRPC). Multiple mechanisms that contribute to castration resistance have been elucidated. One class reactivate AR signaling in the castrate

Significance

Lineage-directed therapy using androgen deprivation has been the mainstay of prostate cancer treatment for 70 years, and can be circumvented by activation of survival programs of alternative lineages. We found that 5% of primary untreated prostate cancers, and 30% of castration-resistant prostate cancers, aberrantly express a gastrointestinal-lineage (PCa-GI) transcriptome. This PCa-GI transcriptome is regulated by master regulators HNF1A and HNF4G. Integrative cistrome, transcriptome, and chromatin analysis shows that HNF4G is a pioneer factor that reprograms the enhancer chromatin landscape and mediates AR therapy resistance.

environment, such as by AR mutations, AR splice variants, AR amplification, or aberrant expression of the glucocorticoid receptor (GR) that activates AR target genes (Antonarakis et al., 2014; Arora et al., 2013; Chen et al., 2004; Taplin et al., 1995; Watson et al., 2015). Alternatively, increased cellular plasticity to bypass the prostate lineage-specific dependence on AR signaling, best exemplified by neuroendocrine transdifferentiation, is increasingly appreciated as a mechanism of castration resistance (Beltran et al., 2016). Nevertheless, even with next-generation AR pathway inhibitors, the majority of CRPC remain histologically adenocarcinoma (Robinson et al., 2015).

Cancer outlier gene expression analysis identified *SPINK1* overexpression in ~10% of all primary prostate cancers, representing a distinct subtype among those prostate cancers without ETS fusion (Tomlins et al., 2008). While prognostic role of *SPINK1* overexpression in resected primary prostate cancer is controversial (Flavin et al., 2014; Tomlins et al., 2008), studies indicate that *SPINK1* overexpression is associated with more rapid progression to castration resistance. In a Finish cohort of prostate cancer patients treated with primary hormone therapy, *SPINK1* immunohistochemistry (IHC) positivity in prostate cancer was associated with a significantly more rapid progression to castration resistance (Leinonen et al., 2010). In a Chinese cohort diagnosed with metastatic prostate cancer to the bone, *SPINK1* IHC positivity was associated with decreased progression-free survival after hormone therapy (Pan et al., 2016). In a Hopkins cohort of intermediate and high-risk prostate cancer patients treated with prostatectomy who subsequently recurred, *SPINK1* expression was associated with more rapid progression to metastasis and to death (Johnson et al., 2016).

SPINK1, also known as pancreatic secretory trypsin inhibitor, protects the gastrointestinal (GI) tract from protease degradation and its expression is normally restricted to the GI organs. The mechanism by which this GI-restricted gene is expressed in prostate cancer is unknown and suggests that an alternative GI-lineage transcriptome is activated and may lead to decreased dependence on AR signaling. In this study we have explored the mechanism and significance of activation of this GI-lineage transcriptome in prostate cancer tumorigenesis and castration resistance.

RESULTS

HNF4G and HNF1A Regulate a GI Transcriptome in *SPINK1*-Positive Prostate Cancer

Since normal *SPINK1* expression is confined to GI tissues, we compared *SPINK1* expression in normal prostate and prostate cancer with that in normal and malignant GI tissues using The Cancer Genome Atlas (TCGA) cohorts. *SPINK1* is highly expressed in normal GI tissues and preserved in corresponding GI cancers. In prostate, *SPINK1* demonstrates an outlier overexpression pattern, with a subset of cancers expressing high levels comparable with GI tissues, while most prostate cancers and normal tissue express very low levels (Figure S1A). To determine whether *SPINK1* overexpression in prostate cancer is part of an aberrant transcriptome, we examined three large high-quality prostate cancer gene expression datasets to identify genes whose expression is correlated with *SPINK1*

(Beltran et al., 2016; Cancer Genome Atlas Research Network, 2015; Taylor et al., 2010). We generated a transcriptome signature of *SPINK1*-correlated genes, consisting of 129 genes that were in the top 500 most correlated genes in two of three gene sets (Figure 1A, Table S1). Examination of normal tissue RNA sequencing data from Genotype Tissue Expression GTEx (GTEx Consortium, 2015) demonstrated that the expression of *SPINK1*-correlated genes is not high in the normal prostate (Figure 1B). Instead, similar to *SPINK1* itself, these genes are enriched in tissues of the GI tract including the liver and intestines. The signature included well-known GI genes such as albumin (*ALB*), complement factor 5 (*C5*), coagulation factor V (*F5*), vitamin D binding protein (*GC*), prealbumin (*TTR*), growth arrest-specific 2 (*GAS2*), and multiple UDP-glucuronosyltransferase and aldo-keto reductase genes involved in detoxification and steroid metabolism (Table S2). This analysis suggests that, in prostate cancer, the outlier expression of *SPINK1* is part of an aberrantly activated GI transcriptome and we thus name the *SPINK1*-correlated genes as the PCa-GI signature.

To identify master regulator transcription factors that may be responsible for aberrant expression of the GI transcriptome in prostate cancer, we noted the presence of hepatocyte nuclear factor 4-gamma gene (*HNF4G*) and hepatocyte nuclear factor 1-alpha gene (*HNF1A*) in the PCa-GI signature (Figure 1B). *HNF4G* is highly homologous to the well-known master regulator hepatocyte nuclear factor 4-alpha (*HNF4A*), and can bind and transactivate a similar set of genes in hepatocytes (Daigo et al., 2011; Parviz et al., 2003). In liver and pancreas, *HNF4A*, *HNF1A*, and *FOXA* (also known as *HNF3*) family transcription factors form a core GI transcriptome regulatory circuit, where they reinforce each other's expression to maintain lineage specification (Odom et al., 2004, 2006). Indeed, exogenous expression of *HNF4A* or *HNF1A* with a *FOXA* member can reprogram murine fibroblasts to gut endoderm, which forms functional liver and colon depending on area of engraftment (Huang et al., 2011; Sekiya and Suzuki, 2011). Examination of *HNF1A* and *HNF4G* in TCGA cohorts of prostate and GI cancers revealed that their expression mirrored that of *SPINK1*, with high endogenous expression in GI tissues and aberrant overexpression in a subset of prostate cancers to levels comparable with GI tissues (Figures S1B and S1C).

To determine the role of *HNF4G* and *HNF1A* in regulation of the PCa-GI transcriptome, we employed 22Rv1 prostate cancer cells, previously characterized to express high levels of *SPINK1*, for our studies (Tomlins et al., 2008). 22Rv1 cells notably also expressed high levels of *HNF4G* and *HNF1A*. To identify *HNF4G*-dependent genes, we generated two 22Rv1 derivatives with doxycycline-inducible *HNF4G* hairpins (*HNF4Gsh1-Dox* and *HNF4Gsh2-Dox*) and a control derivative with doxycycline-inducible scrambled hairpin (*SCR-Dox*). Knock down of *HNF4G* with doxycycline treatment decreased protein and/or mRNA levels of *HNF1A*, as well as several selected PCa-GI genes including albumin, *SPINK1*, *GAS2*, *MUC13*, and *AKR1C3* (Figures 1C and S1D). Next, to identify *HNF1A*-dependent genes, we transduced 22Rv1 cells with three lentiviral hairpins against *HNF1A*. *HNF1A* knockdown decreased protein and/or mRNA levels of *HNF4G*, as well as the same selected PCa-GI genes (Figures 1D and S1E).

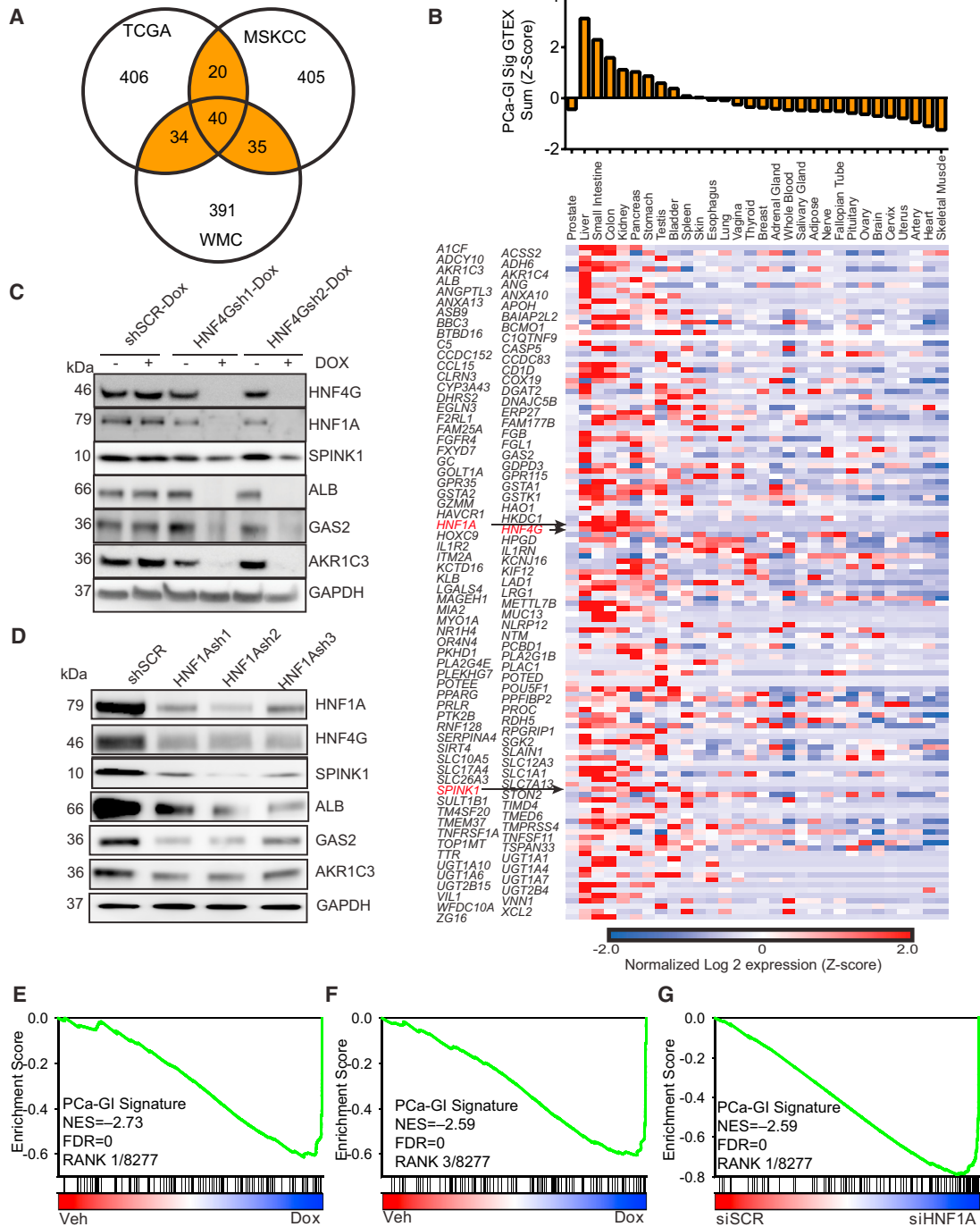


Figure 1. HNF4G and HNF1A Regulate a GI Gene Signature in SPINK1-Positive Prostate Cancer

(A) Venn diagram generated using top 500 genes whose expressions most correlated with *SPINK1* expression in three different gene expression datasets. Highlighted is the PCa-GI signature of 129 genes that are common in two of three datasets. The 40 genes that are in all three sets are called the core PCa-GI signature.

(B) Heatmap of RNA-seq gene expression of the 129 individual *SPINK1* correlated genes in normal tissues from GTEX, expressed as Z score of log2 of read-per-kilobase mapped (RPKM). Top panel shows the sum expression of the 129 genes (Z score).

(C) Immunoblot against indicated PCa-GI signature proteins of indicated derivatives of 22Rv1 cells treated with vehicle or doxycycline for 72 hr.

(D) Immunoblot against indicated PCa-GI signature proteins of 22Rv1 cells 72 hr after transduction with lentiviral shRNAs against HNF1A (HNF1Ash1, HNF1Ash2, and HNF1Ash3) and a scrambled control (shSCR).

(legend continued on next page)

To define the global HNF4G-regulated transcriptome, we performed gene expression profiling of doxycycline treatment in 22Rv1 cells stably expressing doxycycline-inducible hairpins. PCa-GI signature genes are among the most downregulated genes, including *ALB*, *SPINK1*, *MUC13*, *C5*, *GC*, and *TMED6* (Figure S1F). To study the perturbation induced by HNF4G knockdown in an unbiased fashion, we employed gene set enrichment analysis (GSEA), using >8,200 gene sets from the Molecular Signatures Database (Subramanian et al., 2005) and the PCa-GI signature. For both HNF4Gsh1-Dox and HNF4Gsh2-Dox cells, the PCa-GI signature gene set was among the most significantly enriched gene set downregulated by doxycycline treatment (Figures 1E and 1F; Table S3). Next, to define the global HNF1A regulated transcriptome, we performed gene expression profiling of 22Rv1 cells with small interfering RNA (siRNA)-mediated knock down of HNF1A (siHNF1A) or scrambled control (siSCR). As with HNF4G knockdown, many PCa-GI signature genes were among the most downregulated following HNF1A knockdown (Figure S1F). GSEA showed that the PCa-GI signature gene set was the most significantly enriched gene set downregulated by HNF1A siRNA (Figure 1G; Table S4). The PCa-GI signature included another transcription factor nuclear receptor subfamily 1 group H member 4 (*NR1H4*), which was expressed in 22Rv1 cells. NR1H4 knockdown using two different lentiviral hairpins did not show any decrease in select PCa-GI genes, instead we found it to be a downstream target of HNF4G and HNF1A (Figure S1G). These data indicate that HNF1A and HNF4G form a transcriptional regulatory circuitry to reinforce each other to regulate aberrant expression of a GI transcriptome in prostate cancer.

There are no other prostate cancer cell lines that express the PCa-GI transcriptome. We have generated a bank of patient-derived prostate cancer organoids from biopsy specimens of patients with castration-resistant metastatic prostate cancer (Gao et al., 2014). Among these, MSK-PCa10 expresses high levels of HNF4G and HNF1A. To determine if HNF1A and HNF4G regulate a similar PCa-GI-transcriptional program in MSK-PCa10 organoids, we knocked down HNF4G and HNF1A using two different hairpins each. We found that HNF4G and HNF1A regulate the expression of these GI transcriptome genes in MSK-PCa10 prostate cancer cells (Figure S1H). This observation further confirms the role of both these factors in governing the aberrantly activated GI-transcriptional program in prostate cancer.

HNF4G/HNF1A Axis Is Required for Growth in PCa-GI-Positive Prostate Cancer

We examined the requirement of HNF4G and HNF1A for *in vitro* and *in vivo* growth of prostate cancers that express the PCa-GI transcriptome. Downregulation of either HNF1A or HNF4G, each using two independent small hairpin RNA (shRNA) sequences, in 22Rv1 cells caused significant growth suppression compared with scrambled shRNA (Figures 2A and 2B). Consis-

tently, GSEA of transcriptomes from HNF4G or HNF1A knockdown revealed that multiple cell-cycle gene sets were enriched among downregulated genes (Tables S3 and S4).

To investigate the effect of complete depletion of HNF4G and HNF1A, we performed CRISPR/Cas9-mediated knock out of *HNF4G* or *HNF1A* in 22Rv1. Due to outgrowth of cells that escaped knock out in our preliminary experiments, we employed growth competitions assays. 22Rv1-Cas9 cells were then transduced with dual expression vector containing GFP and CRISPR guide RNAs. We verified Cas9-mediated genome editing by next-generation amplicon-sequencing (Figures S2A and S2B) and immunoblotting of HNF4G, HNF1A, and downstream targets (Figures S2C and S2D). Next, we transduced 22Rv1-Cas9 cells at MOI of ~0.4. The relative growth of GFP-positive sgRNA-expressing cells compared with the GFP-negative cells was monitored over time by fluorescence-activated cell sorting analyses (Figure S2E). We observed depletion of GFP-positive sgHNF4G- or sgHNF1A-expressing cells compared with GFP-negative control cells, but not of the sgNTC-expressing GFP-positive cells (Figure S2F), confirming that HNF4G or HNF1A loss leads to significant growth defect in prostate cancer cells.

To further examine the growth inhibitory effect of HNF4G and HNF1A on PCa-GI-positive prostate cancer, we utilized two CRPC prostate cancer organoid lines: MSK-PCa10 (HNF4G⁺ and HNF1A⁺) and MSK-PCa1 (HNF4G⁻ and HNF1A⁻) (Gao et al., 2014; Karthaus et al., 2014). We found that knock down of HNF4G and HNF1A resulted in significant growth suppression in MSK-PCa10, but not in MSK-PCa1 (Figures 2C and 2D).

To determine the role of HNF4G in tumorigenesis *in vivo*, we employed 22Rv1 lines HNF4Gsh1-Dox, HNF4Gsh2-Dox, and SCR-Dox (see Figure 1C). When mice were treated with doxycycline drinking water beginning the same day of grafting, we observed that HNF4Gsh1-Dox and HNF4Gsh2-Dox grafts grew significantly slower than SCR-Dox grafts (Figure 2E). Tumors explanted at the end of the experiment exhibited a decrease in HNF4G and its target proteins AKR1C3, GAS2, and SPINK1 in HNF4Gsh1-Dox and HNF4Gsh2-Dox xenografts compared with SCR-Dox xenografts (Figure 2F). To determine the requirement of HNF4G on tumor growth and maintenance in established tumors, we allowed HNF4Gsh1-Dox, HNF4Gsh2-Dox, and SCR-Dox xenografts to reach a size of 100 mm³ and then started doxycycline or vehicle treatment in drinking water. Tumors explanted 2 days after doxycycline or vehicle administration show that the xenografts retained doxycycline-induced knock down of HNF4G (Figure S2G). In SCR-Dox xenografts, doxycycline had no significant effect on tumor growth. However, in HNF4Gsh1-Dox and HNF4Gsh2-Dox grafts, doxycycline treatment caused a significant reduction in tumor growth rate (Figure 2G). These data indicate that, in the subset of prostate cancers with aberrantly expressed HNF4G/HNF1A transcriptional circuitry, maintaining the circuitry is required for prostate cancer cell growth and tumorigenesis.

(E and F) GSEA plot of PCa-GI gene signature in 22Rv1-HNF4Gsh1-Dox cells (E) or 22Rv1-HNF4Gsh2-Dox cells (F) treated with doxycycline compared with vehicle. NES, normalized enrichment score; FDR, false discovery rate.

(G) GSEA plot of PCa-GI gene signature in 22Rv1 cells transfected with HNF1A siRNA compared with scrambled siRNA. See also Figure S1 and Tables S1, S2, S3, and S4.

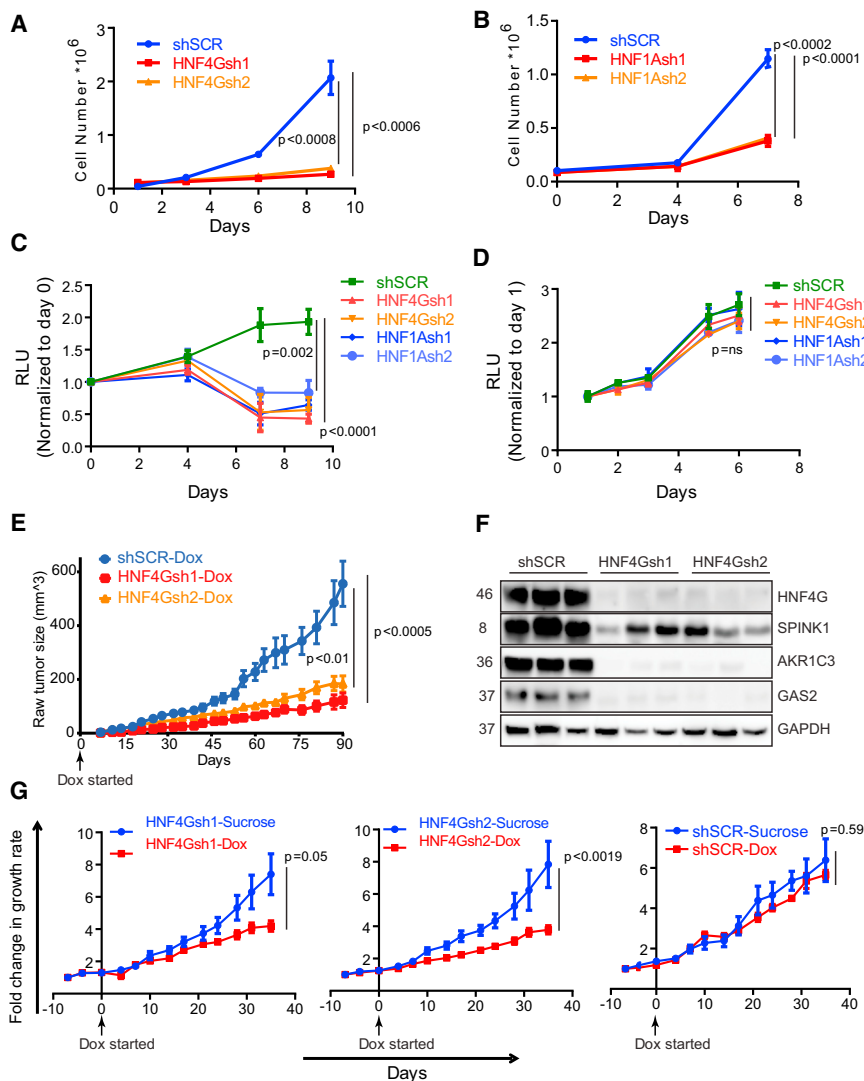


Figure 2. SPINK1-Positive Prostate Cancer Requires HNF4G/HNF1A Axis for Growth

(A and B) Cell growth curve of 22Rv1 following shRNA-mediated HNF4G (A) or HNF1A (B) knockdown and control. Mean \pm SD. Two-tailed unpaired t test, $n = 3$.

(C and D) Cell growth curve of human patient-derived CRPC cell lines MSK-PCa10 (C) and MSK-PCa1 (D) following shRNA-mediated HNF4G and HNF1A knockdown and control. Mean \pm SD. Two-tailed unpaired t test, $n = 3$.

(E) Tumor formation and growth rate of indicated 22Rv1 cells when mice were fed with doxycycline drinking water beginning the same day as grafting. 2.0×10^6 cells were subcutaneously injected into 6–8-week-old CB17-SCID mice; $n = 10$ for all groups. Mean \pm SEM. Two-tailed unpaired t test. (F) Immunoblots of three representative 22Rv1 explants obtained at the end of the experiment shown in (E).

(G) Response of indicated 22Rv1 xenograft tumors in SCID mice upon starting doxycycline water diet when tumors reached approximately 100 mm³; for shSCR-sucrose and shSCR-DOX, $n = 6$; for HNF4Gsh1-sucrose and HNF4Gsh1-DOX, $n = 8$; for HNF4Gsh2-sucrose and HNF4Gsh2-DOX, $n = 6$ and 8, respectively. Fold change in tumor volume over day 0 (start of doxycycline water) is plotted. Mean \pm SEM. Two-tailed unpaired t test.

See also Figure S2.

HNF4G Has a Distinct Cistrome from AR and Maintains Enhancer Chromatin Features at Its Binding Sites in Prostate Cancer

The cellular landscape of enhancers, hubs of permissive chromatin where multiple transcription factors bind, is highly lineage specific and reflects the cell-type-specific gene expression patterns (Heintzman et al., 2009). While many transcription factors bind to existing enhancers to modulate transcription, some master regulators, commonly referred to as “pioneer factors,” can generate and maintain the enhancer sites *de novo* (Zaret and Carroll, 2011). In the prostate lineage, FOXA1 is a well-established pioneer factor that shapes the enhancer landscape and AR cistrome (Iwafuchi-Doi et al., 2016; Lupien et al., 2008; Wang et al., 2011). To determine how HNF4G regulates the GI transcriptome and the interplay of HNF4G, FOXA1, and AR signaling in prostate cancer pathogenesis, we mapped HNF4G, FOXA1, and AR genome-wide binding sites using chromatin immunoprecipitation sequencing (ChIP-seq) and performed cistrome analyses with and without doxycycline-mediated HNF4G knockdown in 22Rv1-HNF4Gsh2-Dox cells.

At baseline (vehicle-treated HNF4Gsh2-Dox cells), we identified $\sim 9,500$ high-confidence ($q < 10^{-5}$) HNF4G peaks, with 90% in enhancer (non-promoter) regions and 10% in promoter regions, consistent with previous observations for the homolog HNF4A (Wallerman et al., 2009) (Figure S3A). Approximately 17% of HNF4G peaks overlapped with FOXA1 peaks, and $\sim 13\%$ overlapped with AR peaks (Figure S3B). In contrast, $\sim 50\%$ of $\sim 24,000$ AR peaks overlapped with FOXA1 (Figure S3B), consistent with prior observations and the prominent role of FOXA1 in defining the AR cistrome (Lupien et al., 2008; Wang et al., 2011). The *de novo* motif of top HNF4G peaks identified the HNF4 motif, centered at the peak summit, consistent with specific HNF4G ChIP (Figure S3C). Similarly, the *de novo* motif of the top AR peaks identified the AR motif, centered at the peak summit. While the FOXA1 motif was the second most enriched motif around both HNF4G and AR peaks, it was substantially more significant in the AR cistrome than the HNF4G cistrome (Figure S3C).

We next examined the effect of HNF4G depletion. At baseline, top HNF4G (blue) and top AR (green) binding sites are largely

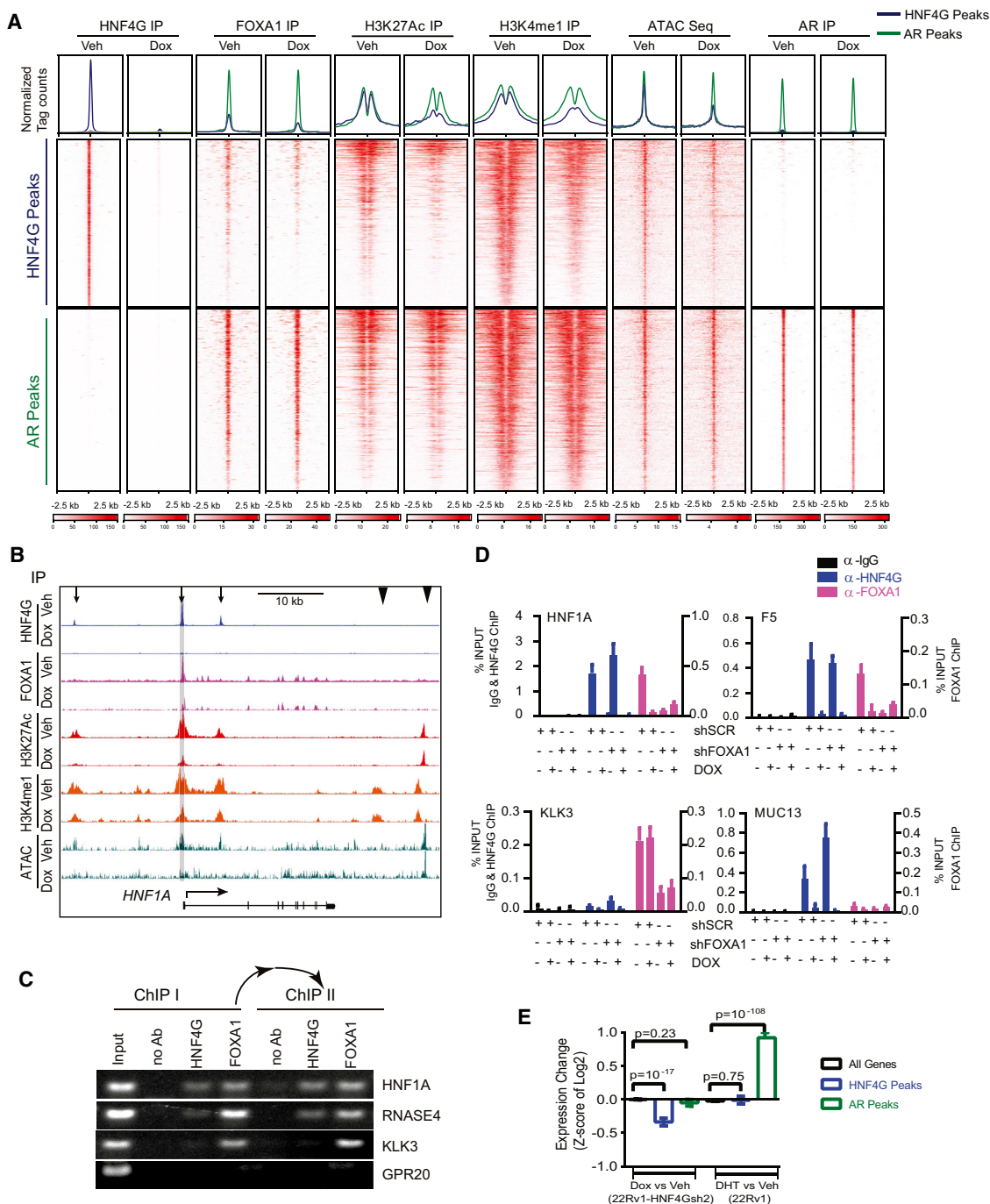


Figure 3. HNF4G Binding Maintains Enhancer Chromatin at Binding Sites

(A) Histograms (top) show the average normalized tag counts of HNF4G, FOXA1, H3K27Ac, H3K4me1, and AR ChIP-seq, as well as ATAC-seq, in vehicle and doxycycline-treated 22Rv1-HNF4Gsh2-Dox cells at HNF4G (blue) and AR (green) binding sites. Heatmap shows the tag densities of HNF4G, FOXA1, H3K27Ac, H3K4me1, AR, and ATAC signal at the top 1,000 HNF4G (middle) or AR (bottom, as internal control) binding sites upon vehicle or doxycycline treatment in 22Rv1-HNF4Gsh2-Dox cells.

(B) ChIP-seq and ATAC-seq profiles of HNF4G, FOXA1, H3K27Ac, and H3K4me1 at the *HNF1A* locus with or without HNF4G knockdown. Arrows indicate enhancers with HNF4G peaks and arrowheads indicate control enhancers without HNF4G peaks. Locus used for ChIP-qPCR is highlighted.

(C) ChIP-re-ChIP showing co-binding of HNF4G and FOXA1 at select HNF4G/FOXA1 co-binding loci (*HNF1A* and *RNASE4*), as well as HNF4G non-occupied locus (*KLK3*) and a HNF4G and FOXA1 non-occupied locus (*GPR20*) as controls. First ChIP was performed in 22Rv1 cells with HNF4G and FOXA1 antibodies and no antibody as a control. Sequential HNF4G and FOXA1 ChIP were then performed with eluates from first ChIP of FOXA1. Input is 0.1% for first ChIP, FOXA1 first ChIP is 10% for subsequent second ChIPs.

(legend continued on next page)

distinct with little overlap (Figure 3A), with a higher percentage of AR binding sites compared with HNF4G binding sites also bound by FOXA1. Both top HNF4G binding sites and top AR binding sites exhibit enrichment for H3K4me1, which mark enhancers, and the majority of these sites also exhibit enrichment of H3K27ac, suggesting that they are active. ATAC-seq reveals that both top HNF4G and AR binding sites are at assessable chromatin. Doxycycline-mediated HNF4G knockdown decreased the mean tag densities of both H3K27ac and H3K4me1 at HNF4G binding sites, with a shift from a bimodal profile to a more unimodal profile. This change in profile suggests loss of the central nucleosome-depleted regions with HNF4G downregulation (Figure 3A, blue) (He et al., 2010). Consistently, HNF4G knockdown also decreased ATAC signal at HNF4G binding sites, indicating decreased chromatin accessibility. In contrast, HNF4G knockdown did not affect the H3K27ac and H3K4me1 ChIP profiles or the ATAC-seq profile at top AR binding peaks (Figure 3A, green).

Examination of ChIP-seq profiles at representative HNF4G target genes *HNF1A*, *F5*, *CLRN3*, *GAS2*, and *MUC13* illustrates a general diminishment of H3K4me1, H3K27ac, ATAC, and FOXA1 binding at many HNF4G binding sites with doxycycline treatment (Figures 3B and S3D, arrows, S3E), whereas non-HNF4G binding sites showed little change (Figures 3B and S3D, arrowheads). Among the spectrum of HNF4G binding sites, some are characterized by loss of H3K4me1 and ATAC signal (*HNF1A*, *CLRN3*) suggesting that HNF4G is required for enhancer maintenance and that some are characterized by loss of H3K27ac (*GAS2*, *RNASE4*) with preservation of H3K4me1, suggesting that HNF4G is required for enhancer activation.

At HNF4G and FOXA1 co-bound sites, FOXA1 binding is decreased in some (*HNF1A*, *GAS2*, and *F5*) and preserved in others (*RNASE4*) (Figures 3B and S3D, arrows). Given the well-established role of FOXA1 as a pioneer factor in the prostate lineage, we were surprised by the requirement of HNF4G for FOXA1 binding, suggesting that the presence of FOXA1 alone is insufficient for maintaining open chromatin at these sites. Using ChIP-reChIP, we confirmed that genomic sites with both HNF4G peaks and FOXA1 peaks were indeed co-bound in the same cell by HNF4G and FOXA1 (Figure 3C). We performed combinatorial depletion of HNF4G (using doxycycline versus vehicle) and FOXA1 (using shSCR versus shFOXA1 lentiviral hairpins) to determine their effect on HNF4G and FOXA1 DNA binding using ChIP-qPCR. HNF4G and FOXA1 did not regulate each other, and knockdown efficiency of one gene was unaffected by the other (Figure S3F). At the *KLK3* site bound only by FOXA1, and the *MUC13* site bound only by HNF4G, depletion of FOXA1 and HNF4G depleted FOXA1 and HNF4G binding, respectively (Figure 3D). At the *HNF1A* and *F5* co-bound sites (Figures 3B and S3D), depletion of HNF4G significantly decreased FOXA1 binding, whereas depletion of FOXA1 did

not decrease HNF4G binding (Figure 3D). These data indicate that HNF4G is required to maintain enhancer chromatin at certain HNF4G binding sites, as well as recruitment of other transcription factors, including FOXA1. Co-immunoprecipitation experiments failed to detect stable interaction between HNF4G and FOXA1, suggesting that the two transcription factors do not form stable complexes prior to DNA binding (Figure S3G).

We next performed integrative analysis of the HNF4G and AR cistrome with the HNF4G- and AR-regulated transcriptome. Compared with all genes, genes mapped to HNF4G peaks were significantly downregulated after HNF4G knockdown ($p < 10^{-17}$), but unaffected by DHT treatment ($p = 0.75$), while genes mapped to AR peaks were significantly upregulated by DHT treatment ($p < 10^{-108}$), but unaffected by HNF4G knockdown ($p = 0.23$) (Figure 3E). These data indicate that HNF4G is required to maintain enhancer chromatin context and allow binding of other transcription factors including FOXA1 at certain HNF4G binding sites and regulate transcriptional targets. Further, HNF4G binds to and maintains a cistrome that regulates a PCa-GI transcriptome distinct from AR signaling and contributes to prostate cancer oncogenesis.

Exogenous HNF4G or HNF1A Expression Activates the PCa-GI Gene Signature and Chromatin Signature, Independent of AR Signaling

The SPINK1-positive subset of prostate cancers have been shown to progress more rapidly to castration resistance in multiple studies (Johnson et al., 2016; Leinonen et al., 2010; Pan et al., 2016), and several other PCa-GI signature genes including *AKR1C3* and *UGT2B15* are clinical biomarkers of castration resistance (Stanbrough et al., 2006). We therefore asked whether activation of the PCa-GI-signature by HNF4G/HNF1A regulatory circuitry can alter AR dependence and lead to castration resistance.

First, to determine if exogenous HNF4G or HNF1A expression can recapitulate the GI-lineage transcriptome in prostate cancer, we stably expressed HNF4G or HNF1A in the LNCaP prostate cancer cell line that is SPINK1-negative and harbors an *ETV1* translocation (Chen et al., 2013; Tomlins et al., 2007). We observed that exogenous expression of HNF4G resulted in expression of the endogenous HNF1A and vice versa. Further, exogenous expression of either transcription factor resulted in upregulation of PCa-GI signature genes including *AKR1C3*, *MUC13*, *TMED6*, *SPINK1*, and *UGT2B15* (Figures 4A and 4B). To characterize the global transcriptome in response to HNF4G or HNF1A expression, we performed gene expression profiling. GSEA revealed that, for each transcriptome, the PCa-GI gene signature was among the most enriched gene sets (Figures 4C, 4D, and S4; Table S5). Other enriched gene sets include steroid metabolism genes, liver- and pancreas-specific genes, and known HNF4A- and HNF1A-dependent genes. These data indicate that expression of either HNF1A or

(D) ChIP-qPCR of HNF4G and FOXA1 at selected HNF4G and FOXA1 co-binding loci (*HNF1A* and *F5*), HNF4G alone locus (*MUC13*), and FOXA1 alone locus (*KLK3*) in 22Rv1 cells. For each bar graph: left axis is fold enrichment over input for IgG, HNF4G ChIP, and right axis is fold enrichment over input for FOXA1 ChIP. Mean \pm SD, $n = 3$.

(E) Bar graph of gene expression change by HNF4G knockdown (dox treatment of 22Rv1-HNF4Gsh2-Dox) or AR activation (DHT treatment of 22Rv1) of all genes (black), genes mapped to top 1,000 HNF4G peaks (blue) and top 1,000 AR peaks (green). Mean \pm 95% confidence. Two-tailed unpaired t test. See also Figure S3.

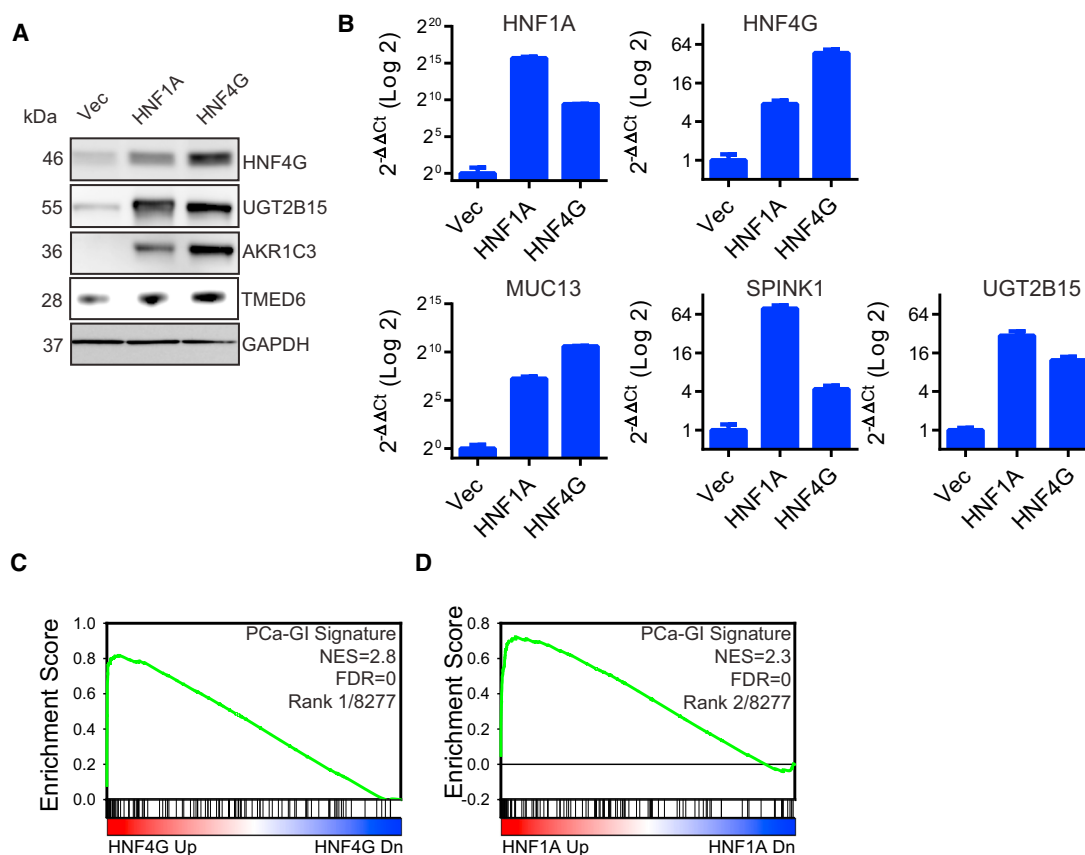


Figure 4. Exogenous HNF4G or HNF1A Expression Recapitulates the PCa-Gl Signature

(A) Immunoblots of indicated proteins in LNCaP cells transduced for stable expression of HNF4G, HNF1A, or empty vector control against the indicated proteins. (B) qRT-PCR showing the expression of selected PCa-Gl signature genes after exogenous expression of HNF4G and HNF1A in LNCaP cells. Data are presented as mean \pm SD.

(C and D) GSEA plot of PCa-Gl signature in LNCaP cells exogenously expressing HNF4G (C) or HNF1A (D) compared with empty vector control. NES, normalized enrichment score; FDR, false discovery rate.

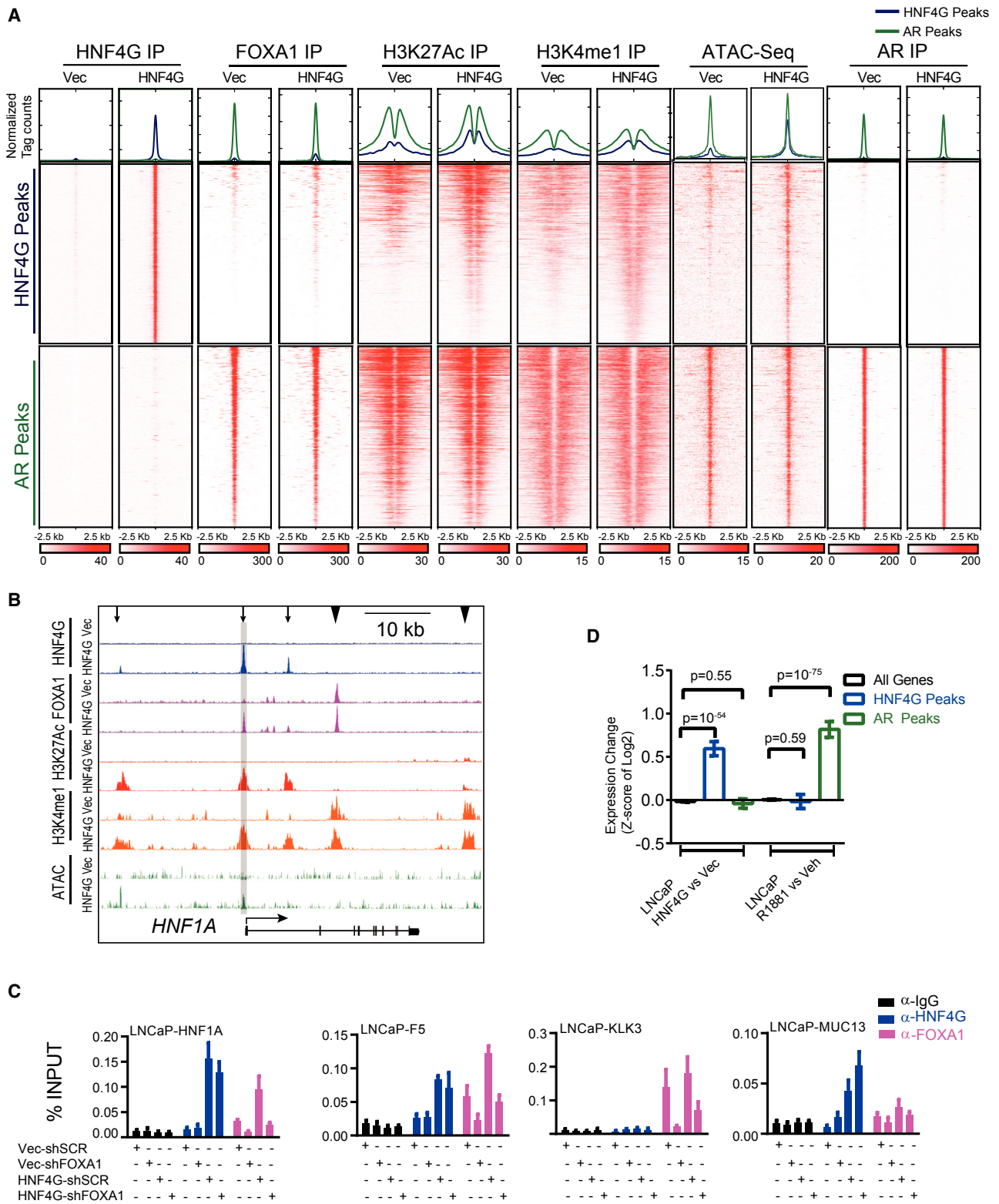
See also Figure S4 and Table S5.

HNF4G in prostate cancer cells can activate the HNF1A/HNF4G circuitry to express the PCa-Gl gene signature.

To determine the effect of activation of HNF1A/HNF4G circuitry on the chromatin enhancer landscape, chromatin accessibility, and the FOXA1 and AR cistrome, we mapped the global localization of HNF4G, AR, FOXA1, H3K4me1, and H3K27ac by ChIP-seq and ATAC-seq in LNCaP cells stably expressing HNF4G versus vector control. In LNCaP cells, there were a large number of FOXA1 peaks (>150,000) and the majority of AR co-localized with FOXA1 (Figure S5A), consistent with prior reports (Jin et al., 2014; Wang et al., 2011). With exogenous HNF4G expression, 34% of HNF4G peaks co-localized with FOXA1 peaks, and 6.5% of HNF4G peaks with AR peaks (Figure S5A). We compared the HNF4G, AR, and FOXA1 peaks of LNCaP-HNF4G cells with 22Rv1-HNF4Gsh2-Dox vehicle-treated cells. We found that ~70% of induced HNF4G peaks in LNCaP cells overlapped with endogenous peaks in 22Rv1 cells, suggesting that we have faithfully recapitulated the HNF4G cistrome (Figure S5B). *De novo* motif analysis of the top 1,000 HNF4G peaks and AR peaks found that the most enriched motifs were HNF4 and AR, respectively, centered at peak center. The

FOXA1 motif was the next most enriched motif for both HNF4G and AR binding sites, but is much more prevalent and significant for AR than that of HNF4G binding sites (Figure S5C).

We next examined the effects of HNF4G exogenous expression on the FOXA1 and AR cistrome, as well as on DNA accessibility and enhancer chromatin marks (e.g., H3K4me1 and H3K27ac) at top HNF4G and AR binding sites in LNCaP cells. HNF4G exogenous expression did not affect the AR cistrome, but enhanced FOXA1 binding at a subset of HNF4G sites (Figure 5A). In addition, the mean ATAC signal, and H3K4me1 and H3K27ac profiles, exhibited increased tag density and a shift in peak profile toward a bimodal distribution upon HNF4G exogenous expression in LNCaP cells at top HNF4G binding sites (Figure 5A, blue). In contrast, exogenous HNF4G expression did not affect the ATAC signal, or H3K4me1 and H3K27ac distribution, at top AR binding sites (Figure 5A, green). The ChIP-seq and ATAC-seq profiles at representative HNF4G target genes (*HNF1A*, *CLRN3*, *F5*, and *MUC13*) illustrates gain of enhancer marks and increased chromatin accessibility specifically at sites of HNF4G binding (Figures 5B and S5D). At some sites (*HNF1A* and *CLRN3*), LNCaP cells were completely devoid



(legend continued on next page)

of H3K4me1 or H3K27ac marks, and HNF4G expression generated them *de novo*. Other sites (*F5*, *MUC13*, and *RNASE4*) were pre-marked by H3K4me1, signifying poised enhancer, and HNF4G increased H3K27ac and ATAC signal, suggesting enhancer activation. Upon HNF4G expression, FOXA1 binding was induced (*HNF1A* and *CLRN3*), enhanced (*F5*), or unaffected (*RNASE4*). These patterns of different effects of exogenous HNF4G expression in LNCaP cells were remarkably similar to HNF4G depletion in 22Rv1 cells. We performed combinatorial exogenous HNF4G expression and FOXA1 knockdown, and assayed FOXA1 and HNF4G binding at specific sites by ChIP-qPCR. At control FOXA1-only *KLK3* and HNF4G-only *MUC13* sites, FOXA1 depletion and HNF4G expression resulted in expected changes in binding. At the co-bound sites at *HNF1A* and *F5*, HNF4G expression caused increased FOXA1 binding, whereas FOXA1 knockdown did not significantly affect HNF4G binding (Figures 5C and S5E).

There was a notable absence of HNF4G binding at the *SPINK1* locus, which contains two canonical HNF1 sites in intron 1 and ~7.5 kb upstream of the promoter. The upstream site is notable for decreased H3K27ac and ATAC signal in 22Rv1 cells after HNF4G knockdown, and increase in H3K27Ac and ATAC signal in LNCaP cells after HNF4G expression (Figure S5F). *SPINK1* expression is induced to a higher extent by HNF1A overexpression compared with HNF4G overexpression (Figures S4 and S5G). These data suggest that *SPINK1* is a direct HNF1A target gene.

Integrating with transcriptome, we found that exogenous HNF4G expression significantly increased expression of genes mapped to HNF4G peaks but not AR peaks, whereas androgen treatment significantly upregulated genes mapped to AR peaks, but not HNF4G peaks (Figure 5D). These data indicate that HNF4G can establish enhancers *de novo* and transcriptionally activate the PCa-GI transcriptome independent of AR signaling in prostate cancer.

HNF4G Expression Imparts Resistance to Androgen Ablation and Enzalutamide Treatment in LNCaP Cells

While the prognostic role of *SPINK1* expression in surgically resected early disease is controversial, three independent studies have shown that *SPINK1* overexpression predicts for decreased response to ADT and more rapid progression to castration resistance and cancer-related death in advanced prostate cancer (Johnson et al., 2016; Leinonen et al., 2010; Pan et al., 2016). We further noted that the *SPINK1* correlated genes (i.e., the PCa-GI signature) is enriched for steroid metabolism genes including *AKR1C3* and *UGT2B15*, which are associated with castration resistance (Table S2) (Stanbrough et al., 2006). We therefore examined the role of HNF4G expression in castration resistance, using the established LNCaP/AR prostate cancer

cell line (Arora et al., 2013; Tran et al., 2009). This LNCaP derivative has exogenous AR overexpression to mimic clinically observed AR overexpression seen in CRPC. It readily forms tumors in castrate mice but is sensitive to the second-generation anti-androgen enzalutamide treatment *in vitro* and *in vivo*.

We performed *in vitro* colony formation assays in full serum (FS) and charcoal-stripped serum (CSS) that is depleted of steroid hormones (e.g., androgen). In FS, there was a slight increase in number of colonies formed. In CSS, LNCaP/AR cells expressing HNF4G were able to form significantly more colonies than those expressing empty vector (Figures 6A and S6A).

While exogenous HNF4G expression increased colony formation in CSS, it did not fully restore colony formation to the level of FS. This suggested that HNF4G primed a subset of cells to become castration resistant. We observed that, when grown long-term in CSS, both LNCaP/AR-Vec and LNCaP/AR-HNF4G cells grew slowly for 10 days. Afterward, multiple LNCaP/AR-HNF4G cell clones grew to take over the culture, while LNCaP/AR-Vec cells arrested (Figure 6B). To study the underlying mechanism of castration-resistant growth we performed transcriptome analysis of LNCaP/AR-Vec and LNCaP/AR-HNF4G at day 9 (D9) of growth in CSS and LNCaP/AR-HNF4G at D32 of growth in CSS to identify the HNF4G transcriptome, as well as determinants of castration-resistant growth among bulk cells with exogenous expression of HNF4G (Figure 6B). As expected, expression of PCa-GI signature genes was increased in LNCaP/AR-HNF4G cells compared with LNCaP/AR-Vec cells (Figures 6C and S6B). Notably, at D32 in CSS when LNCaP/AR-HNF4G cells had started to grow rapidly, there was a further increase in PCa-GI signature gene expression (Figures 6C and S6B). GSEA revealed that the PCa-GI signature is significantly and highly enriched, not only among genes upregulated by HNF4G expression compared with vector control after 9 days in CSS (D9 HNF4G versus Vec CSS), but also among genes upregulated by HNF4G expression after 32 days of CSS compared with 9 days of CSS in HNF4G-expressing cells (D32 versus D9 HNF4G CSS) (Figure S6C; Table S6). These data suggest that further upregulation of the PCa-GI gene signature by HNF4G correlates with progression to castration-resistant growth in prostate cancer cells.

To examine the role of HNF4G expression on AR-directed therapy resistance *in vivo*, we utilized the enzalutamide treatment of LNCaP/AR cells grafted into castrate mice, a well-established *in vivo* system in the laboratory (Arora et al., 2013; Balbas et al., 2013; Tran et al., 2009). We grafted LNCaP/AR-Vec and LNCaP/AR-HNF4G cells, and assessed for response to enzalutamide. LNCaP/AR-Vec tumor grafts exhibited an average tumor size reduction of ~50%, and slow progression after ~60 days of treatment similar to a previous report (Arora et al., 2013). In contrast, LNCaP/AR-HNF4G

(B) Representative ChIP-seq and ATAC-seq profiles of HNF4G, FOXA1, H3K27Ac, and H3K4me1 at *HNF1A* locus in LNCaP cells, with or without exogenous HNF4G expression. Arrows indicate enhancers with HNF4G peaks and arrowheads indicate control enhancers without HNF4G peaks. Locus used for ChIP-qPCR is highlighted.

(C) ChIP-qPCR of HNF4G and FOXA1 at select HNF4G and FOXA1 co-binding loci (*HNF1A* and *F5*), as well as HNF4G alone (*MUC13*) and FOXA1 alone (*KLK3*) loci in LNCaP cells. Mean \pm SD, n = 3.

(D) Bar graph of gene expression change by HNF4G expression or AR activation (R1881 treatment) of all genes (black), genes mapped to top 1,000 HNF4G peaks (blue) and top 1,000 AR peaks (green) in LNCaP cells. Mean \pm 95% confidence. Two-tailed unpaired t test.

See also Figure S5.

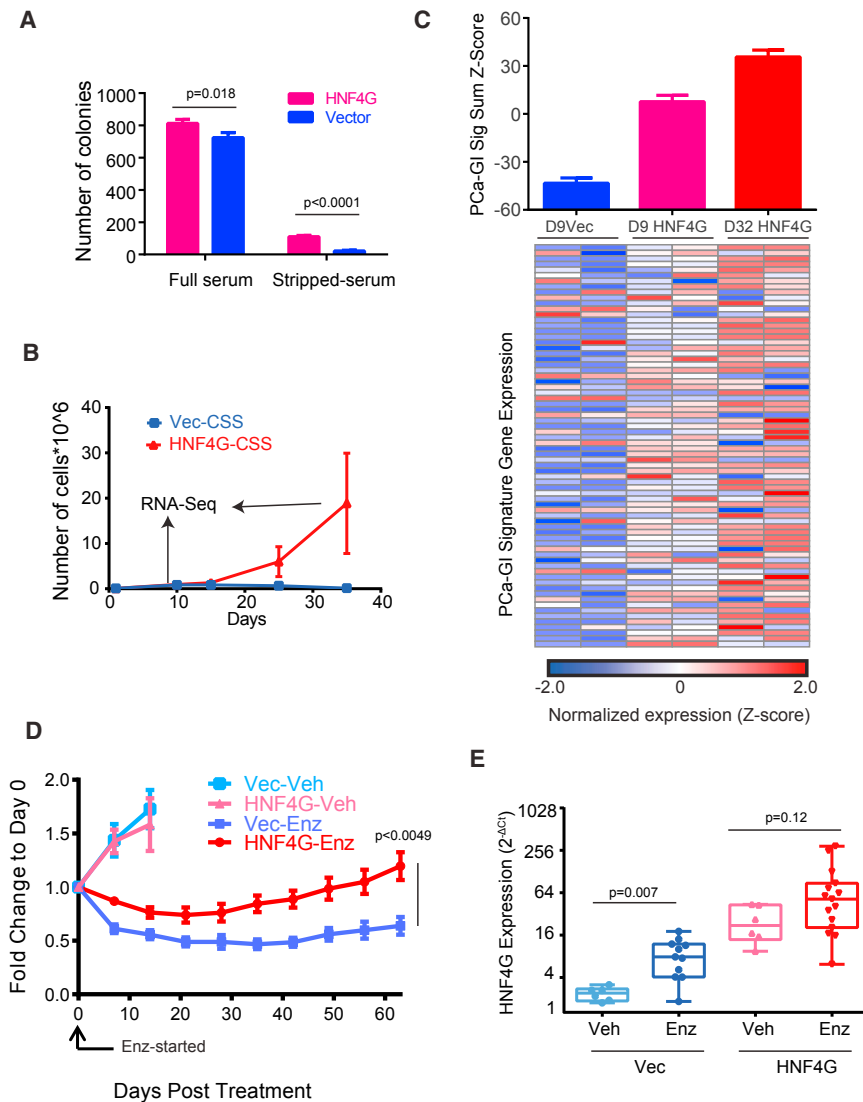


Figure 6. HNF4G Expression Imparts Resistance to Androgen Ablation and Enzalutamide Treatment *In Vitro* and *In Vivo*

(A) Number of colonies formed by LNCaP/AR cells with exogenous expression of HNF4G or vector control in medium with full serum or stripped serum. $n = 3$, Mean \pm SD. Two-tailed unpaired t test.

(B) Growth curve of LNCaP/AR cells exogenously expressing HNF4G or vector control cultured in medium with stripped serum. Arrows shows the time points at which cells were harvested for RNA. $n = 2$, Mean \pm SD.

(C) The sum Z score of individual PCa-GI genes (mean \pm SD) in LNCaP/AR cells expressing vector control at 9 days of growth in CSS medium and HNF4G at day 9 and 32 of growth in CSS medium (top) and heatmap shows the expression of individual SPINK1 signature genes (bottom). $n = 2$.

(D) Treatment response of LNCaP/AR cell xenografts exogenously expressing HNF4G or vector control in SCID mice when treated with enzalutamide (10 mg/kg) or vehicle (1% carboxymethyl cellulose) once a day. For Vec-vehicle and Vec-enzalutamide, $n = 4$ and 13, respectively; for HNF4G-Veh and HNF4G-enzalutamide, $n = 4$ and 18 respectively. Treatment was started when tumors reached a volume of approximately 400 mm³. Fold change in growth rate over day 0 (start of treatment) is shown. Mean \pm SEM. Two-tailed unpaired t test.

(E) Boxplot showing HNF4G mRNA levels of explanted xenografts at the end of the experiment. Box plots show median, quartiles, min, and max, with each sample dot plotted. Statistical analysis was performed using two-tailed unpaired t test. See also Figure S6 and Table S6.

tumor grafts exhibited decreased depth and duration of response, reaching pre-treatment size by 45 days of therapy (Figure 6D). As expected, the *HNF4G* transcript level of explanted tumors at the end of experiment was higher in LNCaP/AR-HNF4G compared with LNCaP/AR-Vec tumor grafts (Figure 6E). Notably, in both LNCaP/AR-Vec and LNCaP/AR-HNF4G grafts, enzalutamide treatment resulted in higher *HNF4G* transcript levels compared with vehicle controls. This suggests that expression of *HNF4G* can be induced in response to AR inhibition, which provides an effective alternative pathway for growth and survival in CRPCs.

Aberrant PCa-GI Activation Is Prevalent in Human CRPC

Since HNF4G expression can lead to activation of PCa-GI expression signature and castration resistance in prostate cancer, we asked if HNF4G and the PCa-signature are activated in human CRPC. In three transcriptome datasets with normal prostate, primary prostate cancer, and CRPC clinical samples (Beltran et al., 2016; Grasso et al., 2012; Taylor et al., 2010), there was both an increase in the percentage of CRPC tumors with

(Figures 7A, 7B, and S7A). Similarly, *HNF1A* is progressively overexpressed from benign prostate, to localized cancer, to CRPC (Figure S7B). To validate this observation, we performed IHC staining of HNF4G on a set of tissue microarrays consisting of benign, primary, and CRPC cases. We found that, in a small percentage of “benign” prostate samples, there were some cells with low HNF4G expression, with the important caveat that the samples are from patients with prostate cancer. As the disease progresses to primary cancer and CRPC, HNF4G increased in both intensity as well as fraction of cases, with ~30% of CRPC cases positive (Figures 7C and 7D).

To examine if HNF4G is transcriptionally active in CRPC, we calculated the PCa-GI score for each sample and found that *HNF4G* expression level and PCa-GI sum expression score was highly correlated (Figures 7E and 7F). These data suggest that HNF4G is not only aberrantly highly expressed and transcriptionally active in a subset of primary prostate cancers that contribute to pathogenesis, but also in a substantial subset of CRPC during clinical progression with AR signaling-targeted therapies.

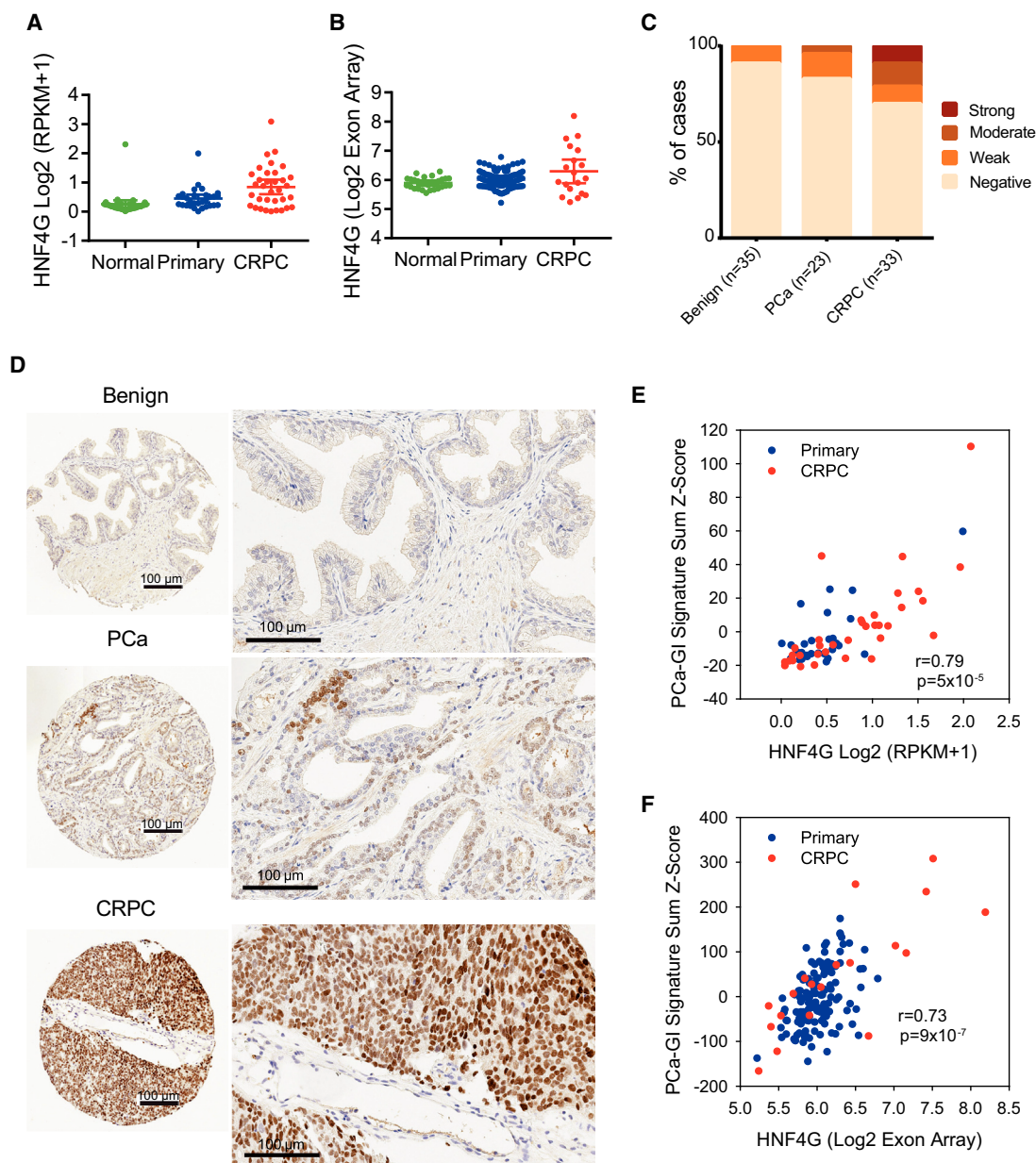


Figure 7. HNF4G Overexpression Is Prevalent in CRPC

(A and B) *HNF4G* expression in normal prostate, primary prostate cancer, and CRPC from the WCMC (A) and the MSKCC (B) datasets. Mean \pm 95% CI.

(C) Quantification of HNF4G nuclear staining by immunohistochemistry (IHC) analysis on tissue microarrays (TMAs) of benign, primary prostate cancer (PCa), and castration-resistant prostate cancer (CRPC) from the WCMC cohort.

(D) Representative images of HNF4G nuclear staining in benign, primary prostate cancer, and CRPC tissue to show negative, weak, and strong HNF4G staining, respectively. Scale bars, 100 μ m.

(E and F) Scatterplot showing correlation between HNF4G expression and PCa-GI signature sum (Z score) in primary and CRPC cases from WCMC dataset (E) and MSKCC dataset (F).

See also [Figure S7](#).

DISCUSSION

During development, cell identity is specified by activation of lineage-specific master regulatory transcription factors, which, in concert with chromatin modifying enzymes, define the enhancer chromatin landscape (Sternberg et al., 2013). Upon

tumorigenic transformation, many cancer types retain gene expression patterns and dependence on master regulators of the cell-of-origin lineage (Bass et al., 2009; Chi et al., 2010; Garraway et al., 2005). Prostate cancer is characterized by dependence on lineage-specific master regulators FOXA1, HOXB13, and AR (Huggins and Hodges, 1941; Pomerantz

et al., 2015), with AR dependence forming the scientific basis and clinical success for ADT.

Here, we have uncovered a mechanism involved in prostate cancer pathogenesis and castration resistance by a HNF4G/HNF1A transcriptional circuitry. HNF4G/HNF1A expression activates PCa-GI signature characteristic of the GI-lineage genes, distinct from the AR-dependent transcriptome, in ~5% of primary prostate cancer and ~30% of CRPC. Prostate cancers that activate this circuitry are dependent on it for growth and survival. Further, exogenous expression of HNF4G to activate this circuit results in more rapid progression to androgen-independent growth *in vitro* and enzalutamide resistance *in vivo*. While exogenous expression of HNF4G does not convey immediate androgen independence to all cells, the cells that eventually propagate in androgen-deplete conditions express higher levels of HNF4G and GI signature genes. These experimental observations correlate with clinical data that SPINK1-positive prostate cancers, which have an activated HNF4G/HNF1A circuit, progress more rapidly to castration resistance (Johnson et al., 2016; Leinonen et al., 2010; Pan et al., 2016). Interestingly, enzalutamide selection pressure *in vivo* increases HNF4G expression. These data are consistent with a prominent role of HNF4G in resistance to AR-targeted therapy. Previous studies of key transcription factors involved in prostate pathogenesis and clinical progression to CRPC, e.g., FOXA1, AR, ETV1/ERG, GR, ROR γ , etc., all involve the activation of AR signaling in primary prostate cancer and reactivation of AR signaling in the presence of ADT. In contrast, HNF4G/HNF1A expression does not alter the AR cistrome or AR signaling. Instead, HNF4G can establish *de novo* enhancers and maintain, as well as augment, pre-existing enhancers at HNF4G binding sites, which can facilitate the binding of other transcription factors and mediate the GI-lineage transcriptome in prostate cancer pathogenesis. Importantly, our data indicate that HNF4G-mediated castration-resistant mechanisms will not only evade the current available ADT and anti-androgen therapies, but also will predict therapeutic resistance to newer generations of therapeutics targeting AR signaling reactivation in CRPCs.

In prostate cancer, ETS translocations serve as a paradigm that aberrant expression of transcription factors, at levels endogenous in other tissue types (e.g., ERG in endothelial cells and ETV1 in interstitial cells of Cajal), can be tumorigenic (Chi et al., 2010; Miettinen et al., 2011; Ran et al., 2015). An important question raised by our work is how *HNF4G* is aberrantly expressed in prostate cancer and whether it is induced in CRPC. IHC studies have identified low HNF4G expression in a small percentage of cells in some benign prostate specimens (with the caveat that they are adjacent to cancerous prostate), raising the possibility that they may be the cell of origin of this subtype of prostate cancer. We further observed that HNF4G and PCa-GI signature genes are induced by androgen deprivation. During development, FOXA1/2 expression specifies the definitive endoderm and establishes competence for subsequent lineage-specific regulators including AR in prostate, HNF4A in liver, and PDX1 in pancreas (Jozwik and Carroll, 2012; Wang et al., 2015). The prostate is derived from the hindgut endoderm and shares expression of master regulators HOXB13 and FOXA1 (Iwafuchi-Doi and Zaret, 2014; Zaret and Carroll, 2011). The suppression of AR activity may therefore activate a latent development program.

HNF4G is a nuclear receptor homologous to retinoid and peroxisome proliferator-activated receptor with fatty acids as its endogenous ligand (Wisely et al., 2002; Yuan et al., 2009). While HNF4A is essential, *Hnf4g* knockout mice exhibit minimal defects (Baraille et al., 2015). These data suggest that HNF4G may be a clinically practical therapeutic target in prostate cancer, especially in CRPCs.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Cell Lines and Cell Culture
 - Clinical Specimen
 - Animal Studies
- METHOD DETAILS
 - Antibodies
 - Lentiviral Knockdown, siRNA Knockdown and CRISPR/Cas9-Mediated Knockout
 - CRISPR Knockout Validation by Sequencing
 - Stable Gene Expression
 - RNA Isolation and qRT-PCR
 - Immunoblot
 - Analysis of Public Human Gene Expression Datasets
 - Gene Expression Analysis
 - Chromatin Immunoprecipitation and Sequencing
 - ChIP-re-ChIP
 - Assay for Transposase-Accessible Chromatin Using Sequencing (ATAC-seq) and Analysis
 - Colony Formation Assay
 - Cell Proliferation Assay
 - Mouse Xenograft Procedures
 - Analysis of HNF4G IHC in Prostate Cancer Tissue Microarrays
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY
 - Gene Expression Omnibus (GEO) Accession Numbers of Datasets Generated

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and seven tables and can be found with this article online at <https://doi.org/10.1016/j.ccell.2017.10.008>.

AUTHOR CONTRIBUTIONS

Conceptualization, S.S., P.C., and Y.C.; Investigation, S.S. and J.C.; Methodology, S.S., L.R., P.A., Y.X., Z.C., D.A.M., E.G.W., S.W., J.S., Y.C., J.W., Y.Z., J.C., J.M.M., D.L., and E.W.P.W.; Formal Analysis, A.S., H.B., R.P.K., A.G., D.Z., and Y.C.; Resources, J.M.M., M.A.R., and H.I.S.; Writing – Original Draft, S.S., P.C., and Y.C.; Writing – Review & Editing, all authors.

ACKNOWLEDGMENTS

We would like to thank Dr. Ling Cai and Daniel Neil Weinberg for reagent development, and Dr. Takao Hamakubo for HNF4G antibody. Next-generation sequencing and gene expression arrays were done at the Memorial Sloan

Kettering Cancer Center (MSKCC) Genomics Core Facility (for ChIP-seq and RNA-seq) and New York Genome Center (for ATAC-seq). ATAC-seq was performed by the Epigenomics Core at MSKCC. Immunohistochemistry was performed at the Translational Research Program facility, Department of Pathology and Laboratory Medicine at Weill Cornell Medicine. We would like to thank support and feedback from the Prostate Cancer Foundation. This work was supported by grants from the NCI (K08CA140946, Y.C.; R01 CA208100, Y.C.; R01CA193837, Y.C.; P50CA092629, Y.C., H.I.S., and A.G.; P30CA008748, P.C., Y.C., and H.I.S.; P50CA140146, P.C.; K08CA151660, P.C.; DP2 CA174499, P.C.; U01 CA111275-09, M.A.R. and J.M.M.); US DOD (W81XWH-10-1-0197) and Starr Cancer Consortium, Y.C., P.C., G.S., and M.A.R.; Geoffrey Beene Cancer Research Center, Y.C. and P.C.; Gerstner Family Foundation, Y.C.; Bressler Scholars Fund, Y.C.; GIST Cancer Research Fund, P.C.; Shuman Fund, P.C.; and GIST Cancer Awareness Fund, P.C. J.C. was supported by the Nuovo-Soldati foundation.

Received: December 16, 2016

Revised: July 13, 2017

Accepted: October 17, 2017

Published: November 16, 2017

REFERENCES

- Antonarakis, E.S., Lu, C., Wang, H., Luber, B., Nakazawa, M., Roeser, J.C., Chen, Y., Mohammad, T.A., Chen, Y., Fedor, H.L., et al. (2014). AR-V7 and resistance to enzalutamide and abiraterone in prostate cancer. *N. Engl. J. Med.* **371**, 1028–1038.
- Arora, V.K., Schenkein, E., Murali, R., Subudhi, S.K., Wongvipat, J., Balbas, M.D., Shah, N., Cai, L., Efsthathiou, E., Logothetis, C., et al. (2013). Glucocorticoid receptor confers resistance to antiandrogens by bypassing androgen receptor blockade. *Cell* **155**, 1309–1322.
- Balbas, M.D., Evans, M.J., Hosfield, D.J., Wongvipat, J., Arora, V.K., Watson, P.A., Chen, Y., Greene, G.L., Shen, Y., and Sawyers, C.L. (2013). Overcoming mutation-based resistance to antiandrogens with rational drug design. *Elife* **2**, e00499.
- Baraille, F., Ayari, S., Carriere, V., Osinski, C., Garbin, K., Blondeau, B., Guillemain, G., Serradas, P., Rousset, M., Lacasa, M., et al. (2015). Glucose tolerance is improved in mice invalidated for the nuclear receptor HNF-4gamma: a critical role for enteroendocrine cell lineage. *Diabetes* **64**, 2744–2756.
- Bass, A.J., Watanabe, H., Mermel, C.H., Yu, S., Perner, S., Verhaak, R.G., Kim, S.Y., Wardwell, L., Tamayo, P., Gat-Viks, I., et al. (2009). SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas. *Nat. Genet.* **41**, 1238–1242.
- Beltran, H., Prandi, D., Mosquera, J.M., Benelli, M., Puca, L., Cyrta, J., Marotz, C., Giannopoulou, E., Chakravarthi, B.V., Varambally, S., et al. (2016). Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. *Nat. Med.* **22**, 298–305.
- Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218.
- Cancer Genome Atlas Research Network. (2015). The molecular taxonomy of primary prostate cancer. *Cell* **163**, 1011–1025.
- Chen, C.D., Welsbie, D.S., Tran, C., Baek, S.H., Chen, R., Vessella, R., Rosenfeld, M.G., and Sawyers, C.L. (2004). Molecular determinants of resistance to antiandrogen therapy. *Nat. Med.* **10**, 33–39.
- Chen, Y., Chi, P., Rockowitz, S., Iaquinata, P.J., Shamu, T., Shukla, S., Gao, D., Sirota, I., Carver, B.S., Wongvipat, J., et al. (2013). ETS factors reprogram the androgen receptor cistrome and prime prostate tumorigenesis in response to PTEN loss. *Nat. Med.* **19**, 1023–1029.
- Chi, P., Chen, Y., Zhang, L., Guo, X., Wongvipat, J., Shamu, T., Fletcher, J.A., Dewell, S., Maki, R.G., Zheng, D., et al. (2010). ETV1 is a lineage survival factor that cooperates with KIT in gastrointestinal stromal tumours. *Nature* **467**, 849–853.
- Daigo, K., Kawamura, T., Ohta, Y., Ohashi, R., Katayose, S., Tanaka, T., Aburatani, H., Naito, M., Kodama, T., Ihara, S., et al. (2011). Proteomic analysis of native hepatocyte nuclear factor-4alpha (HNF4alpha) isoforms, phosphorylation status, and interactive cofactors. *J. Biol. Chem.* **286**, 674–686.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21.
- Flavin, R., Pettersson, A., Hendrickson, W.K., Fiorentino, M., Finn, S., Kunz, L., Judson, G.L., Lis, R., Bailey, D., Fiore, C., et al. (2014). SPINK1 protein expression and prostate cancer progression. *Clin. Cancer Res.* **20**, 4904–4911.
- Gao, D., Vela, I., Sboner, A., Iaquinata, P.J., Karthaus, W.R., Gopalan, A., Dowling, C., Wanjala, J.N., Undvall, E.A., Arora, V.K., et al. (2014). Organoid cultures derived from patients with advanced prostate cancer. *Cell* **159**, 176–187.
- Garraway, L.A., Widlund, H.R., Rubin, M.A., Getz, G., Berger, A.J., Ramaswamy, S., Beroukhi, R., Milner, D.A., Granter, S.R., Du, J., et al. (2005). Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. *Nature* **436**, 117–122.
- Grasso, C.S., Wu, Y.M., Robinson, D.R., Cao, X., Dhanasekaran, S.M., Khan, A.P., Quist, M.J., Jing, X., Lonigro, R.J., Brenner, J.C., et al. (2012). The mutational landscape of lethal castration-resistant prostate cancer. *Nature* **487**, 239–243.
- GTEX Consortium. (2015). Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660.
- He, H.H., Meyer, C.A., Shin, H., Bailey, S.T., Wei, G., Wang, Q., Zhang, Y., Xu, K., Ni, M., Lupien, M., et al. (2010). Nucleosome dynamics define transcriptional enhancers. *Nat. Genet.* **42**, 343–347.
- Heckl, D., Kowalczyk, M.S., Yudovich, D., Belizaire, R., Puram, R.V., McConkey, M.E., Thielke, A., Aster, J.C., Regev, A., and Ebert, B.L. (2014). Generation of mouse models of myeloid malignancy with combinatorial genetic lesions using CRISPR-Cas9 genome editing. *Nat. Biotechnol.* **32**, 941–946.
- Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., et al. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108–112.
- Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589.
- Huang, P., He, Z., Ji, S., Sun, H., Xiang, D., Liu, C., Hu, Y., Wang, X., and Hui, L. (2011). Induction of functional hepatocyte-like cells from mouse fibroblasts by defined factors. *Nature* **475**, 386–389.
- Huggins, C., and Hodges, C.V. (1941). Studies on prostatic cancer. I. The effect of castration, of estrogen and of androgen injection on serum phosphatases in metastatic carcinoma of the prostate. *Cancer Res.* **1**, 293–297.
- Iwafuchi-Doi, M., and Zaret, K.S. (2014). Pioneer transcription factors in cell reprogramming. *Genes Dev.* **28**, 2679–2692.
- Iwafuchi-Doi, M., Donahue, G., Kakumanu, A., Watts, J.A., Mahony, S., Pugh, B.F., Lee, D., Kaestner, K.H., and Zaret, K.S. (2016). The pioneer transcription factor FoxA maintains an accessible nucleosome configuration at enhancers for tissue-specific gene activation. *Mol. Cell* **62**, 79–91.
- Jin, H.J., Zhao, J.C., Wu, L., Kim, J., and Yu, J. (2014). Cooperativity and equilibrium with FOXA1 define the androgen receptor transcriptional program. *Nat. Commun.* **5**, 3972.
- Johnson, M.H., Ross, A.E., Alshalalfa, M., Erho, N., Yousefi, K., Glavaris, S., Fedor, H., Han, M., Faraj, S.F., Bezerra, S.M., et al. (2016). SPINK1 defines a molecular subtype of prostate cancer in men with more rapid progression in an at-risk, natural history radical prostatectomy cohort. *J. Urol.* **196**, 1436–1444.
- Jozwik, K.M., and Carroll, J.S. (2012). Pioneer factors in hormone-dependent cancers. *Nat. Rev. Cancer* **12**, 381–385.
- Karthaus, W.R., Iaquinata, P.J., Drost, J., Gracanin, A., van Boxtel, R., Wongvipat, J., Dowling, C.M., Gao, D., Begthel, H., Sachs, N., et al. (2014).

- Identification of multipotent luminal progenitor cells in human prostate organoid cultures. *Cell* 159, 163–175.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25.
- Leinonen, K.A., Tolonen, T.T., Bracken, H., Stenman, U.H., Tammela, T.L., Saramaki, O.R., and Visakorpi, T. (2010). Association of SPINK1 expression and TMPRSS2:ERG fusion with prognosis in endocrine-treated prostate cancer. *Clin. Cancer Res.* 16, 2845–2851.
- Lupien, M., Eeckhoute, J., Meyer, C.A., Wang, Q., Zhang, Y., Li, W., Carroll, J.S., Liu, X.S., and Brown, M. (2008). FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell* 132, 958–970.
- Machanick, P., and Bailey, T.L. (2011). MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* 27, 1696–1697.
- Miettinen, M., Wang, Z.F., Paetau, A., Tan, S.H., Dobi, A., Srivastava, S., and Sesterhenn, I. (2011). ERG transcription factor as an immunohistochemical marker for vascular endothelial tumors and prostatic carcinoma. *Am. J. Surg. Pathol.* 35, 432–441.
- Nicol, J.W., Helt, G.A., Blanchard, S.G., Jr., Raja, A., and Loraine, A.E. (2009). The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. *Bioinformatics* 25, 2730–2731.
- Odom, D.T., Zizlsperger, N., Gordon, D.B., Bell, G.W., Rinaldi, N.J., Murray, H.L., Volkert, T.L., Schreiber, J., Rolfe, P.A., Gifford, D.K., et al. (2004). Control of pancreas and liver gene expression by HNF transcription factors. *Science* 303, 1378–1381.
- Odom, D.T., Dowell, R.D., Jacobsen, E.S., Neklodova, L., Rolfe, P.A., Danford, T.W., Gifford, D.K., Fraenkel, E., Bell, G.I., and Young, R.A. (2006). Core transcriptional regulatory circuitry in human hepatocytes. *Mol. Syst. Biol.* 2, 2006.0017.
- Pan, X., Zhang, X., Gong, J., Tan, J., Yin, X., Tang, Q., Shu, K., Shen, P., Zeng, H., and Chen, N. (2016). The expression profile and prognostic value of SPINK1 in initially diagnosed bone metastatic prostate cancer. *Prostate* 76, 823–833.
- Parviz, F., Matullo, C., Garrison, W.D., Savatski, L., Adamson, J.W., Ning, G., Kaestner, K.H., Rossi, J.M., Zaret, K.S., and Duncan, S.A. (2003). Hepatocyte nuclear factor 4alpha controls the development of a hepatic epithelium and liver morphogenesis. *Nat. Genet.* 34, 292–296.
- Pinello, L., Canver, M.C., Hoban, M.D., Orkin, S.H., Kohn, D.B., Bauer, D.E., and Yuan, G.C. (2016). Analyzing CRISPR genome-editing experiments with CRISPResso. *Nat. Biotechnol.* 34, 695–697.
- Pomerantz, M.M., Li, F., Takeda, D.Y., Lenci, R., Chonkar, A., Chabot, M., Cejas, P., Vazquez, F., Cook, J., Shivdasani, R.A., et al. (2015). The androgen receptor cistrome is extensively reprogrammed in human prostate tumorigenesis. *Nat. Genet.* 47, 1346–1351.
- Ramirez, F., Ryan, D.P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dundar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* 44, W160–W165.
- Ran, L., Sirota, I., Cao, Z., Murphy, D., Chen, Y., Shukla, S., Xie, Y., Kaufmann, M.C., Gao, D., Zhu, S., et al. (2015). Combined inhibition of MAP kinase and KIT signaling synergistically destabilizes ETV1 and suppresses GIST tumor growth. *Cancer Discov.* 5, 304–315.
- Robinson, D., Van Allen, E.M., Wu, Y.M., Schultz, N., Lonigro, R.J., Mosquera, J.M., Montgomery, B., Taplin, M.E., Pritchard, C.C., Attard, G., et al. (2015). Integrative clinical genomics of advanced prostate cancer. *Cell* 161, 1215–1228.
- Sanjana, N.E., Shalem, O., and Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* 11, 783–784.
- Sarbasov, D.D., Guertin, D.A., Ali, S.M., and Sabatini, D.M. (2005). Phosphorylation and regulation of Akt/PKB by the rictor-mTOR complex. *Science* 307, 1098–1101.
- Sekiya, S., and Suzuki, A. (2011). Direct conversion of mouse fibroblasts to hepatocyte-like cells by defined factors. *Nature* 475, 390–393.
- Stanbrough, M., Bubley, G.J., Ross, K., Golub, T.R., Rubin, M.A., Penning, T.M., Febbo, P.G., and Balk, S.P. (2006). Increased expression of genes converting adrenal androgens to testosterone in androgen-independent prostate cancer. *Cancer Res.* 66, 2815–2825.
- Stergachis, A.B., Neph, S., Reynolds, A., Humbert, R., Miller, B., Paige, S.L., Vernot, B., Cheng, J.B., Thurman, R.E., Sandstrom, R., et al. (2013). Developmental fate and cellular maturity encoded in human regulatory DNA landscapes. *Cell* 154, 888–903.
- Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* 102, 15545–15550.
- Taplin, M.E., Bubley, G.J., Shuster, T.D., Frantz, M.E., Spooner, A.E., Ogata, G.K., Keer, H.N., and Balk, S.P. (1995). Mutation of the androgen-receptor gene in metastatic androgen-independent prostate cancer. *N. Engl. J. Med.* 332, 1393–1398.
- Taylor, B.S., Schultz, N., Hieronymus, H., Gopalan, A., Xiao, Y., Carver, B.S., Arora, V.K., Kaushik, P., Cerami, E., Reva, B., et al. (2010). Integrative genomic profiling of human prostate cancer. *Cancer Cell* 18, 11–22.
- Tomlins, S.A., Laxman, B., Dhanasekaran, S.M., Helgeson, B.E., Cao, X., Morris, D.S., Menon, A., Jing, X., Cao, Q., Han, B., et al. (2007). Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. *Nature* 448, 595–599.
- Tomlins, S.A., Rhodes, D.R., Yu, J., Varambally, S., Mehra, R., Perner, S., Demichelis, F., Helgeson, B.E., Laxman, B., Morris, D.S., et al. (2008). The role of SPINK1 in ETS rearrangement-negative prostate cancers. *Cancer Cell* 13, 519–528.
- Tran, C., Ouk, S., Clegg, N.J., Chen, Y., Watson, P.A., Arora, V., Wongvipat, J., Smith-Jones, P.M., Yoo, D., Kwon, A., et al. (2009). Development of a second-generation antiandrogen for treatment of advanced prostate cancer. *Science* 324, 787–790.
- Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515.
- Wallerman, O., Motallebipour, M., Enroth, S., Patra, K., Bysani, M.S., Komorowski, J., and Wadelius, C. (2009). Molecular interactions between HNF4a, FOXA2 and GABP identified at regulatory DNA elements through ChIP-sequencing. *Nucleic Acids Res.* 37, 7498–7508.
- Wang, D., Garcia-Bassets, I., Benner, C., Li, W., Su, X., Zhou, Y., Qiu, J., Liu, W., Kaikkonen, M.U., Ohgi, K.A., et al. (2011). Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* 474, 390–394.
- Wang, A., Yue, F., Li, Y., Xie, R., Harper, T., Patel, N.A., Muth, K., Palmer, J., Qiu, Y., Wang, J., et al. (2015). Epigenetic priming of enhancers predicts developmental competence of hESC-derived endodermal lineage intermediates. *Cell Stem Cell* 16, 386–399.
- Watson, P.A., Arora, V.K., and Sawyers, C.L. (2015). Emerging mechanisms of resistance to androgen receptor inhibitors in prostate cancer. *Nat. Rev. Cancer* 15, 701–711.
- Wiederschain, D., Wee, S., Chen, L., Loo, A., Yang, G., Huang, A., Chen, Y., Caponigro, G., Yao, Y.M., Lengauer, C., et al. (2009). Single-vector inducible lentiviral RNAi system for oncology target validation. *Cell Cycle* 8, 498–504.
- Wisely, G.B., Miller, A.B., Davis, R.G., Thornquest, A.D., Jr., Johnson, R., Spitzer, T., Sefler, A., Shearer, B., Moore, J.T., Willson, T.M., et al. (2002). Hepatocyte nuclear factor 4 is a transcription factor that constitutively binds fatty acids. *Structure* 10, 1225–1234.
- Yuan, X., Ta, T.C., Lin, M., Evans, J.R., Dong, Y., Bolotin, E., Sherman, M.A., Forman, B.M., and Sladek, F.M. (2009). Identification of an endogenous ligand bound to a native orphan nuclear receptor. *PLoS One* 4, e5609.
- Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors: establishing competence for gene expression. *Genes Dev.* 25, 2227–2241.
- Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-seq (MACS). *Genome Biol.* 9, R137.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit Anti-HNF4G antibody, HNF4G Ab1	Sigma	Cat#HPA005438 RRID:AB_1850822
Rabbit Anti-HNF4G antibody, HNF4G Ab2	Proteintech	Cat#25801-1-AP
Mouse Anti-HNF4G antibody [Clone B6502A] HNF4G Ab3	Abcam, Dr. Takao Hamakubo, University of Tokyo (Daigo et al., 2011)	Cat# ab66179, RRID:AB_1141087
Goat Anti-HNF1A antibody	Santa Cruz Biotechnology	Cat#Sc6547X RRID: AB_648295
Rabbit Anti-H3K4me1 antibody	Abcam	Cat# ab8895 RRID:AB_306847
Rabbit Anti- H3K27ac	Abcam	Cat# ab4729 RRID:AB_2118291
Rabbit Anti-AR antibody	Abcam	Cat# ab108341 RRID:AB_10865716
Mouse Anti- SPINK1 antibody SPINK1 Ab1	R & D Systems	Cat#MAB7496
Rabbit Anti- SPINK1 antibody SPINK1 Ab2	Abcam	Cat#Ab207302
Goat Anti-FOXA1 antibody	Abcam	Cat# ab5089 RRID:AB_304744
Rabbit Anti-GAS2 antibody	Abcam	Cat# ab109762 RRID:AB_10864608
Mouse Anti-AKR1C3 antibody	Sigma	Cat# A6229 RRID:AB_476751
Rabbit Anti-Albumin antibody	Cell signaling	Cat# 4929S; RRID: AB_2225785
Mouse Anti-GAPDH	Abmgood	Cat# G041
Rabbit Anti-UGT2B15	Abcam	ab154864
Rabbit Anti-TMED	Abcam	ab182489
Bacterial and Virus Strains		
MAX Efficiency® Stbl2™ Competent Cells	Thermo Fisher Scientific	Cat#10268019
One Shot™ Stbl3™ Chemically Competent E. coli	Thermo Fisher Scientific	Cat#C737303
Biological Samples		
Prostate TMA	WMC	N/A
Chemicals, Peptides, and Recombinant Proteins		
X-tremegene9	Roche	Cat#06365809001
HNF1A siRNA	Dharmacon	Cat#L-008215-00-0005
SCR siRNA	Dharmacon	Cat#D-001810-10-05
Lipofectamine RNAiMAX	Invitrogen	Cat#13778100
Matrigel	Corning	CB40234C
Enzalutamide	Dr Charles Sawyers (MSKCC, New York)	N/A
Protein A/G Magnetic beads	Thermofisher	Cat#26162
Critical Commercial Assays		
CellTiter-Glo® 2.0 Assay	Promega	Cat# G9242
Re-ChIP-IT®	Active motif	Cat# 53016
ECL kit	Thermo Scientific	Cat# 80196
E.Z.N.A total RNA kit	Omega	Cat# R6834-02
Deposited Data		
GSE85242	This study	GEO database
GSE85244	This study	GEO database
GSE85556	This study	GEO database
GSE85557	This study	GEO database
GSE85558	This study	GEO database

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental Models: Cell Lines		
Human: 22Rv1	ATCC	
Human: LNCaP	ATCC	
Human: LNCaP/AR	Dr Charles Sawyers (MSKCC, New York)	
Human: MSKPCa-1	Gao et al., 2014	
Human: MSKPCa-10	This study	
Experimental Models: Organisms/Strains		
CB17-SCID mice	Taconic	
Oligonucleotides		
see Table S7	This study	N/A
Recombinant DNA		
pLKO-puro	Addgene	Addgene plasmid # 8453
Tet-pLKO-puro	Wiederschain et al., 2009	Addgene plasmid # 21915
Lenti-Cas9-blast	Feng Zhang (Sanjana et al., 2014)	Addgene plasmid # 52962
pLKO5.sgRNA.EFS.GFP	Benjamin Ebert (Heckl et al., 2014)	Addgene plasmid # 57822
pDONR201(HNF4G)	Harvard medical school Plasmid database	ID:HsCD00022314
HNF1A overexpression plasmid	Origene	Cat # RC211201L1
pMSCV-Puro	Addgene	Addgene plasmid #68469
pMSCVhygro	Clontech	Cat # 634401
Software and Algorithms		
GENE-E software	Broad Institute	http://www.broadinstitute.org/cancer/software/GENE-E
JAVA GSEA 2.0 program	(Subramanian et al., 2005)	http://software.broadinstitute.org/gsea/downloads.jsp
Bowtie alignment software	Langmead et al., 2009	http://bowtie-bio.sourceforge.net/index.shtml
ImageJ	NIH	https://fiji.sc
PRISM	Graphpad	https://www.graphpad.com
MACS 2.1 callpeak	Zhang et al., 2008	
Homer mergePeaks algorithm	Heinz et al., 2010	
Integrated Genome Browser (IGB)		
DeepTool2	Ramirez et al., 2016	
MEME-ChIP suite	Machanick and Bailey, 2011	
Homer annotatePeaks program		
STAR v2.3	(Dobin et al., 2013)	
Cufflinks	Trapnell et al., 2010	
CRISPR design tool	Feng Zhang	http://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design
CRISPResso		http://crispresso.rocks/
Flow Jo	FLOWJO LLC	FlowJo

CONTACT FOR REAGENT AND RESOURCE SHARING

Please direct all requests for reagents and resource sharing to the Lead Contact, Yu Chen (cheny1@mskcc.org).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell Lines and Cell Culture

The LNCaP and 22Rv1 cell lines were obtained from the American Type Culture Collection (ATCC). The LNCaP/AR line was a kind gift from Dr. Charles Sawyers (Memorial Sloan-Kettering Cancer Center). All these cell lines were maintained in RPMI supplemented with 10% Fetal bovine serum (Omega), L-glutamine (2 mM), penicillin (100 U/ml), and streptomycin (100 μ g/ml). MSK-PCa1 and MSK-PCa10 were generated from patient samples by organoid culture technique and cultured as described previously (Gao et al., 2014). Cell lines were authenticated by RNA-seq analysis of SNP's compared to exome data from Cancer Cell Line Encyclopedia. Cell lines were confirmed mycoplasma free by PCR testing.

Clinical Specimen

For Prostate tissue microarray staining, archival Formalin-fixed paraffin-embedded (FFPE) material was used under an IRB-approved protocol (Weill Cornell Medicine IRB 1007011157A015). For benign prostate tissue, TMA cores were obtained for areas localized at a distance from prostate cancer. For hormone naïve, organ-confined prostate adenocarcinomas (23 patients in total). Tumor tissue was obtained from radical prostatectomy specimens. The Gleason scores ranged from 6 (3+3) to 9 (4+5). For castration-resistant prostate cancer (33 patients in total), castration resistance was defined as disease progression in spite of an androgen-deprivation treatment. Seven samples were obtained from metastatic sites, the remaining ones represented locally advanced tumor. Each case was represented at least in duplicate (two cores per case) on the TMA. Most cases were represented in triplicate (three cores per case).

Animal Studies

All mice procedures were performed under MSKCC approved IACUC protocol #11-12-027. 6-8 weeks old CB17-SCID mice were purchased from Taconic and maintained under standard pathogen free conditions. For Enzalutamide treatment, we calculated that 7 xenografts are required to give 90% power to detect a 50% difference in growth with $\alpha=0.05$. We used an intragroup variance of 50% and a treatment effect of 2-fold.

METHOD DETAILS

Antibodies

Antibodies to the following were used for Western blotting, ChIP, IHC and IF: rabbit anti-HNF4G (Sigma; HPA005438; 1:50 for IHC, 10 μ g for ChIP-seq), rabbit anti-HNF4G (Proteintech; 25801-1-AP; 10 μ g for ChIP-Seq) mouse anti-HNF4G (Clone B6502A; a kind gift from Dr. Takao Hamakubo, University of Tokyo (Daigo et al., 2011); 1:2000 for Western blotting), goat anti-HNF1A (Santa Cruz Biotechnology, sc6547X; 1:2000 for Western blotting), rabbit anti-H3K4me1 (for ChIP; Abcam; ab8895), rabbit anti-H3K27ac (for ChIP; Abcam; ab4729), rabbit anti-AR (Abcam, ab108341; 1:1000 for Western blotting, 5 μ g for ChIP-seq), goat anti-FOXA1 (for ChIP; Abcam; ab5089), mouse anti-SPINK1 (R&D Systems; MAB7496; 2 μ g/ml for Western Blotting), rabbit anti-GAS2 (Abcam; ab109762; 1:200 for Western blotting), mouse anti-AKR1C3 (Sigma; A6229; 1:1000 for Western blotting), rabbit anti-Albumin (Cell signaling; 4929S; 1:1000 for Western blotting), mouse anti-GAPDH (abmgood; G041; 1:5000 for Western blotting), rabbit anti-UGT2B15 (Abcam; ab154864; 1:1000 for Western blotting), rabbit anti-TMED6 (Abcam; ab182489; 1:500 for Western blotting).

Lentiviral Knockdown, siRNA Knockdown and CRISPR/Cas9-Mediated Knockout

The following hairpins were used to knockdown HNF4G:

HNF4Gsh1: TRCN0000019243, HNF4Gsh2: TRCN0000420190 and shSCR (Addgene plasmid # 1864) (Sarbasov et al., 2005). See Table S7 for hairpins sequences. These were sub-cloned into Tet-pLKO-puro, a gift from Dmitri Wiederschain (Addgene plasmid # 21915) (Wiederschain et al., 2009). Lentiviral particles were generated by co-transfecting the lentiviral constructs with psPax2 and pVSV-G into 293FT cells using XtremeGene 9 (Roche) or Fugene 6 (Promega; E269A). Viral supernatants were collected 48-72 hours post transfection. Stable cell lines were generated after puromycin selection. Doxycycline at a concentration of 100 ng/ml was used to achieve HNF4G knockdown.

The following hairpins were used to knockdown HNF1A:

HNF1Ash1: TRCN0000017193, HNF1Ash2: TRCN0000017196, HNF1Ash3: TRCN0000017195 and shSCR (as above). See Table S7 for hairpins sequences.

Cell lysates and RNA were isolated 72 hours after transduction for immunoblot and qRT-PCR. For siRNA knockdown, 22Rv1 cells were transfected with pooled siRNA (20 nM) against HNF1A (Dharmacon #L-008215-00-0005) or scrambled control (Dharmacon #D-001810-10-05) using Lipofectamine RNAiMAX transfection reagent (Invitrogen #13778100). RNA was harvested 72 hours after transfection.

For CRISPR/Cas9 knock-out of HNF4G and HNF1A in 22Rv1 cell line, single guide RNA (sgRNA) sequences were designed using CRISPR design tool from Feng Zhang lab (<http://portals.broadinstitute.org/gpp/public/analysis-tools/sgRNA-design>). See Table S7 for sgRNA sequences. We first generated 22Rv1 cells stably expressing Cas9 using Lenti-Cas9-blast, a gift from Feng Zhang (Addgene plasmid # 52962) (Sanjana et al., 2014). We next transduced the 22Rv1-Cas9 cells with sgRNAs and GFP co-expression. The sgRNAs were cloned into the pLKO5.sgRNA.EFS.GFP a gift from Benjamin Ebert (Addgene plasmid # 57822) (Heckl et al., 2014).

CRISPR Knockout Validation by Sequencing

For HNF4G and HNF1A knockout validation, 22Rv1-Cas9 cells were transduced with either sgNTC, sgHNF4G or sgHNF1A expressing viruses at a high MOI. After two days of puromycin selection and at seven days of infection genomic DNA was isolated from each population. Region specific primers were designed to amplify sgHNF4G and sgHNF1A target regions using genomic DNA extracted from each of the three cell population. The amplicons were subjected to 125 bp paired-end deep sequencing. The reads obtained were aligned to Refseq sequence using software CRISPResso (Pinello et al., 2016) and were analyzed for indels at or around the sgRNA target regions.

Stable Gene Expression

cDNA for HNF4G in pDONR201 vector was obtained from Harvard medical school Plasmid database (ID:HsCD00022314) and cloned into an murine stem cell virus (MSCV)-based retroviral vector with puromycin selection marker (Addgene) using Gateway technology. The HNF1A overexpression plasmid was purchased from Origene (RC211201L1). The ORF was subsequently sub-cloned into pMSCVhygro (Addgene). Retroviruses were produced in 293FT cells by standard methods using amphoteric packaging vector. Virus-containing supernatant was harvested at 48 and 72 h after transfection, pooled and filtered through a 0.45 μm PVDF membrane, and used for transduction in the presence of polybrene ($8 \mu\text{g ml}^{-1}$). LNCaP cells were selected with $2 \mu\text{g/ml}$ puromycin or $400 \mu\text{g/ml}$ of hygromycin for 4 days at 48 hours after infection. The lines generated were used for subsequent studies.

RNA Isolation and qRT-PCR

To isolate RNA from cell lines, E.Z.N.A total RNA kit (Omega) was used. To isolate RNA from xenograft tumor explants, the tumor samples were ground in 1 ml Trizol (Invitrogen) using a PowerGen homogenizer (Fisher Scientific), followed by the addition of 200 μL chloroform. The samples were then centrifuged at 10,000 g for 15 minutes. The upper phase was mixed with an equal volume of 70% ethanol, and the RNA was further purified using the E.Z.N.A total RNA kit (Omega). For qRT-PCR, RNA was reverse transcribed using the High-Capacity CDNA Reverse Transcription Lit (ABI). Power SYBR Master Mix (ABI) was used to run PCR on a ViiA7 Real Time PCR System (Life Technologies).

Immunoblot

Cell lysates were prepared in RIPA buffer supplemented with proteinase/phosphatase inhibitor. Proteins were resolved on NuPAGE Novex 4–12% Bis-Tris Protein Gels (Life Technologies) or Tris -glycine 4-20% acrylamide gels (Invitrogen) and transferred electrophoretically onto a PVDF 0.45 μm membrane (BioRad). Membranes were blocked for 1 hour at room temperature in blocking buffer consisting of 5% milk or 1 % BSA diluted in Tris buffer saline plus 0.1% Tween 20 (TBST) and were incubated overnight at 4 °C with the primary antibodies diluted in the same buffer. After 3 washes of 10 min in TBST, membranes were incubated with secondary antibodies diluted in blocking buffer for 1 hour at room temperature. After 3 washes of 10 minutes in TBST, Enhanced Chemiluminescence (ECL) was performed using ECL kit (Thermo Scientific, 80196).

Analysis of Public Human Gene Expression Datasets

To identify *SPINK1* correlated genes, we obtained normalized Affymetrix Human Exon 1.0 ST array gene expression data from MSKCC (GEO GSE21034), normalized RNA-seq data from prostate cancer TCGA (www.firebrowse.org) and RNA-seq prostate cancer data from Weill Cornell (dbGap phs000909.v.p1). We identified top 1,000 *SPINK1* correlated genes by Pearson analysis and overlapped them. We defined the PCa-GI signature as the 129 genes in two of three datasets and the core PCa-GI signature as the 40 genes in all three datasets. To determine the expression of PCa-GI signature genes in normal tissue, we downloaded the GTEx v1.18 RNA-seq gene expression data in RPKM (<http://www.gtexportal.org/home/datasets>). We used the mean expression of each tissue type. Z-score was calculated as the standard deviations away from mean. To determine the expression of *SPINK1*, *HNF1A*, *HNF4G* in normal and cancerous tissue in TCGA datasets or prostate, colon, liver, pancreas, rectum and stomach cancers, we used Tukey box-and-whisker plots from <http://firebrowse.org/>.

Gene Expression Analysis

For gene expression profiling of 22Rv1 cells after *HNF4G* knockdown, doxycycline inducible 22Rv1-HNF4Gsh1-dox, 22Rv1-HNF4Gsh2-Dox, and 22Rv1-shSCR-dox lines were treated with vehicle or doxycycline (100 ng/ml) for 72 hours in triplicate prior to RNA isolation. Microarray was performed on an Illumina HumanHT-12 platform.

For gene expression profiling of 22Rv1 cells after *HNF1A* knockdown, 22Rv1 cells were transfected with pooled siSCR and siHNF1A. At 72 hours after transfection, RNA was isolated and gene expression profiling was performed using RNA-seq.

For gene expression profiling of 22Rv1 cells after DHT treatment, 22Rv1 cells treated with vehicle (DMSO) or 1 nM DHT for 8 hours. RNA was isolated and gene expression profiling was performed using RNA-seq. For gene expression profiling of LNCaP cells after R1881 treatment, LNCaP cells treated with vehicle (DMSO) or 1 nM R1881 for 24 hours in triplicate. RNA was isolated and gene expression profiling was performed using Illumina HumanHT-12.

For gene expression profiling of LNCaP cells expressing exogenous *HNF4G*, LNCaP cells were transduced with viruses containing MSCV-HNF4G or empty vector control in triplicate. After 48 hours, cells were selected with puromycin ($2 \mu\text{g/ml}$) for 4 days. RNA was isolated 2 days later for gene expression analysis using Illumina HumanHT-12 platform.

For gene expression profiling of LNCaP cells expressing exogenous *HNF1A*, LNCaP cells were transduced with viruses containing MSCV-HNF1A or empty vector control in duplicate. After 48 hours, cells were selected with hygromycin (400 $\mu\text{g}/\text{ml}$) for 4 days. RNA was isolated 2 days later for gene expression analysis using RNA-seq.

To generate the transcriptome of LNCaP/AR-HNF4G and LNCaP/AR-Vec cells after 9 days of growth in charcoal-stripped serum and of LNCaP/AR-HNF4G cells after 32 days of growth in charcoal-stripped serum, we performed the entire 32-day experiment in duplicate. Cell growth analysis was performed using cell counting and gene expression was analyzed using RNA-seq.

RNA-seq was performed by MSKCC genomics core facility using poly-A capture. The libraries were sequenced on an Illumina HiSeq-2500 platform with 50 bp single reads to obtain a minimum yield of 40 million reads per sample. The sequence data were processed and mapped to the human reference genome (hg19) using STAR v2.3 (Dobin et al., 2013). Gene expression was quantified to reads-per-kilobase mapped (RPKM) using the Cufflinks (Trapnell et al., 2010) and Log₂ transformed. GSEA was performed using JAVA GSEA 2.0 program (Subramanian et al., 2005), using difference of mean between replicates and gene permutation. The gene sets used were the Broad Molecular Signatures Database gene sets v5, c2 (curated gene sets), c5 (gene ontology gene sets), c6 (oncogenic signatures), c7 (immunologic signatures) as well as PCa-GI custom gene set generated by us.

Heatmaps were generated using GENE-E software

(<http://www.broadinstitute.org/cancer/software/GENE-E>).

Chromatin Immunoprecipitation and Sequencing

Chromatin isolation from cell lines and immunoprecipitation was performed following the protocol previously described (Chi et al., 2010). For HNF4G knockdown experiments, chromatin was isolated from 22Rv1-HNF4Gsh2-Dox cell lines expressing doxycycline inducible HNF4Gsh2 hairpin 72 hours post doxycycline or mock treatment. For HNF4G overexpression, chromatin was isolated from LNCaP cells expressing HNF4G or a vector control. HNF4G, AR, FOXA1, H3K4me1 and H3K27ac ChIP were performed using the antibodies described in reagents section. Input DNA was also sequenced. For AR ChIP, cells were treated with 10 nM DHT for 8 hours to induce maximal AR binding sites. Next generation sequencing was performed on an Illumina HiSeq2000 platform with 50-bp or 100 bp single reads.

Reads were aligned to the human genome (hg 19) using the Bowtie alignment software (Langmead et al., 2009). Duplicate reads were eliminated for subsequent analysis. For FOXA1 and AR in both 22Rv1 and LNCaP cells and for HNF4G in LNCaP cells, peak calling was performed using MACS 2.1 callpeak function comparing immunoprecipitated chromatin with input chromatin, using a standard parameters and a q-value cutoff of 10^{-2} (Zhang et al., 2008). For HNF4G in 22Rv1-HNF4Gsh2-Dox cells, we used MACS2.1 bdgdiff function using profiles of vehicle and doxycycline treated HNF4G ChIP and input DNA to identify differential peaks in vehicle treated cells compared to doxycycline treated cells. After running MACS2.1, we further filtered the peaks file for q-value $< 10^{-5}$ for downstream application of high-confidence peaks.

To determine overlap of AR, FOXA1, and HNF4G peaks we used Homer mergePeaks algorithm (Heinz et al., 2010) and consider two peaks overlapping if the summits are within 200 bp of each other. The ChIP-seq profiles presented were generated using Integrated Genome Browser (IGB) software of bigWig format files, generated using the bamCoverage tool from deepTool2 (Nicol et al., 2009; Ramirez et al., 2016). We identified top 1,000 AR and top 1,000 HNF4G peaks, based on lowest q-values and plotted the ChIP-seq profiles around the sites using deepTools2. For motif analysis, we employed the MEME-ChIP suite (Machanic and Bailey, 2011), 400 bp sequences (200 bp upstream and downstream of the peak summit) of the top 1,000 peaks. To annotate peaks as promoter, gene body, and intergenic, we used Homer annotatePeaks program.

For integrative analysis of gene expression and ChIP-seq in 22Rv1 cells, we determined the gene expression change of HNF4G knockdown (22Rv1-HNF4Gsh2-Dox treated with doxycycline or vehicle) and AR activation (22Rv1 cells treated with vehicle or 10 nM DHT for 8 hours) at the closest genes which mapped to top HNF4G and AR peaks, using Homer annotatePeaks program. Similarly, for LNCaP cells, we determined the gene expression change of HNF4G overexpression (LNCaP-HNF4G vs LNCaP-Vec) and AR activation (LNCaP cells treated with vehicle or 1nM R1881 for 24 hours) at the closest genes that mapped to top HNF4G and AR peaks.

ChIP-re-ChIP

The ChIP-re-ChIP assay was performed with Re-ChIP-IT kit according to manufacturer's instructions (Active Motif, 53016). Anti-HNF4G (Proteintech; 25801-1-AP) and anti-FOXA1 (Abcam; ab5089) were used for the first ChIP; eluates from FOXA1 ChIP were used to perform 2nd ChIP with same anti-HNF4G, anti-FOXA1 or no antibody control.

Assay for Transposase-Accessible Chromatin Using Sequencing (ATAC-seq) and Analysis

ATAC-seq was performed as previously described (Buenrostro et al., 2013) with exception that digitonin was used instead of NP-40 for nuclei isolation. For each sample, cell nuclei were prepared from 50,000 cells, and incubated with 2.5 μl of transposase (Illumina) in a 50 μl reaction for 30 min at 37 °C. After purification of transposase-fragmented DNA, the library was amplified by PCR and subjected to high-throughput sequencing on an Illumina HiSeq2500 platform.

ATAC-seq reads were quality and adapter trimmed using 'trim_galore' before aligning to human genome assembly hg19 with bowtie2 using the default parameters. Aligned reads with the same start position and orientation were collapsed to a single read before subsequent analysis. Density profiles were created by extending each read to the average library fragment size and then

computing density using the BEDTools suite, with subsequent normalization to a sequencing depth of ten million reads for each library. Subsequent data analysis and display is as described in the ChIP-seq analysis section.

Colony Formation Assay

LNCaP/AR cells overexpressing HNF4G or vector control were pre-grown in media containing charcoal-stripped serum for about 7 days. After 7 days, 2500 cells from both group were plated on a well of a six-well tissue culture dish and allowed to grow for an additional 10–12 days. The colonies obtained were then washed in PBS buffer, fixed in 70% ethanol for 15 minutes and then stained with a 0.5% crystal violet solution in 25% ethanol. The dishes were washed in a gentle stream of water through the side of the wells and were then air dried. The colonies obtained were then counted by an automated colony counter.

Cell Proliferation Assay

For HNF4G and HNF1A knockdown in 22Rv1 cells, cells were plated at 0.1×10^6 cells per well in a 12-well plate on day 0 and infected with pLKO.1 hairpin viruses at MOI ~ 2.0 on day 1. Cells were not further puromycin selected. Cells were counted in triplicate using Vicell XR cell viability analyzer at the indicated days.

For LNCaP/AR growth curve in charcoal-stripped serum with exogenous HNF4G or Vector expression, cells were plated at 0.1 million in a six well plate in duplicates. The cells were counted and replated at a density of 0.1 million at indicated days. The difference in split ratios to plate the same number of cells was noted and was used in growth curve plotting to extrapolate the total number of cells obtained every week.

For competition growth assay, 22Rv1 cells were stably transduced for Cas9 expression. 22Rv1-Cas9 cells were then transduced at ~ 0.4 MOI for dual expression of GFP and CRISPR guide RNAs against HNF4G, HNF1A or non target control (NTC). FACS analysis was performed at regular intervals to determine any changes in percentage of GFP-positive cells over the course of the experiment.

For HNF4G and HNF1A knockdown in patient derived organoid cell lines, MSK-PCa1 and MS-PCa10, about 2 million cells were plated per well of a six-well plate and were transduced next day with desired pLKO.1 hairpin viruses at MOI ~ 2.0 . After a day of infection, cells were split and plated at a density of 20,000 cells per well of a 96 well plate. At indicated time points, the number of cells was determined using a CelltiterGlo assay (Promega)

Mouse Xenograft Procedures

For LNCaP/AR xenograft studies, 2.0×10^6 cells resuspended in 100 μ L of 1:1 mix of growth media and Matrigel (BD Biosciences) were subcutaneously injected into 6–8 weeks old CB17-SCID castrated male mice (Taconic). Tumor sizes were measured weekly with callipers starting 10 weeks after xenografting and were calculated using the following formula: tumor volume = $(D^2 \times d \times h^2)/6$, whereby D, d and h refers to long diameter, short diameter and height of the tumor, respectively. Treatment with enzalutamide (10 mg/kg) or vehicle was begun at a tumor size of 400 mm³. Mice were treated once daily until the end of the experiments.

For 22Rv1 xenografts, 2.0×10^6 cells resuspended in 100 μ L of 1:1 mix of growth media and Matrigel (BD Biosciences) were subcutaneously injected into 6–8 weeks old CB17-SCID mice (Taconic). The mice were fed with doxycycline water (200 mg/L in 0.5% sucrose) from the beginning of the grafting or when the tumors reached a size of 100 mm³. Tumors were measured twice a week after 1 week of grafting.

Analysis of HNF4G IHC in Prostate Cancer Tissue Microarrays

Immunohistochemistry (IHC) was performed using the anti-HNF4G antibody HPA005438 (MilliporeSigma, Billerica, MA, USA) at a 1:50 dilution, on a Bond III automated immunostainer (Leica Microsystems, IL, USA). Formalin-fixed paraffin-embedded (FFPE) tissue sections were de-paraffinized and endogenous peroxidase was inactivated. Antigen retrieval was performed using the Bond Epitope Retrieval Solution 1 (ER1) at 99–100°C for 60 minutes (Leica Microsystems). Sections were then incubated sequentially with the primary antibody overnight, post-primary for 15 minutes and polymer for 25 minutes, followed by a 10 minute colorimetric development with diaminobenzidine (DAB) (Bond Polymer Refine Detection; Leica Microsystems). FFPE material from 22Rv1 prostate cancer cell line xenografts with known levels of HNF4G expression were used as controls.

IHC was performed on tissue microarrays composed of representative cases of benign human prostate tissue, hormone naïve prostatic carcinoma (HNPca), castration-resistant prostatic adenocarcinoma (CRPC). Staining was evaluated by a pathologist with experience in genitourinary pathology (JC). Nuclear HNF4G expression was considered positive if at least 10% of nuclei in a given tissue type showed brown staining. Nuclear staining intensity was scored following a three-tiered system (negative=0, weak=1, medium=2 and strong=3).

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical comparisons between two groups were performed by GraphPad Prism software 6.0 using a two-tailed unpaired t-test. The variance between the statistically compared groups was similar.

DATA AND SOFTWARE AVAILABILITY

Gene Expression Omnibus (GEO) Accession Numbers of Datasets Generated


- GSE85242: The accession number for Illumina Human HT-12 bead array expression profile of doxycycline mediated HNF4G knockdown in 22Rv1 cells reported in this paper is GEO: [GSE85242].
- GSE85244: The accession number for Illumina Human HT-12 bead array expression profile of HNF4G expression in LNCaP cells reported in this paper is GEO: [GSE85244].
- GSE85556: The accession number for RNA-seq expression profile of HNF1A knockdown in 22Rv1 cells and HNF1A overexpression in LNCaP cells respectively reported in this paper is GEO: [GSE85556].
- GSE85557: The accession number for RNA-seq expression profile of LNCaP/AR cells with or without HNF4G exogenous expression grown in charcoal-stripped serum reported in this paper is GEO: [GSE85557].
- GSE85558: The accession number for HNF4G, AR, FOXA1, H3K4me1, H3K27acetyl ChIP-seq and ATAC-seq in 22Rv1 and LNCaP cells with the knockdown and overexpression of HNF4G respectively reported in this paper is GEO: [GSE85558].

ARTICLE

DOI: 10.1038/s41467-018-07511-4

OPEN

NSD2 is a conserved driver of metastatic prostate cancer progression

Alvaro Aytes^{1,2}, Arianna Giacobbe^{1,3}, Antonina Mitrofanova^{4,5}, Katia Ruggero², Joanna Cyrt⁶, Juan Arriaga^{1,3}, Luis Palomero², Sonia Farran-Matas², Mark A. Rubin^{6,7}, Michael M. Shen^{1,3,4,8,9} , Andrea Califano^{4,9,10} & Cory Abate-Shen^{1,3,4,9,11}

Deciphering cell-intrinsic mechanisms of metastasis progression in vivo is essential to identify novel therapeutic approaches. Here we elucidate cell-intrinsic drivers of metastatic prostate cancer progression through analyses of genetically engineered mouse models (GEMM) and correlative studies of human prostate cancer. Expression profiling of lineage-marked cells from mouse primary tumors and metastases defines a signature of de novo metastatic progression. Cross-species master regulator analyses comparing this mouse signature with a comparable human signature identifies conserved drivers of metastatic progression with demonstrable clinical and functional relevance. In particular, *nuclear receptor binding SET Domain Protein 2 (NSD2)* is robustly expressed in lethal prostate cancer in humans, while its silencing inhibits metastasis of mouse allografts in vivo. We propose that cross-species analysis can elucidate mechanisms of metastasis progression, thus providing potential additional therapeutic opportunities for treatment of lethal prostate cancer.

¹Department of Urology, Columbia University Irving Medical Center, 160 Fort Washington Ave, New York, NY 10032, USA. ²Programs of Molecular Mechanisms and Experimental Therapeutics in Oncology (ONCOBell), and Cancer Therapeutics Resistance (ProCURE), Catalan Institute of Oncology, Bellvitge Institute for Biomedical Research, L'Hospitalet de Llobregat, Gran Via de L'Hospitalet, 199, 08908 Barcelona, Spain. ³Department of Medicine, Columbia University Irving Medical Center, 630W 168th Street, New York, NY 10032, USA. ⁴Department of Systems Biology, Columbia University Irving Medical Center, 1130 Saint Nicholas Ave, New York, NY 10032, USA. ⁵Department of Health Informatics, Rutgers School of Health Professions, Rutgers, The State University of New Jersey, 65 Bergen Street, Newark, NJ 07101, USA. ⁶Department of Pathology and Laboratory Medicine, Weill Cornell Medicine, 1300 York Avenue, New York, NY 10065, USA. ⁷Department for BioMedical Research, University of Bern, Murtenstrasse 35, CH-3008 Bern, Switzerland. ⁸Department of Genetics and Development, Columbia University Irving Medical Center, 701 West 168th Street, New York, NY 10032, USA. ⁹Herbert Irving Comprehensive Cancer Center, Columbia University Irving Medical Center, 1130 Saint Nicholas Ave, New York, NY 10032, USA. ¹⁰Department of Biochemistry and Molecular Biophysics, Columbia University Irving Medical Center, 701 West 168th Street, New York, NY 10032, USA. ¹¹Department of Pathology and Cell Biology, Columbia University Irving Medical Center, 630W 168th Street, New York, NY 10032, USA. These authors contributed equally: Alvaro Aytes, Arianna Giacobbe, Antonina Mitrofanova. Correspondence and requests for materials should be addressed to A.A. (email: aaytes@idibell.cat) or to A.C. (email: ac2248@cumc.columbia.edu) or to C.A.-S. (email: cabateshen@columbia.edu)

Metastasis is a complex process that culminates in the progressive accumulation of molecular alterations of cancer cells, which allow them to escape the confines of the tumor, survive during dissemination, and ultimately reside at distant sites, wherein requisite adaptive changes ensue in their new microenvironment^{1,2}. Therefore, it would be most informative to study the biological processes and molecular mechanisms underlying metastatic progression as occur in the context of the whole organism in vivo. However, inherent challenges in accessing primary tumors and their metastases from cancer patients have made it difficult to study de novo metastasis formation. Moreover, most in vivo studies of metastasis have utilized transplantation models wherein cells or tumors are implanted into host organisms, usually immune-deficient ones. While such investigations have advanced our understanding of metastasis mechanisms and have elucidated factors that promote organ tropism³, they may not ideally model the cell-intrinsic mechanisms of de novo metastatic progression. Analyses of genetically engineered mouse models (GEMMs) can overcome these obstacles, since they enable access to tumors and their resultant metastases as they arise de novo during cancer progression in the whole organism^{4–7}.

Virtually all prostate cancer deaths are due to metastasis, which arises at advanced disease stages and is often resistant to treatment. Indeed, while patients with locally confined disease have highly favorable outcomes (>95%), the 5-year survival for metastatic prostate cancer is less than 30%⁸. Standard treatment for advanced prostate cancer involves androgen deprivation therapy, which is initially effective but ultimately leads to disease recurrence in the form of castration resistant prostate cancer (CRPC), which is highly aggressive and prone to metastasis^{9–12}. While second generation anti-androgens, such as enzalutamide and abiraterone acetate, are now being used for treatment of CRPC^{10,11}, treatment failure is often associated with progression to even more aggressive subtypes, including neuroendocrine prostate cancer (NEPC)^{12–15}. Frequent sites of prostate cancer metastasis are bone and lymph nodes, however, visceral metastasis, such as to lungs and liver, are becoming more prevalent in aggressive variants and associated with increased lethality and poor prognosis¹⁶.

Several recent studies have identified the landscape of recurrent genomic alterations in prostate tumors and metastases^{17–30}. The culmination of these analyses has revealed that metastatic prostate cancer has a significantly higher burden of mutational and somatic copy number alterations compared with primary tumors^{20,21,25,27–30}. These include increased frequency of alterations of key oncogenic and tumor suppressor genes such as *AR*, *PTEN*, *TP53*, and *RBI*, and aberrant activation of key signaling pathways such as the PI-3 kinase, FGF receptor, and RAS signaling pathways^{5,31,32}. However, functional analyses of causal drivers of metastatic prostate cancer progression have been hindered by the lack of experimental models that enable biological and molecular investigations of de novo metastasis in context of the whole organism.

In the current study, we have investigated metastatic progression in vivo in a GEMM of prostate cancer. Employing lineage tracing to isolate tumor and metastatic cells, we have defined a molecular signature of metastasis progression. Cross-species computational analyses comparing this mouse signature with a comparable human signature of metastatic prostate cancer progression have identified conserved master regulators of metastasis progression that drive these processes. In particular, we have identified *NSD2* as a conserved master regulator of metastatic prostate cancer progression and a robust marker of lethal prostate tumors. Our findings suggest that cross-species investigations based on analyses of de novo metastasis in GEMMs can be

broadly used to elucidate mechanisms of metastatic progression and identify potential new therapeutic opportunities for treatment of lethal cancer.

Results

A molecular signature of de novo metastasis progression. To elucidate mechanisms of metastasis progression, we utilized a previously described GEMM of highly penetrant metastatic prostate cancer based on an inducible Cre (*Cre^{ERT2}*) expressed under the control of the promoter of the *Nkx3.1* homeobox gene⁵. This *Nkx3.1^{CreERT2}* allele drives Cre-mediated recombination in an appropriate cell of origin of prostate cancer^{33,34} while simultaneously resulting in heterozygosity for *Nkx3.1*, which is prevalent in human prostate cancer^{17,27}. We crossed the *Nkx3.1^{CreERT2}* allele with a *Pten* floxed allele (*Pten^{lox/flox}*) and an activatable mutant *K-Ras* allele (*Kras^{LSL-G12D/+}*) to generate *NPK* mice (for *Nkx3.1^{CreERT2}*; *Pten^{lox/flox}*; *Kras^{LSL-G12D/+}*). Tumor induction of these *NPK* mice leads to co-activation of *PI3K* and *RAS* signaling, as frequently occurs in lethal prostate cancer in human^{5,32}, while these mice develop metastasis with 100% penetrance⁵. These *NPK* mice also contain a conditionally activatable fluorescent reporter allele, *R26R^{YFP}*, which enables in vivo lineage tracing of primary tumors and their resultant metastases with high efficiency and specificity (Fig. 1a)⁵.

To identify a molecular signature of metastatic progression, we compared expression profiles of primary tumors and metastases from *NPK* mice. Since these mice exhibit temporal progression from pre-invasive (~1 month), to invasive prostate cancer (~3 months), and ultimately to metastasis (~5 months)⁵, we analyzed expression profiles of primary tumors from mice prior to the occurrence of overt metastasis (pre-metastatic, <3 months, *n* = 8), as well as primary tumors from mice that had developed overt metastases (post-metastatic, ~5 months, *n* = 8). Further, since these *NPK* mice metastasize primarily to soft tissues, including lung, liver, and lymph node⁵, we analyzed metastases from these various sites (*n* = 8, 5, 7, respectively); however, since lung is the most prevalent metastatic site⁵, we focused our molecular analyses primarily on lung. As controls, we performed comparable analyses using non-metastatic primary tumors from *NP* mice (for *Nkx3.1^{CreERT2}*; *Pten^{lox/flox}*; *n* = 7)³⁵.

To focus on cell-intrinsic molecular features of primary tumor and metastatic cells free of the surrounding stromal or other components of the microenvironment, which are likely to differ for each tissue, we isolated YFP-lineage-traced cells from tumors and metastases using fluorescence activated cell sorting (FACS) (Fig. 1b). We then performed RNA sequencing on the purified YFP-labeled cells to generate expression profiles corresponding to pre- or post-metastatic primary tumors (*n* = 8/group) as well as lung, liver, and lymph node metastases (*n* = 8, 5, 7, respectively; Supplementary Data 1).

Interestingly, we found that the expression profiles of the pre- and post-metastatic primary tumor cells were highly dissimilar, whereas expression profiles from the post-metastatic primary tumors were very similar to those from lung, liver, and lymph node metastases (Supplementary Data 1; Fig. 2, Supplementary Fig. 1). Specifically, unsupervised principal component analysis (PCA) revealed that expression profiles from the post-metastatic *NPK* primary tumors clustered more closely with the lung metastases, as well as the metastases to other sites, and further from the pre-metastatic tumors from these *NPK* mice, whereas the pre-metastatic *NPK* tumors tended to cluster more closely with non-metastatic *NP* primary tumors (Fig. 2a; Supplementary Fig. 1a). This relationship was further confirmed by gene set enrichment analyses (GSEA) wherein a differential expression signature comparing post- versus pre-metastatic primary tumors

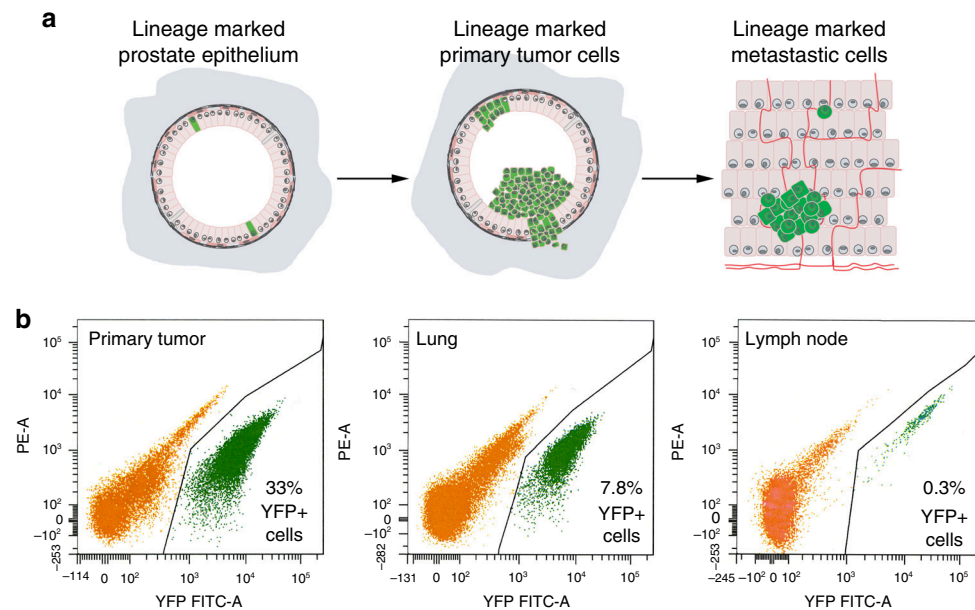


Fig. 1 Strategy for molecular profiling of tumors and metastases. **a** Lineage tracing of YFP-labeled (green) prostate epithelial cells at the time of tumor induction leads to YFP-labeled cells in tumors and metastases. **b** Lineage-marked cells from primary tumors, or lung or lymph node metastases were isolated by fluorescence activated cell sorting (FACS). Shown are representative images with percentages of YFP-labeled cells indicated; axes show fluorescent intensity of the fluorescein isothiocyanate (FITC-A) and phycoerythrin (PE) channels

was significantly enriched with a signature of lung metastases versus pre-metastatic tumors in both the positive (NES = 19.64, $p < 0.001$) and negative (NES = -7.52, $p < 0.001$) leading edges (Fig. 2b). Moreover, expression signatures from the other metastatic sites, namely lung, liver and lymph nodes, were also highly enriched relative to each other (Supplementary Fig. 1b–d), providing further evidence of their similarity.

These observations suggest that the most prominent cell-intrinsic molecular changes that occur during metastatic progression in the *NPK* mice are those that distinguish pre-metastatic from post-metastatic tumors. Hence, taking into consideration: (1) the distinct molecular changes between pre- and post-metastatic *NPK* tumors; (2) the overall similarity of gene signatures of metastatic cells at the various tissue sites (i.e., lung, liver and lymph node); and (3) that the lung is the major metastatic site in the *NPK* mice, our subsequent analyses was done using a signature of metastasis progression based on the differentially expressed genes between the pre-metastatic tumors and lung metastases ($n = 8/\text{group}$, respectively; two-sample two-tailed Welch *t*-test; Supplementary Data 1).

Notably, this mouse metastasis progression signature shares molecular features that have been associated with the hallmarks of metastasis progression in other cancer contexts^{1,2}. In particular, differentially expressed genes from the positive leading edge of the GSEA (from Fig. 2b) include those associated with epithelial-mesenchymal transition (e.g., *Vim*, *Zeb1*, and *Twist2*), cell and focal adhesion (e.g., *Itga5*, *Col4a1*, and *Col4a2*), membrane type matrix metalloproteinases (e.g., *Mmp14*), and developmental pathways (e.g., *Wnt5A*) as well as other genes known to promote metastasis of prostate (e.g., *Etv4*⁵) or other cancers (e.g., *Hmga2*⁷) (Fig. 2c). Similarly, genes from the negative leading edge (from Fig. 2b) include genes associated with the immune response, such as the interferon regulatory factor *Irf7* which has been shown to be a critical regulator of immunosurveillance in cancer metastasis³⁶ (Fig. 2d).

Furthermore, GSEA of biological pathways comparing the mouse metastasis progression signature with the MSigDB Hallmarks dataset revealed a significant enrichment of pathways that

are commonly associated with metastatic progression in other tumor contexts, including epithelial to mesenchymal transition, E2F targets, Myc targets, TGF beta, and P53 pathway among others ($p < 0.05$; Fig. 2e; Supplementary Data 2). Notably, many of the pathways enriched in this metastasis progression signature based on tumor versus lung metastases from the *NPK* mice were also enriched in analogous signatures based on tumor versus liver or lymph node metastases ($p < 0.05$; Supplementary Fig. 1e, f), further emphasizing the overall similarity of cell-intrinsic molecular programs associated with metastasis progression across these various organ sites.

Most notably, the mouse metastasis progression signature was highly conserved with a corresponding signature of human metastatic prostate cancer progression reported by Balk and colleagues, which compares primary tumors with metastatic bone biopsies³⁷ (Supplementary Table 1). In particular, GSEA comparing a “humanized” version of the mouse metastasis progression signature with the Balk human prostate cancer metastasis progression signature demonstrated their significant similarity in both the positive (NES = 8.02, $p < 0.001$) and negative (NES = -3.38, $p = 0.002$) leading edges (Fig. 2f). Furthermore, GSEA comparing a “humanized” version of the mouse post-metastatic versus pre-metastatic progression signature with this human prostate cancer metastasis progression signature also demonstrated strong enrichment in both the positive (NES = 12.12, $p < 0.001$) and negative (NES = -2.67, $p = 0.0035$) leading edges (Supplementary Fig. 1g). Further, the leading edge genes between these GSEA comparisons (i.e., from Fig. 2f and Supplementary Fig. 1g) were highly similar (overlap of the positive leading edges was 90.5% and the negative leading edge was 80%; $X^2 p < 0.0001$). Taken together, these molecular analyses define a cell-intrinsic signature of de novo metastasis progression in the *NPK* mouse model that is highly conserved with metastasis progression of human prostate cancer.

Conserved master regulators of metastasis progression. We performed cross-species computational analyses to identify

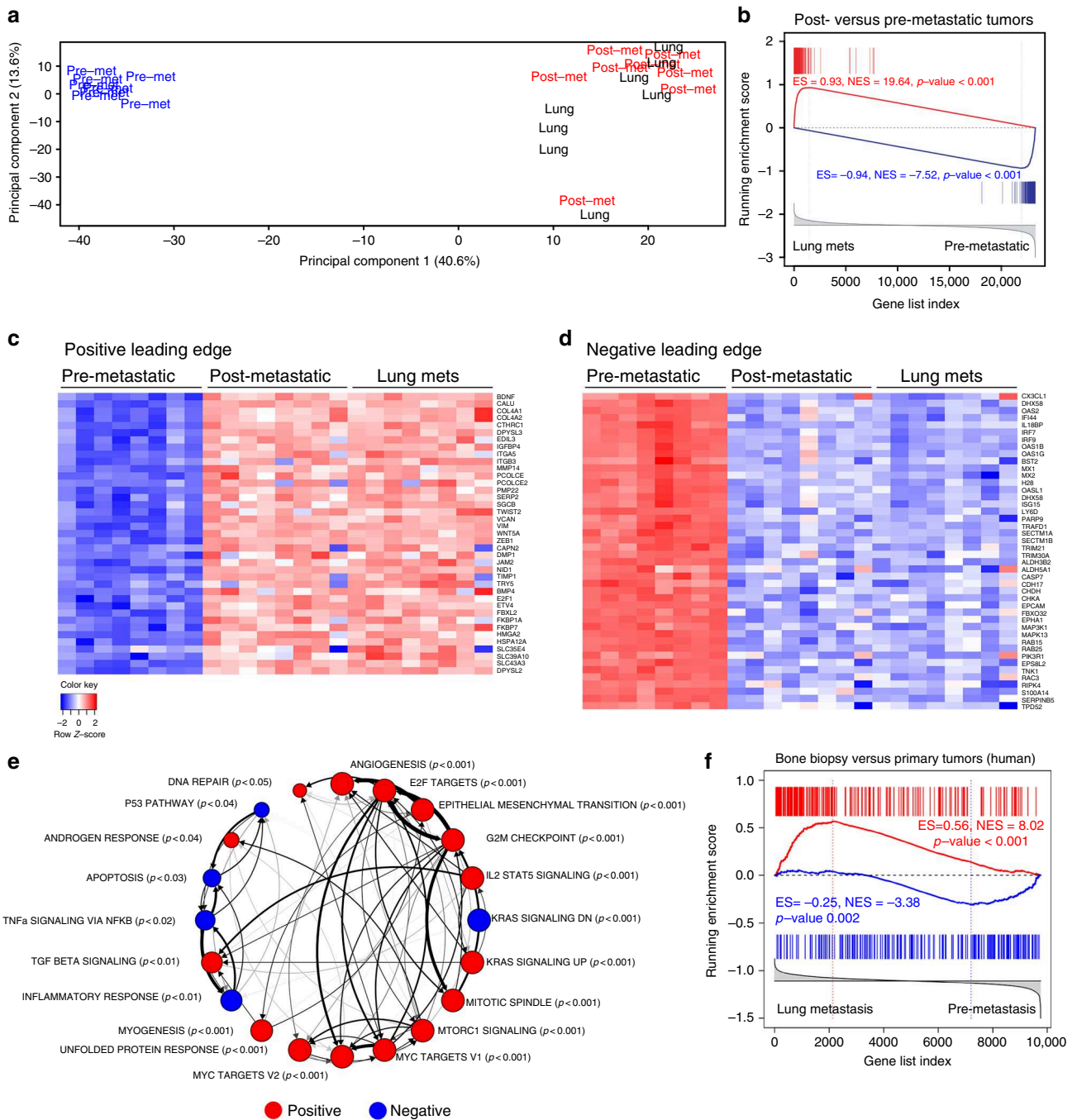


Fig. 2 A molecular signature of de novo metastasis progression. **a** Principal component analysis (PCA) comparing expression profiles from pre-metastatic (pre-met, blue) or post-metastatic (post-met, red) primary tumors, or lung metastases (lung, black) from *NPK* mice ($n = 8$ /group). Note that Principal Component 1, capturing 40.6% of gene expression variance, effectively distinguishes pre-metastatic tumors from post-metastatic tumors and lung metastases. **b** GSEA comparing a reference signature of mouse lung metastases (lung mets) versus pre-metastatic tumors to a query signature of mouse post-metastatic versus pre-metastatic tumors. **c, d** Heat map representations of differentially expressed genes from the positive and negative leading edges, respectively, from the GSEA in panel **b**. The color key shows relative expression levels of the differentially expressed genes (red corresponds to overexpressed genes while blue corresponds to underexpressed genes). **e** Pathway enrichment analysis using the mouse metastasis progression signature defined between lung metastasis versus pre-metastatic tumors, as in panel **b**, to query the hallmark pathways dataset from the molecular signatures database (MSigDB). Red and blue nodes indicate positive and negative enrichment, respectively ($p < 0.05$). Thickness of arrows indicate the overlap of genes in the leading edges. The p -values correspond to the GSEA enrichment, and the relative size of the node indicates the relative p -value, as shown. **f** Cross-species GSEA comparing a reference mouse metastasis progression signature (lung metastasis versus pre-metastatic tumors, as in panel **b**) with a query gene set from a human metastasis signature defined between bone metastasis biopsies versus primary tumors from Balk et al. (Supplementary Table 1). For GSEA, red vertical bars indicate overexpressed query genes and blue vertical bars indicate underexpressed query genes. GSEA were done using the top 200 differentially expressed genes; p -values were calculated using 1000 gene permutations. ES: enrichment score, NES: normalized enrichment score

conserved master regulators (MRs) of metastasis progression by interrogating genome-wide regulatory networks, or interactomes, for mouse and human prostate cancer³⁸, using the master regulator inference analysis (MARINa) algorithm³⁹. First, we interrogated the individual mouse and human prostate cancer interactomes with their respective metastatic progression signatures, which defined independent lists of mouse and human MRs of metastatic progression (Fig. 3a). We subsequently integrated these individual mouse and human MR lists using Stouffer integration to define the subset of conserved candidate MRs ($n = 485$ MRs with Stouffer integrated $p < 0.0001$; Supplementary Data 3). Gene ontology analysis of these integrated MRs revealed an over-representation of genes associated with all aspects of epigenetic regulation, including histone modification, DNA methylation, and chromatin remodeling^{40–42} ($n = 136/485$ genes, 28%; Supplementary Data 3). Because of the potential significance of perturbations of epigenetic regulation for metastatic progression^{43,44}, particularly in lethal prostate cancer²¹, and since epigenetic regulators are potential therapeutic targets^{40–42}, we focused on the subset of conserved MRs that are predicted to function as epigenetic regulators.

To further prioritize these candidate MRs, we used a Cox proportional hazard model to assess the association of their MR activity with prostate cancer-specific survival (where activity for a given MR is defined based on the expression levels of its MR transcriptional targets, see Methods). In particular, we used a human prostate cancer cohort described by Sboner et al., which has more than 30 years of clinical follow-up data based on death due to prostate cancer⁴⁵ (Supplementary Table 1). These analyses identified a subset of 41 MRs whose activities were significantly associated with prostate cancer-specific survival (Wald test Cox $p < 0.05$; Fig. 3b). Among these, we focused our subsequent analysis on 8 candidate MRs: (1) that are associated with adverse disease outcome and prostate-cancer specific lethality (Fig. 3b, c); (2) that are broadly activated in multiple metastatic organ sites (Supplementary Fig. 2a); and (3) whose activities are up-regulated (rather than repressed) in metastasis progression (Fig. 3b, c; Supplementary Fig. 2a), and therefore are potentially targets for treatment inhibition. In particular, these 8 candidate MRs are predicted to be highly activated across multiple metastatic sites, namely lung, liver and lymph node (MARINa $p < 0.001$; Supplementary Fig. 2a), and are significantly associated with adverse disease outcome (Wald test Cox $p \leq 10^{-5}$; Fig. 3c).

We further assessed the clinical relevance of these 8 candidate metastasis MRs using several independent cohorts of advanced prostate cancer patients, including those with clinical endpoints of metastasis or lethality due to prostate cancer (Supplementary Table 1). First, we performed hierarchical clustering on the activity levels of the 8 candidate MRs using the Grasso et al. cohort⁴⁶, which showed that each of these MRs robustly stratify metastases ($n = 35$) from primary tumors ($n = 59$) (Fig. 3d). We observed similar findings with a second cohort that included primary tumors from The Cancer Genome Atlas (TCGA; $n = 497$)²⁷ and metastases from the SU2C cohort ($n = 51$)²¹ (Supplementary Fig. 2b).

Additionally, using two independent patient cohorts with clinical follow-up, we found that these 8 candidate MRs predict disease outcome as evidenced by Kaplan–Meier survival analyses based on MR activity levels. In particular, activity of the candidate MRs stratified prostate cancer patients based on their risk of biochemical recurrence using the Glinsky et al. cohort⁴⁷ ($n = 79$ primary prostate tumors; log-rank $p = 0.00605$; Hazard ratio = 2.452; Fig. 3e). Furthermore, we found that the activities of these candidate MRs also stratified patients based on the risk of death due to prostate cancer in the Sboner et al. cohort⁴⁵ ($n = 281$ primary tumors; log-rank $p = 8.66 \times 10^{-8}$; Hazard ratio = 2.635;

Fig. 3f). Notably, the predictive ability of the 8 candidate MRs for adverse disease outcome in both of these cohorts was highly specific when compared to other MRs selected at random (significance for 8 candidate MRs versus randomly selected MRs was p -value = 0.0011 for Sboner et al. and p -value = 0.0214 for Glinsky et al.; Supplementary Fig. 2c, d).

To evaluate the functional relevance of the 8 candidate MRs for tumor growth and metastasis progression in vivo, we used an allograft cell model derived from the *NPK* mice, which recapitulates the pattern of *NPK* primary tumor growth and metastasis when engrafted into host mice⁵. In particular, we performed shRNA-mediated silencing of each of the candidate MRs using a minimum of 2 shRNAs for each gene (Supplementary Fig. 3a; Supplementary Table 3). Analysis in vitro revealed that *NPK* cells having individually silenced MRs displayed reduced colony formation and reduced invasive potential compared with the control cells, albeit to varying extents for each MR ($p < 0.05$, two-tailed Student's *t*-test; Supplementary Fig. 3b–d). Furthermore, when engrafted in vivo, these MR-silenced *NPK* cells displayed reduced tumor growth ($p < 0.05$, two-way ANOVA) and/or reduced incidence of metastasis compared with the control cells, also to varying degrees for each MR ($p < 0.05$, two-tailed Student's *t*-test; Supplementary Fig. 3e–g). Taken together, these cross-species computational systems analyses have identified conserved master regulators of metastasis progression that are associated with adverse disease outcome and functionally relevant for prostate cancer progression.

NSD2 is a driver of metastatic prostate cancer progression.

Among the candidate MRs, the highest level of MR activity (Fig. 3c) as well as experimentally determined functional activity (Supplementary Fig. 3) were observed for the histone methyltransferase, *Nuclear receptor binding SET Domain protein 2* (*NSD2*). Notably, *NSD2* is a putative cofactor of androgen receptor⁴⁸ that has been previously implicated in advanced prostate cancer^{49–51}, and has been shown to collaborate with *RAS* signaling in other tumor contexts⁵². Therefore, we examined the expression of *NSD2* at the mRNA and protein levels in non-metastatic and metastatic contexts in both mouse and human prostate cancer (Fig. 4). In the mouse prostate, we found that *Nsd2* protein is expressed at low levels in non-metastatic tumors from the *NP* mice, while it is highly expressed in metastatic tumors from the *NPK* mice, as well as corresponding metastases from these mice ($n = 4$ /group; Fig. 4a). Notably, *Nsd2* was robustly expressed in nuclei of *NPK* tumors and lung metastases, coincident with high levels of Ki67, a marker of cell proliferation, strong expression of nuclear androgen receptor (AR), and robust expression of pan-cytokeratin (Pan-Ck) (Fig. 4a).

In human prostate cancer, we found that *NSD2* expression is increased during cancer progression at both the mRNA and protein levels (Fig. 4b–e). In particular, expression of *NSD2* mRNA levels were significantly higher in more advanced (Gleason $\geq 4 + 4$; $n = 104$) versus earlier stage (Gleason $< 4 + 4$; $n = 173$) prostate primary tumors reported in TCGA²⁷ ($p = 2.65 \times 10^{-9}$ two-sample two-tailed Welch *t*-test; Fig. 4b). Further, *NSD2* expression was significantly higher in prostate cancer metastases reported in the SU2C cohort²¹ ($n = 51$) as compared with primary tumors from TCGA ($n = 333$; $p = 1.64 \times 10^{-10}$ two-sample two-tailed Welch *t*-test; Fig. 4c).

To evaluate expression of *NSD2* protein in human prostate cancer, we performed immunohistochemistry on a human prostate cancer tissue microarray ($n = 100$) comprised of benign tumors ($n = 26$), non-lethal prostate adenocarcinoma ($n = 25$), lethal castration-resistance adenocarcinomas (CRPC-Adeno;

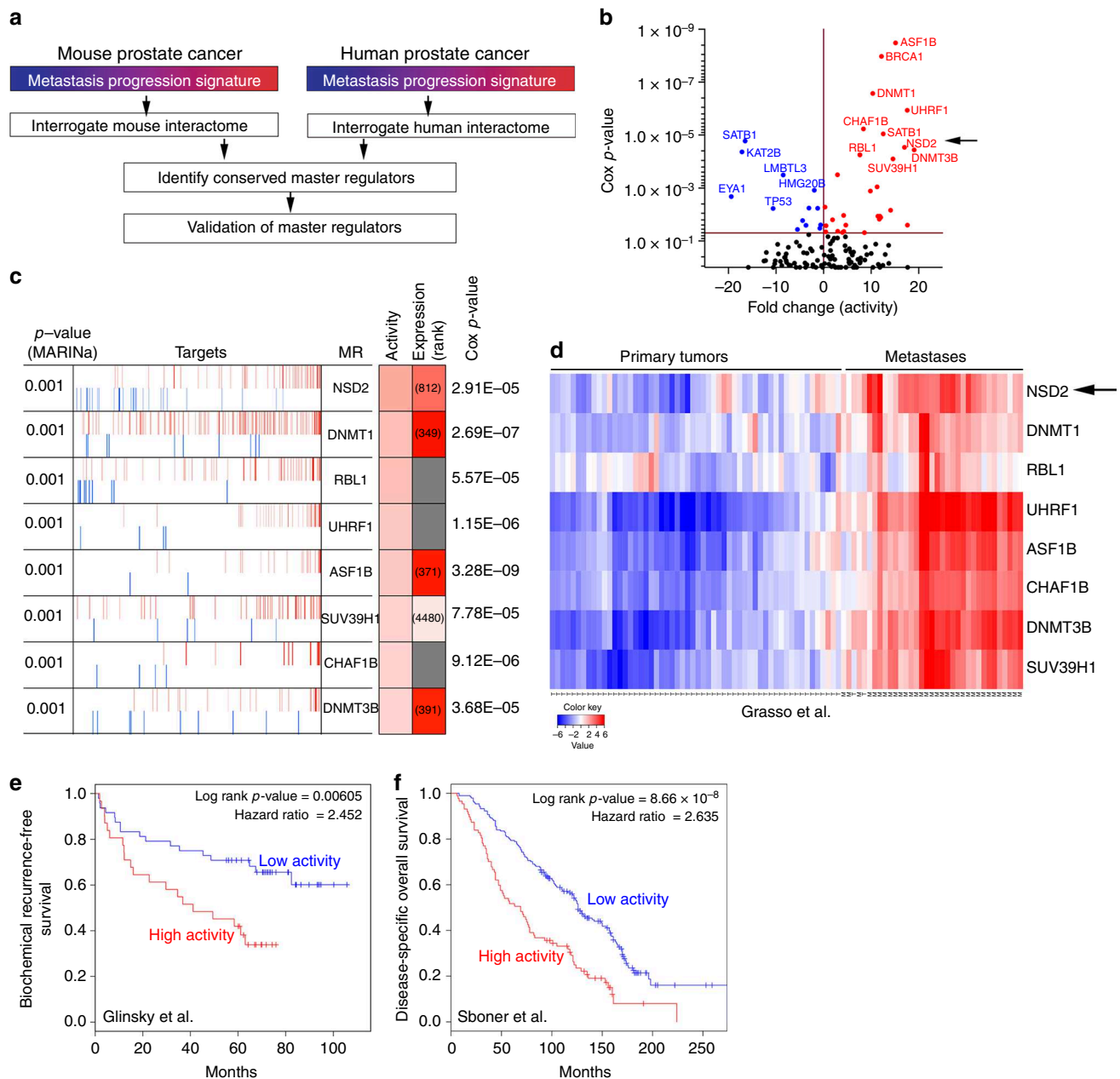


Fig. 3 Conserved master regulators of metastasis progression. **a** Strategy: mouse and human metastasis progression signatures (as in Fig. 2f) were used to interrogate mouse and human prostate cancer interactomes, respectively, using the MARINa algorithm. Independent lists of mouse and human master regulators (MRs) were integrated to identify conserved MRs, which were prioritized by clinical and functional validation. **b** Scatter plot showing the association of the 136 conserved MRs (Supplementary Data 3) to clinical outcome using the Sboner et al. dataset, which reports prostate cancer-specific survival as the clinical endpoint (Supplementary Table 1). The Y axis represents the Cox proportional hazard p -value and the X axis represents the fold change based on MR activity. MRs that are inactive (blue) relative to primary tumors have negative fold change values and those that are active (red) have positive fold change values. **c** Summary of the 8 candidate MRs depicting their positive (activated; red bars) and negative (repressed; blue bars) targets. Shaded boxes show the ranks of differential activity and differential expression (darker is higher and lighter is lower); the numbers indicate their rank in the differential expression signature (gray indicates that a specific gene is not present on the utilized gene expression platform, yet its targets are present). P -values for Cox proportional hazard were estimated using a Wald test based on time to prostate cancer-specific death in Sboner et al. **d** Heatmap showing hierarchical clustering of primary tumors and metastasis from the Grasso et al. cohort (Supplementary Table 1) based on the activities of the 8 candidate MRs. The color key shows activity levels of MRs (i.e., NESs), where red corresponds to increased activity and blue correspond to decreased activity of the MRs. **e, f** Kaplan-Meier survival analysis based on the activity levels of the 8 candidate MRs in: **e** Glinsky et al. ($n = 79$), with biochemical recurrence as the clinical end-point; and **f** Sboner et al. ($n = 281$), with prostate cancer-specific survival as the clinical endpoint (Supplementary Table 1). P -values were estimated using a log-rank test to determine the difference in outcomes between patients with higher MR activity levels (red) versus those with lower/no MR activity (blue)

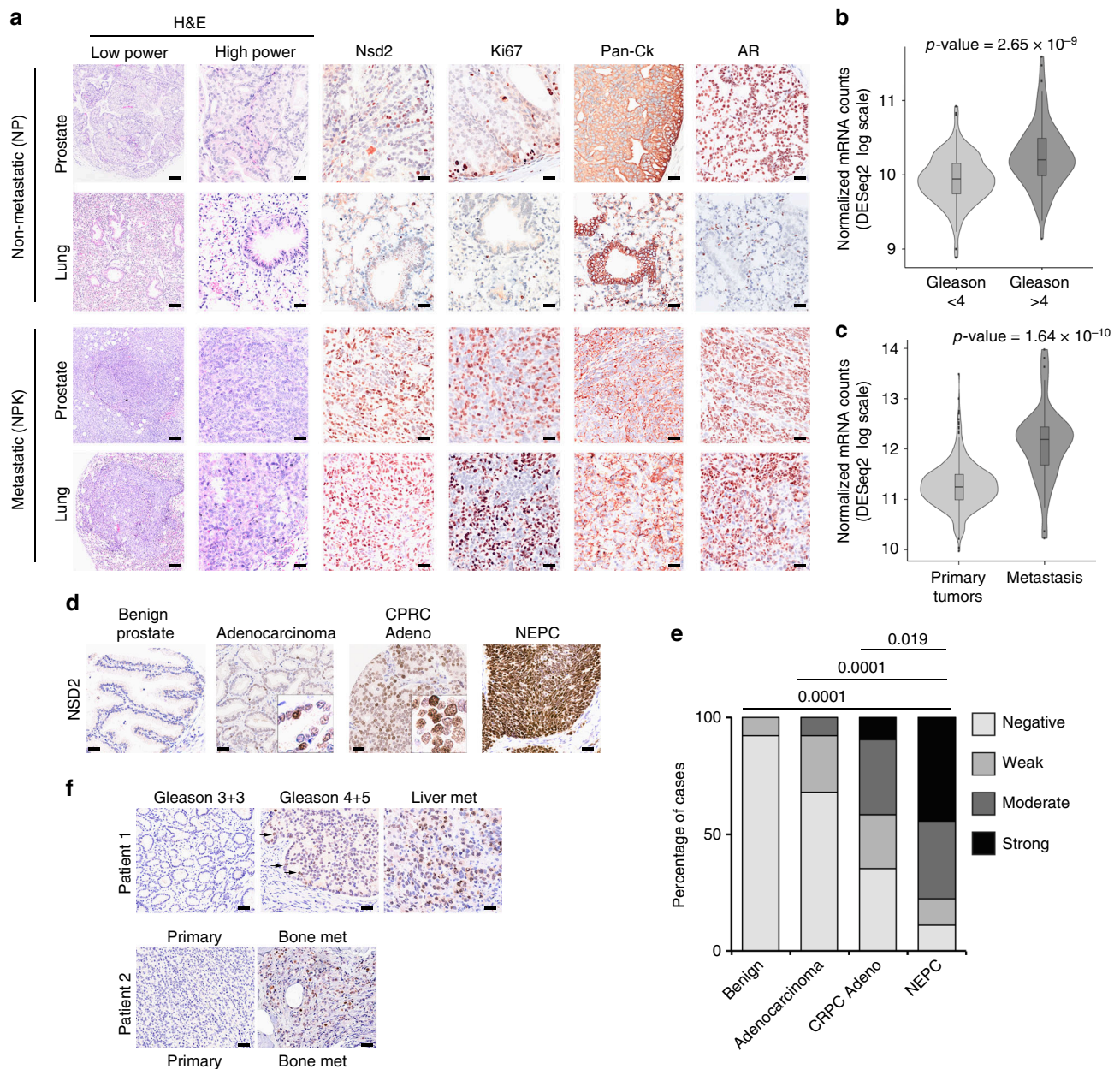


Fig. 4 Expression of *NSD2* in prostate cancer and metastases. **a** Immunostaining of *Nsd2* and other markers on mouse primary tumors and metastases. Shown are representative H&E images and immunostaining for the indicated antibodies from non-metastatic *NP* mice and metastatic *NPK* mice ($n = 4/\text{group}$). Scale bars in the low power H&E images represent 100 microns, and all other panels 50 microns. **b, c** Violin plots comparing mRNA expression levels of *NSD2* in TCGA and SU2C human prostate cancer cohorts (Supplementary Table 1). **b** compares primary tumors from TCGA divided based on pathological grade [Gleason <4 ($n = 104$) or ≥ 4 ($n = 173$)], as indicated. **c** compares primary tumors and metastases from a cohort combining primary tumors from TCGA (all Gleason scores; $n = 333$) and metastases from SU2C ($n = 51$). *P*-values were estimated using the two-sample two-tailed Welch *t*-test. **d, e** Immunostaining of *NSD2* on a human prostate tissue microarray (TMA) ($n = 100$ independent cases). **d** shows representative images representing benign prostate, untreated localized adenocarcinoma, castration-resistant adenocarcinoma (CRPC-Adeno) and neuroendocrine prostate cancer (NEPC). Nuclear staining intensity was evaluated blinded by a pathologist and scored as negative (or present in <5% of nuclei), weak, moderate or strong. Scale bars represent 50 microns. **e** shows quantification of nuclear intensity staining for each score (negative, weak, moderate, and strong). The *p*-values compare negative/weak staining versus moderate/strong staining in each group and were calculated using a two-tailed Fisher's exact test. **f** Immunostaining of *NSD2* on matched patient sets of primary prostate cancer and distant metastasis to soft tissues or bone, as indicated. Patient 1 shows representative images of lower pathological grade (Gleason 3 + 3), which is negative for *NSD2*, and higher pathological grade (Gleason 4 + 5) and a liver metastasis that have increasing expression of *NSD2*. Patient 2 shows a high grade primary tumor (primary) that is negative for *NSD2* and a matched bone metastasis in which *NSD2* staining is readily detected. Scale bars represent 50 μ

($n = 31$), and neuroendocrine prostate tumors (NEPC; $n = 18$) (Fig. 4d). While *NSD2* was either not expressed or expressed at low levels in the non-lethal tumors, its expression increased dramatically in advanced disease stages and was particularly

robust in the most aggressive phenotypes, namely CRPC adenocarcinomas and NEPC ($p < 0.01$, two-tailed Fisher's exact test; Fig. 4e). To further evaluate the relationship of *NSD2* expression with progression to lethal prostate cancer, we

examined matched sets of primary tumors and metastases from the same patient ($n = 3$). Whereas expression in the primary tumors was scattered and focal, NSD2 was robustly expressed in metastasis from these patients (Fig. 4f). These findings extend previous studies showing increased expression of NSD2 in advanced prostate cancer⁴⁹.

To evaluate the functional consequences of NSD2 for disease progression and metastasis, we used the mouse *NPK* metastatic allograft model, as described above, as well as human DU145 prostate cancer cells, which model aggressive disease^{5,38}. In particular, we used lentiviral gene delivery to introduce a minimum of two independent shRNAs to silence NSD2 in either the mouse or human cells, which resulted in effective silencing of NSD2 as evident both at the mRNA and protein levels (Fig. 5a–c).

NSD2 has been reported to function as a histone methyltransferase that targets the histone H3 di-methyl mark on lysine 36 (H3K36me2)^{51,53–55}. Accordingly, we found that silencing of NSD2 in either human or mouse cells resulted in a modest but reproducible reduction of the H3K36me2 mark, while not altering the mono-methyl marks on lysine 36 (H3K36me1) or other histone marks such as tri-methyl lysine 27 (H3K27me3) or lysine 9 (H3K9me3) (Fig. 5c). Furthermore, NSD2 silencing in either mouse *NPK* cells or human DU145 cells in vitro resulted in a 5–10 fold inhibition of colony formation ($p < 0.0001$, two-tailed Student's *t*-test), as well as significantly decreased invasion ($p < 0.01$, two-tailed Student's *t*-test; Fig. 5d–g).

Moreover, analyses of *NPK* metastatic allografts in vivo revealed that *Nsd2* silencing resulted in increased overall survival ($n = 10$ /group; $p = 0.0005$, log-rank; Fig. 6a) as well as a significant reduction of metastatic burden while not affecting primary tumor growth ($n = 9$ /group; $p < 0.03$, Mann–Whitney *U* test; Fig. 6b–d). Notably, these *Nsd2*-silenced tumors had profoundly reduced expression of Nsd2 protein compared with the control tumors, as well as reduced expression of the corresponding H3K36me2 mark, but not other histone marks, such as H3K9me3 or H3K27me3 (Fig. 6e, f). Taken together, these observations demonstrate that increased expression of NSD2 is associated with lethal and metastatic prostate cancer, and establish the functional relevance of NSD2 for metastatic prostate cancer progression.

To consider whether it might be feasible to pharmacologically target NSD2 activity to inhibit prostate cancer progression and tumor growth, we used a small molecule inhibitor of NSD2 called MCTP-39 (3-hydrazinoquinoxaline-2-thiol), which has been reported to be a lysine-HMTase inhibitor that is a competitor of the SAM (Sterile Alpha Motif) domain⁵⁶. We found that MCTP-39 inhibited the H3K36me2 mark, while not affecting other histone marks such as H3K9me3 or H3K27me3 (Fig. 7a); notably, the degree of reduction the H3K36me2 mark following treatment with MCTP-39 in vitro was comparable to the degree that the H3K36me2 mark was reduced following silencing of NSD2 in vitro (Fig. 5c).

Treatment with MCTP-39 in human DU145 cells in vitro resulted in a significant dose-dependent reduction in colony formation (>10 fold; $p < 0.01$ two-tailed Student's *t*-test; Fig. 7b). Since we found that this inhibitor was well tolerated in vivo (Supplementary Fig. 4), we evaluated the effect of MCTP-39 on tumor growth of human prostate cells in vivo by establishing DU145 xenografts (Fig. 7c–f). We found that DU145 xenografts treated with MCTP-39 had a significant decrease in tumor volume ($n = 14$ vehicle-treated and $n = 15$ MCTP-39-treated; $p < 0.001$, two-way ANOVA; Fig. 7c, d). The resulting MCTP-39 treated tumors had reduced expression of the H3K36me2 mark, but not other histone marks such as H3K9me3 or H3K27me3 (Fig. 7e, f). Together with the results of silencing NSD2 in vivo, these findings regarding MCTP-39 treatment

suggest that NSD2 may be a target for intervention in advanced prostate cancer.

Discussion

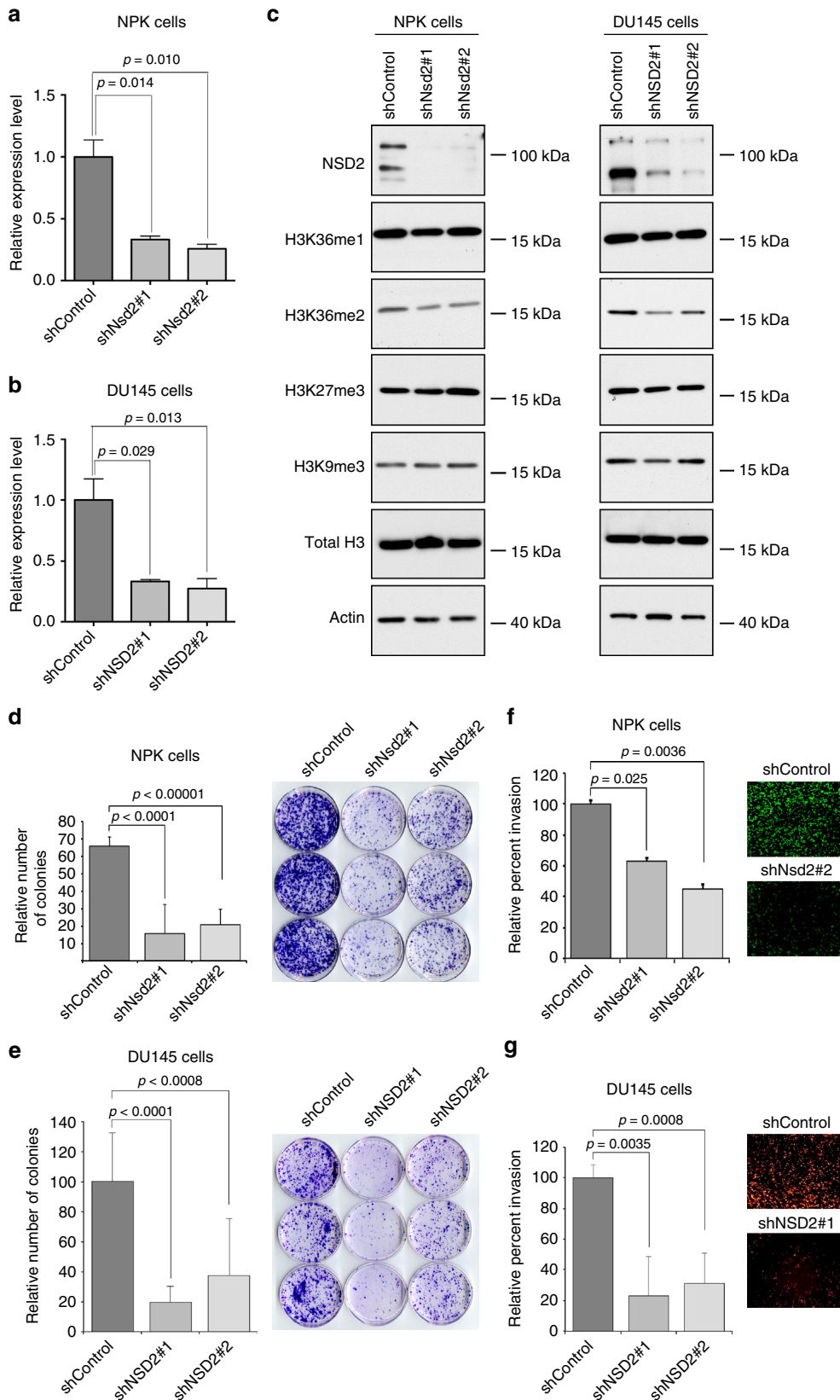
Our study demonstrates the value of cross-species integration of molecular data from genetically engineered mouse models (GEMM) and human cancer to elucidate cell-intrinsic mechanisms of de novo metastasis progression. Notably, our current study, which identifies conserved drivers of metastatic progression by isolation of lineage-marked cells directly from tumors and metastases from a prostate cancer GEMM, complements and extends previous work that identified mechanisms of lung cancer metastasis using cell lines generated from tumors and metastases of lung cancer GEMM⁷. Among its advantages for our current investigations, the *NPK* mouse model displays a highly penetrant metastatic phenotype with consistent temporal progression, and it incorporates in vivo lineage tracing of the primary tumors and metastases. Thus, by comparing metastasis progression signatures from *NPK* mice with comparable signatures from human prostate cancer, we have identified conserved master regulators (MRs) of metastasis progression that are associated with adverse disease outcome. We propose that the general strategy of integrating molecular analyses of tumors and metastases from relevant GEMMs with cross-species computational analyses of human cancer can be broadly adopted to identify new targets for prevention, detection, and potentially treatment of metastasis progression for other cancer types.

The capability of generating transcriptomic data from lineage-marked tumor and metastatic cells from different organ sites that are free from other stromal and tissue-specific cells, allowed us to elucidate cell-intrinsic gene expression changes that occur during cancer progression. Surprisingly, we found that the predominant gene expression differences that occur during metastasis progression arise in the transition from pre-metastatic to metastatic tumors and are shared among metastases from various organs. This is similar to findings of a recent study of metastasis progression in mouse model of pancreatic cancer⁶.

Notably, our analyses of purified tumor and metastatic cells free of other tissue components reveals an overall similarity of cell-intrinsic metastasis progression across the various metastatic sites, thus supporting the concept that organ-site tropic factors may be contributed by the tumor microenvironment at the metastatic site². We speculate that organ-site specific factors act in collaboration with cell-intrinsic drivers of metastasis progression, such as those identified herein. Furthermore, our findings, which suggest that there are common cell-intrinsic drivers of metastasis progression across organ sites in the mouse model, is consistent with a study of human prostate cancer, which reported the inherent similarity of tumors and metastases from the same patient²⁵, and therefore support the feasibility of investigating agents that target metastatic progression in advanced prostate cancer patients.

Interestingly, we find that conserved master regulators of metastasis progression are highly enriched for genes that are predicted to function as regulators of the epigenome, including those that modify DNA and histones, or remodel chromatin architecture. Consistent with our findings, genomic sequencing of prostate tumors has identified several mutations in epigenetic genes particularly in advanced prostate cancer^{20,21,27}. Furthermore, dysregulation of the epigenome is associated with metastatic progression of human prostate cancer^{57,58}.

In particular, we have demonstrated that the *Nuclear receptor binding SET Domain Protein 2* (NSD2) is a robust marker of lethal metastatic prostate cancer and a key driver of prostate cancer metastasis, extending previous studies that have reported



the relevance of NSD2 in prostate cancer^{49–51}. NSD2 was discovered as the overexpressed product of the t(4;14)(p16.3;q32.3) translocation in multiple myeloma, and alternatively named *Multiple Myeloma SET domain containing protein (MMSET)*, and was identified as a target gene on the 4p16 deletion for the Wolf-

Hirschhorn Syndrome, and alternatively called *Wolf-Hirschhorn Syndrome Candidate 1 (WHSC1)*⁵³. Previous studies have shown that genomic alterations occur in other cancer types in addition to multiple myeloma including pediatric leukemia and laryngeal tumors^{59,60}. In prostate cancer, NSD2 has been shown to be up-

Fig. 5 Silencing of *NSD2* abrogates tumorigenicity in vitro. Panels **a–g** show in vitro analyses of *NSD2* silencing in a mouse metastatic cell line (*NPK* cells) and a human advanced prostate cancer cell line (*DU145* cells). Cells were infected with control shRNA or two independent shRNAs for mouse or human *NSD2*, respectively. **a, b** Validation of *NSD2* silencing in *NPK* and *DU145* cells, as indicated, using quantitative real-time PCR (qPCR). **c** Western blot analyses of *NSD2*-silenced or control *NPK* and *DU145* cells, as indicated, showing reduced expression of *NSD2*, which is accompanied by reduction of the H3K36me2 mark, but not the H3K36m1 or the other histone marks shown. The position of a molecular marker is shown; uncropped images are provided in Supplementary Figure 5. **d, e** Colony formation assays in *NPK* and *DU145* cells, as indicated showing quantification (left) representative images (right). **f, g** Invasion assays in *NPK* and *DU145* cells, as indicated showing quantification (left) and representative images (right). Experiments were done in three independent biological replicates each in triplicate; *p*-values were calculated using a two-tailed Student's *t*-test. Error bars represent the standard deviation (s.d.) from the mean

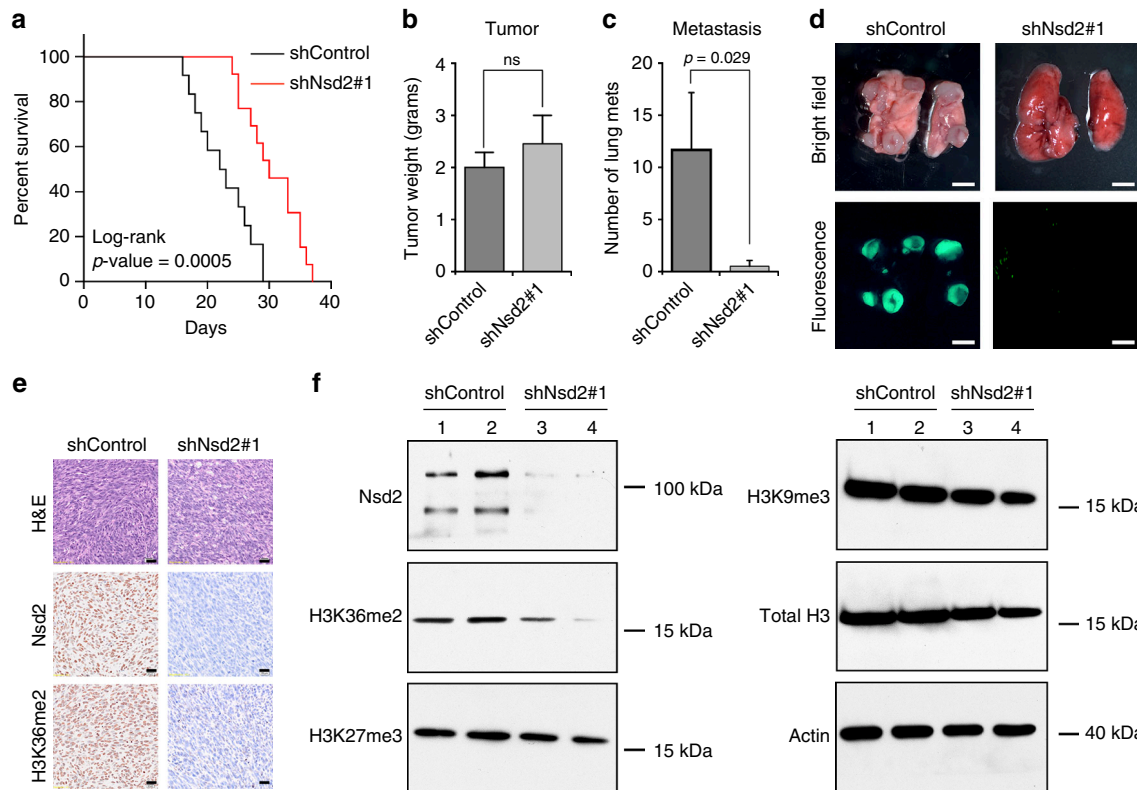


Fig. 6 Silencing of *Nsd2* abrogates metastasis in vivo. Panels **a–f** show in vivo analyses of *Nsd2* silencing in a mouse metastatic cell line (*NPK* cells). Cells (1×10^6 cells) were engrafted subcutaneously into the flank of *nude* mice and the mice were monitored for up to 40 days. Studies were done using 2 independent shRNA for *Nsd2*; representative data for shRNA#1 is shown. **a** Survival analyses with the endpoint being tumor volume of 1.5 cm^3 ($n = 10/\text{group}$). The *p*-value was calculated using a log-rank test. **b** Analyses of tumor weights (in grams) at the time of killing (total $n = 9/\text{group}$). **c** Number of lung metastases per mouse (total $n = 9/\text{group}$). **b, c** *p*-values were calculated using the Mann-Whitney *U* test; error bars represent the standard deviation (s.d.) from the mean. **d** Representative whole mount and epifluorescence images of lung metastases. Scale bar represent 100 microns. **e** Representative immunostaining of shControl and sh*Nsd2* tumors using the indicated antibodies ($n = 4/\text{group}$). Scale bars represent 50 μ . **f** Western blot analysis showing representative cases from the shControl (lanes 1, 2) and sh*Nsd2* (lanes 3, 4) tumors using the indicated antibodies (total $n = 4/\text{group}$). The position of a molecular marker is shown; uncropped images are provided in Supplementary Figure 5

regulated in advanced tumors coordinating with the activation of PI-3 kinase signaling⁴⁹, and to be a cofactor of androgen receptor⁴⁸.

Notably, the role of *NSD2* in cancer has been shown to be dependent on its activity as a histone methyltransferase for the histone H3 di-methyl K36 (H3K36me2)^{50,61,62}. In the current study, we show that MCTP-39, a putative inhibitor of *NSD2*⁵⁶, inhibits prostate tumor growth in vivo. However, several caveats preclude us from drawing the direct conclusion that MCTP-39 is acting to inhibit *NSD2* activity in this context, including the potential activity of unknown metabolites and the potential lack of specificity of MCTP-39 given its relatively simple chemical structure⁵⁶.

Nonetheless, our study demonstrates that *NSD2* is a functional driver of prostate cancer metastasis and suggests that it may be

target for treatment of advanced prostate cancer. Notably, the activity of *NSD2* as a histone methyltransferase has been shown to be coordinately regulated by EZH2⁵¹, a major component of the histone methyltransferase polycomb repressive complex 2 (PRC2), which is also dysregulated in prostate cancer. Additionally, in multiple myeloma, *NSD2* has been shown to be a regulator of DNA damage response that impacts resistance to chemotherapy⁶². These previous studies suggest that combination therapy targeting *NSD2* together with inhibition of PI-3 Kinase, AR, EZH2, and/or DNA repair mechanisms, all of which are themselves targetable and highly relevant for prostate cancer, may prove to be efficacious for treatment of metastatic prostate cancer. We further proposed that these combination treatments can be evaluated in co-clinical assays using the *NPK* mouse model described herein.

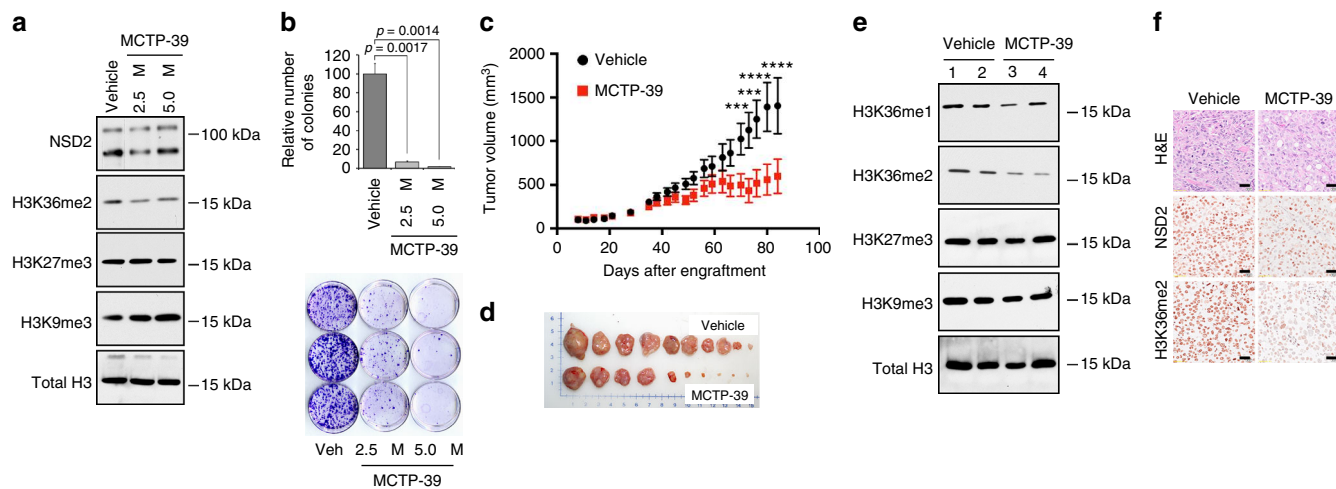


Fig. 7 Pharmacological treatment. **a, b** Pharmacological treatment in vitro. DU145 cells were treated with MCTP-39 at the indicated concentrations for 72 h. Panel a shows western blot data using the indicated antibodies. The position of a molecular marker is shown; uncropped images are provided in Supplementary Figure 6. **b** depicts colony forming assays showing quantification (top) and representative images (bottom). Shown are representative data from 3 independent experiments, each done in triplicate. Error bars represent the standard deviation (s.d.) from the mean; p -values were calculated using a two-tailed Student's t -test. **c-f** Pharmacological treatment in vivo. DU145 cells (5×10^6 cells) were engrafted subcutaneously into male nude mouse hosts. After 1 week of growth, the tumor-bearing mice were randomized by cage to the vehicle (black) or MCTP-39 (red) treatment groups and treated with 10 mg/kg with MCTP-39 (or vehicle only) for up to 3 months. Tumor volume was monitored using calipers and calculated using the formula [Volume = (width)² x length/2]. Total mice analyzed for vehicle were 14 and for MCTP-39-treatment were 15 in two independent experiments. **c** Tumor volume. Two-way analysis of variance (ANOVA) was used to calculate the significance (p -value) of the difference between the vehicle and treatment group; *** $p < 0.001$ and **** $p < 0.0001$. **d** Representative tumors collected at the time of euthanasia. **e** Western blot showing 2 examples from vehicle (lanes 1, 2) and MCTP-39 (lanes 3, 4) treated tumors using the indicated antibodies (total $n = 4$ /group). The position of a molecular marker is shown; uncropped images are provided in Supplementary Figure 6. **f** Representative immunostaining for NSD2 and H3k36me3 from vehicle and MCTP-39 treated mice ($n = 4$ /group). Scale bars represent 50 μ

Methods

Expression profiling of lineage-marked cells. All experiments using animals were performed according to protocols approved by and following all ethical guidelines required by the Institutional Animal Care and Use Committee (IACUC) at Columbia University Irving Medical Center or the Ethics Committee for Animal Research (CEIC) at Bellvitge Biomedical Research Institute. For molecular profiling analyses, lineage-marked cells from primary tumors and/or metastases were collected from *Nkx3.1^{CreERT2/+}; Pten^{flxed/flxed}; R26R^{YFP} (NP)* and *Nkx3.1^{CreERT2/+}; Pten^{flxed/flxed}; Kras^{Isl-G12D/+}; R26R^{YFP} (NPK)* mice, which have been previously published^{5,33,35}. Note that inclusion of the lox-stop-lox *R26^{YFP}* allows for lineage tracing specifically in prostate cells at the time of tumor induction⁵. These NP, and NPK mice have been maintained in our laboratory on a predominantly C57BL/6 background. All studies were done using littermates that were genotyped prior to enrollment; mice were randomly enrolled to treatment or control groups and only male mice were used because of the focus on prostate.

Mice were induced to form tumors at 2–3 months of age by administration of tamoxifen (Sigma-Aldrich, Allentown, PA) using 100 mg/kg once daily for 4 consecutive days and monitored for 1 to 9 months, during which time the NPK mice develop prostate adenocarcinoma that progresses to overt metastasis⁵. At the time of killing, prostate tumors from NP or NPK mice, as well as tissues with overt metastases from the NPK mice, as detected by ex-vivo fluorescence, were collected in ice cold phosphate-buffered saline (PBS). Tissues were digested in one part of collagenase/hyaluronidase (Stem Cell Technologies, Cambridge, MA) and nine parts of DMEM-F12 and 10% fetal bovine serum (FBS) at 37 °C for 3 h. Samples were pelleted at 350XG in an Eppendorf 5810 R tabletop centrifuge for 5 min at 4 °C, re-suspended in 0.25% trypsin/EDTA (Stem Cell Technologies, Cambridge, MA), and incubated for 1 h on ice. Cells were collected by centrifugation as above, and incubated in a cocktail of pre-warmed dispase (5 mg/ml) plus DNaseI (1 mg/ml) (Stem Cell technologies, Cambridge, MA) for 10 min at 37 °C; after which, cells were filtered through a 40 μ m cell strainer, pelleted and re-suspended in 1% PBS/FBS and proceed to the sorter. The YFP-lineage marked cells were purified using a BD FACS Aria II sorter and the YFP + population isolated using the PE/FITC (R-Phycoerythrin/Fluorescein isothiocyanate) channels to gate the YFP + population. Cell pellets were resuspended in 10 μ l of Trizol and flash-frozen in liquid nitrogen.

RNA was prepared using a MagMAX-96 total RNA isolation kit (Life technologies). Total RNA was enriched for mRNA using poly-A pull-down; only samples having between 200 ng and 1 μ g with an RNA integrity number (RIN) > 8 were used. Library preparation was done using an Illumina TruSeq RNA prep-kit, and the libraries were sequenced using an Illumina HiSeq2500 by multiplexing samples in each lane, which yields targeted number of single-end/paired-end 100 bp reads for each sample, as a fraction of 180 million reads for the whole lane. Raw

counts were normalized and the variance was stabilized using DESeq2 package (Bioconductor) in R-studio 0.99.902, R v3.3.0 (The R Foundation for Statistical Computing, ISBN 3-900051-07-0). A complete list of differentially expressed genes is provided in Supplementary Data 1.

Cross species computational analyses. Differential gene expression signatures were defined as a list of genes ranked by their differential expression between any two phenotypes of interest (e.g., metastases vs primary tumors) estimated using a two-sample two-tailed Welch t -test. For cross species analyses, the human gene expression signatures were defined based on published prostate cancer cohorts (Supplementary Table 1) and the mouse gene expression signatures were generated from the RNA sequencing analyses as described above (Supplementary Data 1). For the mouse signatures, a minimum of 5 samples were used for each group as necessary to estimate statistical significance in the two-sample two-tailed Welch t -test and GSEA.

For comparison with human genes, mouse genes were mapped to their corresponding human orthologs based on the homoloGene database (NCBI) so that mouse-human comparisons were done using the “humanized” mouse signatures. For gene set enrichment analysis (GSEA) normalized enrichment score (NES) and p -values were estimated using 1,000 gene permutations. Pathway enrichment analysis was done using GSEA to query the Hallmark Pathways dataset from the MSigDB (i.e., Molecular Signatures Database) collections available from the Broad Institute, where NES and p -value were estimated using 1000 gene permutations (Supplementary Data 2). Master regulator (MR) analysis was performed using the MAster Regulator INference algorithm (MARINA) to query the mouse and human prostate cancer interactomes, respectively, as published previously³⁸ (Supplementary Data 3).

Master regulator activity analyses. Transcriptional activity of master regulators (MRs) was estimated using expression levels of their transcriptional targets and reflects their enrichment in the signature being queried. In particular, targets of a particular MR are used as a query gene set to estimate their enrichment in the reference signature of interest (e.g., metastatic progression signature). If positive targets are overexpressed (i.e., enriched in the overexpressed tail of the reference signature) and/or negative targets are underexpressed (i.e., enriched in the underexpressed tail of the reference signature), such MR is considered active (i.e., its transcriptional activity is positive). If the converse is the case, the MR is repressed (i.e., its transcriptional activity is negative).

The relationship of MR activity levels for clinical outcome was assessed using four independent datasets (Supplementary Table 1): Sboner et al., which reports

death due to prostate cancer as the clinical end-point⁴⁵; Grasso et al., which reports metastases versus primary tumors as a binary outcome⁴⁶; Glinsky et al., which reports biochemical recurrence (BCR) as the clinical end-point⁴⁷; and a combined SU2C²¹ and TCGA²⁷ cohort (i.e., cohorts were combined on raw count levels and normalized using DESeq2 package), which report castration-resistant metastases and primary prostate tumors as a binary outcome, respectively. Sboner et al. was utilized for Cox proportional hazard model analysis (MR filtering/discovery step) and subsequent confirmatory Kaplan–Meier survival analysis. Glinsky et al. was utilized as an independent validation dataset for Kaplan–Meier survival analysis. Grasso et al. and combined SU2C and TCGA cohort were utilized to evaluate the efficacy of MRs stratification of primary tumor versus metastasis.

Immunohistochemical analysis. All studies involving human subjects were approved by the Institutional Review Board at Weil Cornell Medical School. Only anonymized tissues were used and patient consent was obtained. The cohort included benign prostate tissue ($n = 26$), untreated localized adenocarcinoma (with a representative range of different Gleason scores) ($n = 25$), castration-resistant adenocarcinoma (CRPC-Adeno) ($n = 31$) and neuroendocrine prostate cancer (NEPC) ($n = 18$). Subtype and grading were assigned as defined by pathology and clinical criteria as described⁶³. Immunohistochemistry was performed on formalin-fixed paraffin-embedded sections using a Bond III automated immunostainer and the Bond Polymer Refine detection system (Leica Microsystems, IL, USA). Slides were de-paraffinized and heat-mediated antigen retrieval was performed using the Bond Epitope Retrieval 2 solution at pH9 and incubated with the anti-NSD2 primary antibody (Supplementary Table 2). Nuclear staining intensity was evaluated by a pathologist and scoring was done blinded and defined as negative (or present in <5% of nuclei), weak, moderate, or strong. Immunostaining of mouse prostate tissues and metastases was done as described previously^{5,35}. Briefly, 3 μ m paraffin sections were deparaffinized in xylene, followed by antigen retrieval in antigen unmasking solution (Vector Labs, Burlingame, CA). Slides were blocked in 10% normal goat serum, then incubated with primary antibody overnight at 4 °C, followed by incubation with secondary antibodies for 1 h. For immunostaining, the signal was enhanced using the Vectastain ABC system and visualized with NovaRed Substrate Kit (Vector Labs). All antibodies used in this study, as well as antibody dilutions, are described in Supplementary Table 2.

Functional validation studies. Mouse cell lines were isolated from lung metastases from NPK mice and their genotype was authenticated as described previously⁵. Human cell lines were purchased from and authenticated by ATCC (American Type Culture Collection). Cells were grown in RPMI media supplemented with 10% FBS (ThermoFisher, Bridgewater, NJ). Only early passage cells were used for all studies herein. Cells were routinely tested to ensure that they are free of mycoplasma using the MycoFluor Mycoplasma Detection Kit (Invitrogen™, Carlsbad, CA). For shRNA-mediated silencing, a minimum of two independent shRNA clones were used for each gene using the pLKO.1 lentiviral vector system following manufacturer's instructions (Sigma-Aldrich, Allentown, PA). The sequences for all mouse and human shRNA used in this study are provided in Supplementary Table 3. As a control, we used a pLKO.1 lentiviral vector with shRNA targeting the Luciferase gene (SHC007, Sigma-Aldrich).

Analysis of RNA expression was done by quantitative real time PCR using the QuantiTect SYBR Green PCR kit (Qiagen, Germantown, MD) using mouse or human *GADPH* as the control^{5,35}. Sequences of all primers are provided in Supplementary Table 3.

Western blot analysis was performed using total protein extracts as described^{5,35}. Briefly, cells were lysed with 1X radioimmunoprecipitation assay (RIPA) buffer (0.1% SDS, 1% deoxycholate sodium salt, 1.0% Triton-X 100, 0.15 M NaCl, 10 mM Tris-HCl (pH 7.5), 1 mM EDTA supplemented with protease inhibitor cocktail (Roche, 11 836 153 001), 1% phosphatase inhibitor cocktail 3 (Sigma, P0044), and 0.5 % PMSF (Sigma). Protein lysates (5 μ g per lane) were resolved by SDS-PAGE and transferred onto a PVDF membrane (GE Healthcare, Amersham), then blocked with PBS-T (phosphate-buffered saline and 0.1 Tween-20) containing 5% non fat dry milk. Incubation with primary antibody was done at 4° overnight, followed by incubation with secondary antibody for 1 h at room temperature. Detection was performed using the ECL Plus Western Blotting Detection Kit (GE Healthcare/Amersham, New York). A list of all antibodies and antibody dilutions is provided in Supplementary Table 2. Uncropped images are provided in Supplementary Figures 5 and 6.

Colony formation and invasion assays were done as described⁵. Briefly, for colony formation, 1×10^3 cells were seeded in triplicate in three independent experiments (aggregate total $n = 9$) in 60-mm plates and grown for 10 days in RPMI-1640 media (Gibco, Bridgewater, NJ) supplemented with 10% FBS (BenchMark™ Gemini Bio-Products, Sacramento, CA). Colonies were visualized by staining with crystal violet and quantified using ImageJ software (National Institute of Health website). For matrigel invasion assays, 2.5×10^4 cells were seeded in three independent experiments (aggregate total $n = 9$) in BD FluoroBlok inserts (BD Biosciences, Billerica, MA) in FBS-free media and media supplemented with 10% FBS was used in the lower chamber as a chemoattractant.

For in vivo assays, 1×10^6 cells were injected subcutaneously into the flank of immunodeficient NCr nude mice (male, Taconic, Rensselaer, NY); we performed

two independent experiments each done using 10 mice/group. Tumors were monitored by caliper measurement twice weekly for approximately 4 weeks and tumor volumes were calculated using the formula [Volume = (width)² x length/2]. At the time of euthanasia, tumors were harvested and weighed, and the number of lung metastases was documented ex vivo by visualizing their fluorescence using an Olympus SZX16 microscope equipped with epifluorescence capabilities. The total number of metastatic nodules in the lung was assessed for each mouse and the p -value was calculated by comparing the control and each shRNA silences using a two-tailed Student's t -test or the Mann–Whitney U test.

For analyses of NSD2 inhibition, MCTP-39 (3-hydrazinylquinoxaline-2-thiol) was purchased from Enamine L.t.d. (Ukraine)⁵⁶. For studies performed in mice, the MCTP-39 was further purified at the Organic Synthesis Core Facility (OSCF) at Memorial Sloan Kettering Cancer Center. The MCTP-39 was subjected to silica gel chromatography, microfiltered on a 0.2 μ m Teflon membrane and lyophilized. The resulting solid was analyzed by UPLC (ultrahigh pressure liquid chromatography) on reverse phase C18 silica gel in addition to ¹H and ¹³C NMR (Bruker 600 MHz machine).

For in vitro assays, 3×10^5 DU145 cells were seeded in triplicate in three independent experiments (aggregate total $n = 9$) in a 100 mm dish. 24 h following plating, the compound MCTP-39 (5 mM in DMSO) was added to a final concentration of 2.5 μ M or 5 μ M and incubated for 72 h; DMSO was used as a vehicle control. Cells were collected and lysed for western blot analysis or were seeded for colony formation assays as described above.

For in vivo studies, 5×10^6 DU145 cells in 50% matrigel were injected subcutaneously into the flanks of nude mice. When tumors reached 200 mm³, mice were allocated into the vehicle (1% Carboxymethylcellulose, 0.1% Tween80, 5% DMSO) or MCTP-39 (10 mg/kg in vehicle) groups using cage-based randomization, which were administered by oral gavage daily for 3 months. Tumor volume was measured by calipers twice weekly, and estimated by the formula [Volume = (width)² x length/2]. Total mice analyzed were: vehicle = 14; MCTP-39 = 15 (two independent experiments). Two-way analysis of variance (ANOVA) was used to calculate the significance (p -value) of the difference between the vehicle and treatment groups.

Statistical analyses. The Cox model and Kaplan–Meier analysis were done using the *surv* and *coxph* functions from the *survcomp* R package (Bioconductor). Statistical significance was estimated with Wald test and log-rank test, respectively. For Kaplan–Meier survival analysis, k -means clustering was done on the activity levels of the MRs to cluster patients into two groups: one group with increased activity of the candidate MRs and one group with decreased MR activity. To compare the predictive ability of candidate MRs to results at random, we selected a random (equally sized, $n = 8$) group of MRs and utilized their activity levels for Kaplan–Meier survival analysis. This procedure was repeated 1000 times and log-rank p -values from all iterations were used to build a Null model. The empirical p -value was then estimated as a number of times log-rank p -values for a random group of 8 MRs reached or outperformed our original log-rank p value for the identified 8 MRs. R-studio 0.99.902, R v3.3.0, were used for statistical calculations and data visualization.

Data availability

The unique raw and normalized RNAseq files are available at Gene Expression Omnibus (GSE111291). A Reporting Summary for this Article is available as a Supplementary information file. All the other data supporting the findings of this study are available within the Article and its Supplementary Information files or from the corresponding authors upon reasonable request.

Received: 3 February 2018 Accepted: 6 November 2018

Published online: 05 December 2018

References

- Lambert, A. W., Pattabiraman, D. R. & Weinberg, R. A. Emerging biological principles of metastasis. *Cell* **168**, 670–691 (2017).
- Vanharanta, S. & Massague, J. Origins of metastatic traits. *Cancer Cell*. **24**, 410–421 (2013).
- Bos, P. D., Nguyen, D. X. & Massague, J. Modeling metastasis in the mouse. *Curr. Opin. Pharmacol.* **10**, 571–577 (2010).
- Maddipati, R. & Stanger, B. Z. Pancreatic cancer metastases harbor evidence of polyclonality. *Cancer Discov.* **5**, 1086–1097 (2015).
- Aytes, A. et al. ETV4 promotes metastasis in response to activation of PI3-kinase and Ras signaling in a mouse model of advanced prostate cancer. *Proc. Natl Acad. Sci. USA* **110**, E3506–E3515 (2013).
- Chiou, S. H. et al. BLIMP1 induces transient metastatic heterogeneity in pancreatic cancer. *Cancer Discov.* **7**, 1184–1199 (2017).
- Winslow, M. M. et al. Suppression of lung adenocarcinoma progression by Nkx2-1. *Nature* **473**, 101–104 (2011).

8. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2017. *CA Cancer J. Clin.* **67**, 7–30 (2017).
9. Scher, H. I. & Sawyers, C. L. Biology of progressive, castration-resistant prostate cancer: directed therapies targeting the androgen-receptor signaling axis. *J. Clin. Oncol.* **23**, 8253–8261 (2005).
10. Watson, P. A., Arora, V. K. & Sawyers, C. L. Emerging mechanisms of resistance to androgen receptor inhibitors in prostate cancer. *Nat. Rev. Cancer* **15**, 701–711 (2015).
11. Attard, G. et al. Prostate cancer. *Lancet* **387**, 70–82 (2016).
12. Chang, A. J., Autio, K. A., Roach, M. 3rd & Scher, H. I. High-risk prostate cancer-classification and therapy. *Nat. Rev. Clin. Oncol.* **11**, 308–323 (2014).
13. Aparicio, A. M. et al. Combined tumor suppressor defects characterize clinically defined aggressive variant prostate cancers. *Clin. Cancer Res.* **22**, 1520–1530 (2016).
14. Aggarwal, R., Zhang, T., Small, E. J. & Armstrong, A. J. Neuroendocrine prostate cancer: subtypes, biology, and clinical outcomes. *J. Natl. Compr. Canc. Netw.* **12**, 719–726 (2014).
15. Vlachostergios, P. J., Puca, L. & Beltran, H. Emerging variants of castration-resistant prostate cancer. *Curr. Oncol. Rep.* **19**, 32 (2017).
16. Halabi, S. et al. Meta-analysis evaluating the impact of site of metastasis on overall survival in men with castration-resistant prostate cancer. *J. Clin. Oncol.* **34**, 1652–1659 (2016).
17. Baca, S. C. et al. Punctuated evolution of prostate cancer genomes. *Cell* **153**, 666–677 (2013).
18. Barbieri, C. E. et al. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nat. Genet.* **44**, 685–689 (2012).
19. Cooper, C. S. et al. Analysis of the genetic phylogeny of multifocal prostate cancer identifies multiple independent clonal expansions in neoplastic and morphologically normal prostate tissue. *Nat. Genet.* **47**, 367–372 (2015).
20. Beltran, H. et al. Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. *Nat. Med.* **22**, 298–305 (2016).
21. Robinson, D. et al. Integrative clinical genomics of advanced prostate cancer. *Cell* **161**, 1215–1228 (2015).
22. Gundem, G. et al. The evolutionary history of lethal metastatic prostate cancer. *Nature* **520**, 353–357 (2015).
23. Abida, W. et al. Prospective genomic profiling of prostate cancer across disease states reveals germline and somatic alterations that may affect clinical decision making. *JCO Precis. Oncol.* **2017**, 1–26 (2017).
24. Berger, M. F. et al. The genomic complexity of primary human prostate cancer. *Nature* **470**, 214–220 (2011).
25. Kumar, A. et al. Substantial interindividual and limited intraindividual genomic diversity among tumors from men with metastatic prostate cancer. *Nat. Med.* **22**, 369–378 (2016).
26. Fraser, M. et al. Genomic hallmarks of localized, non-indolent prostate cancer. *Nature* **541**, 359–364 (2017).
27. Cancer Genome Atlas Research, N. The molecular taxonomy of primary prostate. *Cancer Cell* **163**, 1011–1025 (2015).
28. Espiritu, S. M. G. et al. The evolutionary landscape of localized prostate cancers drives clinical aggression. *Cell* **173**, 1003–1013 (2018).
29. Quigley, D. A. et al. Genomic hallmarks and structural variation in metastatic prostate. *Cancer Cell* **174**, 758–769 (2018).
30. Viswanathan, S. R. et al. Structural alterations driving castration-resistant prostate cancer revealed by linked-read genome sequencing. *Cell* **174**, 433–447 (2018).
31. Bluemn, E. G. et al. Androgen receptor pathway-independent prostate cancer is sustained through FGF signaling. *Cancer Cell* **32**, 474–489 (2017).
32. Taylor, B. S. et al. Integrative genomic profiling of human prostate cancer. *Cancer Cell* **18**, 11–22 (2010).
33. Wang, X. et al. A luminal epithelial stem cell that is a cell of origin for prostate cancer. *Nature* **461**, 495–500 (2009).
34. Wang, Z. A., Toivanen, R., Bergren, S. K., Chambon, P. & Shen, M. M. Luminal cells are favored as the cell of origin for prostate cancer. *Cell Rep.* **8**, 1339–1346 (2014).
35. Zou, M. et al. Transdifferentiation as a mechanism of treatment resistance in a mouse model of castration-resistant prostate cancer. *Cancer Discov.* **7**, 736–749 (2017).
36. Bidwell, B. N. et al. Silencing of Irf7 pathways in breast cancer cells promotes bone metastasis through immune escape. *Nat. Med.* **18**, 1224–1231 (2012).
37. Stanbrough, M. et al. Increased expression of genes converting adrenal androgens to testosterone in androgen-independent prostate cancer. *Cancer Res.* **66**, 2815–2825 (2006).
38. Aytes, A. et al. Cross-species regulatory network analysis identifies a synergistic interaction between FOXM1 and CENPF that drives prostate cancer malignancy. *Cancer Cell* **25**, 638–651 (2014).
39. Walsh, L. A. et al. An integrated systems biology approach identifies TRIM25 as a key determinant of breast cancer metastasis. *Cell Rep.* **20**, 1623–1640 (2017).
40. Brien, G. L., Valerio, D. G. & Armstrong, S. A. Exploiting the epigenome to control cancer-promoting gene-expression programs. *Cancer Cell* **29**, 464–476 (2016).
41. Dawson, M. A. & Kouzarides, T. Cancer epigenetics: from mechanism to therapy. *Cell* **150**, 12–27 (2012).
42. Popovic, R. & Licht, J. D. Emerging epigenetic targets and therapies in cancer medicine. *Cancer Discov.* **2**, 405–413 (2012).
43. Flavahan, W. A., Gaskell, E., & Bernstein, B. E., Epigenetic plasticity and the hallmarks of cancer. *Science* **357**, eaal2380 (2017).
44. Ell, B. & Kang, Y. Transcriptional control of cancer metastasis. *Trends Cell Biol.* **23**, 603–611 (2013).
45. Sboner, A. et al. Molecular sampling of prostate cancer: a dilemma for predicting disease progression. *BMC Med. Genom.* **3**, 8 (2010).
46. Grasso, C. S. et al. The mutational landscape of lethal castration-resistant prostate cancer. *Nature* **487**, 239–243 (2012).
47. Glinisky, G. V., Gliniskii, A. B., Stephenson, A. J., Hoffman, R. M. & Gerald, W. L. Gene expression profiling predicts clinical outcome of prostate cancer. *J. Clin. Invest.* **113**, 913–923 (2004).
48. Kang, H. B. et al. The histone methyltransferase, NSD2, enhances androgen receptor-mediated transcription. *FEBS Lett.* **583**, 1880–1886 (2009).
49. Li, N. et al. AKT-mediated stabilization of histone methyltransferase WHSC1 promotes prostate cancer metastasis. *J. Clin. Invest.* **127**, 1284–1302 (2017).
50. Ezponda, T. et al. The histone methyltransferase MMSET/WHSC1 activates TWIST1 to promote an epithelial-mesenchymal transition and invasive properties of prostate cancer. *Oncogene* **32**, 2882–2890 (2013).
51. Asangani, I. A. et al. Characterization of the EZH2-MMSET histone methyltransferase regulatory axis in cancer. *Mol. Cell* **49**, 80–93 (2013).
52. Garcia-Carpizo, V. et al. NSD2 contributes to oncogenic RAS-driven transcription in lung cancer cells through long-range epigenetic activation. *Sci. Rep.* **6**, 32952 (2016).
53. Bennett, R. L., Swaroop, A., Troche, C., & Licht, J. D., The role of nuclear receptor-binding SET domain family histone lysine methyltransferases in cancer. *Cold Spring Harb. Perspect. Med.* **7**, a026708 (2017).
54. Sankaran, S. M. & Gozani, O. Characterization of H3.3K36M as a tool to study H3K36 methylation in cancer cells. *Epigenetics* **12**, 917–922 (2017).
55. Sankaran, S. M., Wilkinson, A. W., Elias, J. E. & Gozani, O. A PWWP domain of histone-lysine N-methyltransferase NSD2 binds to dimethylated Lys-36 of histone H3 and regulates NSD2 function at chromatin. *J. Biol. Chem.* **291**, 8465–8474 (2016).
56. Chinnaiyan, A. M., Lnu, S., Cao, Q. & Asangani I. The reagents of the University of Michigan, Ann Arbor, MI (US). Compositions and methods for inhibiting MMSET. US Patent 8,697,407 B2 (2014).
57. Bianco-Miotto, T. et al. Global levels of specific histone modifications and an epigenetic gene signature predict prostate cancer progression and development. *Cancer Epidemiol. Biomark. Prev.* **19**, 2611–2622 (2010).
58. Geysels, M. S. et al. Epigenetic signature of Gleason score and prostate cancer recurrence after radical prostatectomy. *Clin. Epigenetics* **8**, 97 (2016).
59. Jaffe, J. D. et al. Global chromatin profiling reveals NSD2 mutations in pediatric acute lymphoblastic leukemia. *Nat. Genet.* **45**, 1386–1391 (2013).
60. Peri, S. et al. NSD1- and NSD2-damaging mutations define a subset of laryngeal tumors with favorable prognosis. *Nat. Commun.* **8**, 1772 (2017).
61. Kuo, A. J. et al. NSD2 links dimethylation of histone H3 at lysine 36 to oncogenic programming. *Mol. Cell* **44**, 609–620 (2011).
62. Shah, M. Y. et al. MMSET/WHSC1 enhances DNA damage repair leading to an increase in resistance to chemotherapeutic agents. *Oncogene* **35**, 5905–5915 (2016).
63. Epstein, J. I. et al. Proposed morphologic classification of prostate cancer with neuroendocrine differentiation. *Am. J. Surg. Pathol.* **38**, 756–767 (2014).

Acknowledgements

We acknowledge support from the Flow Cytometry Core, the JP Sulzberger Columbia Genome Center, and the Small Animal Imaging Facility, which are supported in part by NIH/NCI grant #P30 CA013696. This research was supported by funding from the National Cancer Institute to C.A.S. (CA173481), M.M.S., C.A.S. (CA196662), A.C. and C.A.S. (U54 CA209997), M.A.R. and M.M.C. (P50 CA211024), and A.C. (R35 CA197745). C.A.S. and M.M.S. are supported by the TJ Martell Foundation for Leukemia, Cancer and AIDS Research. A.A. was supported by grants from the Instituto de Salud Carlos III (PI16/01070 and CP15/00090), the European Association of Urology Research Foundation (EAURF/407003/XH), Fundacion BBVA, the CERCA Program / Generalitat de Catalunya, and FEDER funds/ European Regional Development Fund (ERDF)-a way to Build Europe. AG was supported by post-doctoral training grants from the DOD Prostate Cancer Research Program (PC150959) and an International Cancer Research Fellowship Outgoing funded by the Associazione Italiana per la Ricerca sul Cancro (AIRC) and Marie Curie Actions COFUND and a Department of Defense Award (W81XWH-18-1-0193). J.C. was supported by a fellowship from the Nuovo-Soldati Foundation. J.A. was supported by a post-doctoral training grant from the DOD Prostate Cancer Research Program (PC141549), an Irving Institute/Clinical Trials Office (CTO) Pilot Award (UR002765-01), and a Precision Medicine Research Fellow funded by the Irving Institute for Clinical and Translational Research (UL1TR001873). C.A.S. is an American Cancer Society Research Professor supported in part by a generous gift from the F.M. Kirby Foundation.

Author contributions

A.A., A.M., and A.G. designed and performed experiments and analyzed data. K.R., J.C., J. A., L.P., and S.F.-M. performed experiments and analyzed data. M.A.R. and M.M.S. supervised the study. A.C., A.A., and C.A.S. designed experiments and supervised the study. C.A.S., A.A., A.G., and A.M. wrote the manuscript; all authors provided comments.

Additional information

Supplementary Information accompanies this paper at <https://doi.org/10.1038/s41467-018-07511-4>.

Competing interests: M.A.R. receives research support from Jansen, Eli Lilly, and Sanofi-Aventis; is a co-inventor on gene fusion prostate cancer for diagnostic and therapeutic uses, and is a co-founder Thucydx, LLC. A.A. and L.P. receive support from Roche Pharma and Astellas Pharma directed to support the ProCURE research program. A.C. is a founder, equity holder, and serves on the advisory board of DarwinHealth Inc. The remaining authors declare no competing interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018

ARTICLE

DOI: 10.1038/s41467-018-04495-z

OPEN

Patient derived organoids to model rare prostate cancer phenotypes

Loredana Puca^{1,2,3}, Rohan Bareja^{3,4}, Davide Prandi⁵, Reid Shaw⁶, Matteo Benelli⁵, Wouter R. Karthaus⁷, Judy Hess¹, Michael Sigouros¹, Adam Donoghue¹, Myriam Kossai⁸, Dong Gao⁷, Joanna Cyrta³, Verena Sailer³, Aram Vosoughi³, Chantal Pauli³, Yelena Churakova³, Cynthia Cheung³, Lesa Dayal Deonaraine², Terra J. McNary³, Rachele Rosati⁶, Scott T. Tagawa^{1,2}, David M. Nanus^{1,2}, Juan Miguel Mosquera^{2,3,8}, Charles L. Sawyers⁷, Yu Chen⁷, Giorgio Inghirami⁸, Rema A. Rao³, Carla Grandori⁶, Olivier Elemento^{2,3,4}, Andrea Sboner^{3,4}, Francesca Demichelis^{3,5}, Mark A. Rubin^{3,8} & Himisha Beltran^{1,2,3}

A major hurdle in the study of rare tumors is a lack of existing preclinical models. Neuroendocrine prostate cancer is an uncommon and aggressive histologic variant of prostate cancer that may arise de novo or as a mechanism of treatment resistance in patients with pre-existing castration-resistant prostate cancer. There are few available models to study neuroendocrine prostate cancer. Here, we report the generation and characterization of tumor organoids derived from needle biopsies of metastatic lesions from four patients. We demonstrate genomic, transcriptomic, and epigenomic concordance between organoids and their corresponding patient tumors. We utilize these organoids to understand the biologic role of the epigenetic modifier EZH2 in driving molecular programs associated with neuroendocrine prostate cancer progression. High-throughput organoid drug screening nominated single agents and drug combinations suggesting repurposing opportunities. This proof of principle study represents a strategy for the study of rare cancer phenotypes.

¹ Department of Medicine, Division of Hematology and Medical Oncology, Weill Cornell Medicine, New York, NY 10021, USA. ² Meyer Cancer Center, Weill Cornell Medicine, New York, NY 10021, USA. ³ Englander Institute for Precision Medicine, Weill Cornell Medicine-New York Presbyterian Hospital, New York, NY 10021, USA. ⁴ Institute for Computational Biomedicine, Weill Cornell Medicine, New York, NY 10021, USA. ⁵ Center for Integrative Biology, University of Trento, 38123 Trento, Italy. ⁶ Cure First and SEngine Precision Medicine, Seattle, WA 98109, USA. ⁷ Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA. ⁸ Department of Pathology and Laboratory Medicine, Weill Cornell Medicine, New York, NY 10021, USA. Correspondence and requests for materials should be addressed to H.B. (email: hip9004@med.cornell.edu)

Prostate cancer is the most common cancer in men and second leading cause of male cancer death in the United States¹. Nearly all prostate cancer patients are diagnosed with prostate adenocarcinoma, which arises as an androgen-driven disease. Therefore, a highly effective therapeutic approach for patients with advanced disease is androgen deprivation therapy with gonadal suppression with or without the addition of chemotherapy or the potent androgen synthesis inhibitor abiraterone acetate^{2,3}. However despite initial responses, castration resistance ultimately ensues. With recent therapeutic advances including more effective and earlier use of androgen receptor (AR)-targeted therapies, the landscape of castration-resistant prostate cancer (CRPC) is evolving⁴. While the majority of CRPC tumors remain AR-driven through the acquisition of activating AR mutations, amplification, splice variants, bypass, or other means, up to 10–20% of CRPC tumors lose AR dependence as a means to evade AR-targeted therapy⁴. One extreme manifestation is transformation from an AR-positive adenocarcinoma to an AR-negative small cell neuroendocrine carcinoma characterized by distinct morphologic features⁵. While small cell carcinoma of the prostate rarely arises de novo, castration-resistant small cell neuroendocrine prostate cancer evolves clonally from prostate adenocarcinoma during disease progression retaining early prostate cancer genomic alterations and acquiring distinct genomic, epigenetic, and pathway changes⁶. Patients with either de novo small cell neuroendocrine prostate cancer or castration-resistant neuroendocrine prostate cancer (CRPC-NE) are often treated with platinum-based chemotherapy similar to patients with small cell lung cancer; however, prognosis is poor and there are no known effective therapies beyond platinum.

While in vivo models have been described to model small cell neuroendocrine prostate cancer, the only widely available cell line is the NCI-H660 cell line, derived from a patient initially thought to have small cell lung cancer but later reclassified as prostate based on the presence of the prostate cancer-specific *TMPRSS2-ERG* gene fusion⁷. To expand on this unmet need, we developed patient-derived organoids from metastatic biopsies from patients with CRPC-NE. We molecularly characterized these new models and illustrate how they may be utilized to manipulate the expression and activity of oncogenes involved in the establishment of the neuroendocrine phenotype. High-throughput drug screening of patient-organoids nominated novel drug targets and combinations for CRPC-NE.

Results

Development of patient-derived tumor organoid and xenograft models. Fresh tumor tissue from 25 patients with metastatic prostate cancer was used for organoid development with an overall patient success rate of 16% (4/25) (Fig. 1a). Both three-dimensional (3D) and two-dimensional monolayer (2D) organoid-derived cell lines were successfully developed from four patients (liver, lymph node, soft tissue, and bone biopsy sites; Fig. 1b) and propagated (median 12 months) (Fig. 1c). During early passages, a cytology smear was performed to confirm the presence of tumor cells in the culture⁸ (Fig. 1d) and cancer-associated fibroblasts were isolated and propagated separately for further planned studies on the tumor microenvironment (Supplementary Fig. 1). The organoids were also engrafted as patient-derived organoid xenografts (PDOXs) using NOD scid gamma (NOD.Cg-Prkdc^{scid} Il2rg^{tm1Wjl}/SzJ) mice and subsequently re-passaged in vitro as organoids from PDOXs (PDOX-ORG) (Supplementary Fig. 2).

The pathology of each of the four patient's metastatic tumor and their matched organoids and PDOXs was classified as neuroendocrine prostate cancer based on tumor morphology,

including both pure small cell carcinoma and high-grade carcinoma with extensive neuroendocrine differentiation, and were characterized by the presence of small- to medium-sized round cells with fine chromatin pattern and nuclear molding^{9,10}. All four organoids and their PDOXs lacked AR protein expression and expressed classical neuroendocrine markers by immunohistochemistry (Fig. 1e–h).

Molecular characterization of neuroendocrine models. To determine how genomically stable the organoids are, we performed whole-exome sequencing (WES) of organoids in both 2D and 3D culture conditions and at serial timepoints (passages 10 and 35) and compared these results with the patient's matched metastatic tumor biopsy and PDOX. Tumor purity of all models including organoids, PDOXs, and PDOX-derived organoids estimated by CLONET¹¹ was high (median 98%) (Supplementary Fig. 3). Genome-wide copy number alterations were concordant across models and with time including genes commonly altered in advanced prostate cancer^{6,12} (Fig. 2a, Supplementary Fig. 4).

Using RNA-seq and principal component analysis, we compared the patient organoids and matched PDOX transcriptome profiles with a published cohort of 26 localized prostate adenocarcinoma (PCA), 33 metastatic castration-resistant adenocarcinoma (CRPC-Adeno), and 13 CRPC-NE patient tumors, and found consistent segregation of the CRPC-NE organoids with CRPC-NE patient tumors (Fig. 2b). The organoids and PDOXs clustered based on their shared expression of CRPC-NE signature genes⁶ (Fig. 2c) including overexpression of *MYCN*¹³, *PEG10*¹⁴, *SRRM4*¹⁵, *EZH2*⁶, *SOX2*¹⁶, *BRN2*¹⁷, and *FOXA2*² (Supplementary Fig. 5), and low expression of AR signaling genes¹⁸ (Supplementary Fig. 6). There were no significant differences in gene expression between 2D and 3D cultures (correlation coefficient 0.934, Supplementary Fig. 7) or in media with or without DHT (Supplementary Fig. 8).

Although the retinoblastoma gene *RB1*, commonly deleted in CRPC-NE and other small cell carcinomas, was not lost at the genomic level in any of our CRPC-NE organoids, transcriptome analysis revealed pathway dysregulation consistent with *RB1* loss¹⁹. This suggests a loss of function of the *RB1* pathway by other means, which we found in these cases due to aberrant phosphorylation of *RB1* and/or inactivation/deletion of CDK-inhibitor p16ink4a (*CDKN2A*) (Supplementary Fig. 9), mechanisms previously described^{20,21}.

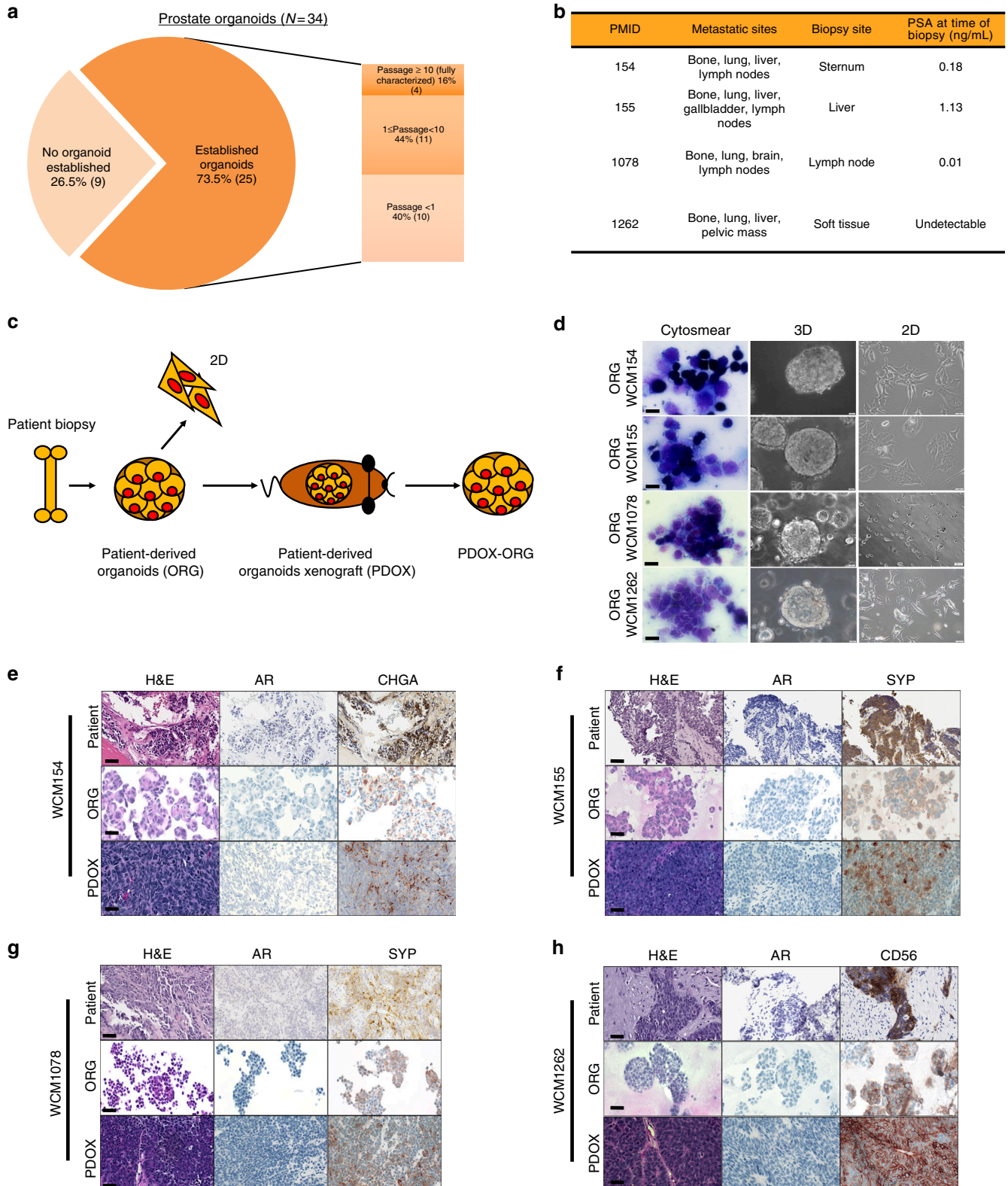
Based on the marked epigenomic changes previously reported in CRPC-NE patient biopsies⁶, we also evaluated CpG-rich methylation in organoids on a genome-wide scale using enhanced reduced representation bisulfite sequencing (ERRBS). The patient-derived organoid models clustered with their corresponding patient tumors based on DNA methylation as well as with other CRPC-NE cases using our published datasets (Fig. 2d).

Based on the presence of genomic alterations involving common cancer-associated genes (Supplementary Fig. 10) and mRNA and DNA methylation clustering with patient tumors of the same disease state, the models appeared representative of their matched patient and of CRPC-NE. The high tumor purity of the models, consistent IHC analysis of common markers across tumor cells, and lack of expression of benign markers (i.e., benign liver marker Hep Par 1 in the liver biopsy-derived organoid (Supplementary Fig. 11)) suggested limited cellular heterogeneity and supported a lack of contamination by benign tissues but also the inability to maintain features of the microenvironment or multiple tumor populations with time.

Effects of EZH2 inhibition. The histone methyltransferase enhancer of zeste 2 (*EZH2*) is an epigenetic modifier frequently

overexpressed in many cancer types including prostate cancer and supports cancer cell proliferation and survival^{22–25}. Recent work by our group and others has identified EZH2 as a potential mediator of CRPC-NE progression^{6,13,26–28}. We evaluated a larger cohort of patients and identified overexpression of EZH2 protein expression in the majority (87%) of CRPC-NE tumors ($n = 15$) compared with 46% of CRPC-Adeno ($n = 26$),

5% localized prostate adenocarcinoma ($n = 21$), and minimal to no expression in benign prostate tissues (0%, $n = 34$). Overexpression of EZH2 was associated with concomitant increased EZH2 activity (i.e., H3K27me3 expression). The levels of EZH2 and H3K27me3 were comparable in CRPC-NE organoid and PDOX models (Fig. 3a, b and Supplementary Fig. 12) and expression of EZH2 in the nuclei was visualized using



immunofluorescence (Fig. 3c). The cell cycle regulator E2F1 positively controls EZH2 transcription²⁹. As expected, the expression of EZH2 and E2F1 in our patient cohorts and organoids was highly correlated ($r = 0.879$, p value $< 2.2 \times 10^{-16}$) (Supplementary Fig. 13).

To gain further insights into how epigenetic modulation might affect neuroendocrine prostate cancer progression, we successfully infected human CRPC-NE organoids with short hairpin RNA targeting EZH2 or a scramble sequence. Knockdown of EZH2 resulted in a reduction of its activity measured through the H3K27 methylation and a decrease in expression of classical neuroendocrine markers including synaptophysin (SYP) but remained AR-negative (Fig. 3d, e). By gene set enrichment analysis (GSEA), we found a significant upregulation of EZH2-suppressed target genes and downregulation of stem cell and neuronal programs after knockdown (Fig. 3f). EZH2 has been associated with stem cell properties and tumor-initiating cell function in different cancer types including glioblastoma, breast, and pancreatic cancers^{24,30}. GSEA of organoids treated with the EZH2 inhibitor GSK503 demonstrated similar results as with shRNA (Fig. 3f, Supplementary Fig. 14 and Supplementary Data1), though neuronal pathways did not reach statistical significance (p value < 0.05 and FDR < 0.25). Again, no upregulation was observed in AR expression or AR signaling by interfering with EZH2 activity (Fig. 3g). Taken together, these data suggest that EZH2 is associated with CRPC-NE program dysregulation, but suppression of EZH2 alone is not sufficient to re-express AR in this late-stage AR-negative CRPC-NE state. This differs from what has been recently described in other prostate cancer models of lineage plasticity in which EZH2 inhibition resulted in re-expression of the AR^{26,28}, possibly due to an earlier more “plastic” disease state in those models where AR was not completely absent prior to therapy.

To understand whether inhibition of EZH2 activity could be considered as a treatment option for CRPC-NE, we treated CRPC-NE organoids with the EZH2 inhibitors GSK343 and GSK503. This resulted in a reduction of H3K27me3 expression (Supplementary Fig. 15) and a preferential decrease in the viability in CRPC-NE organoids at high doses compared to CRPC-Adeno organoids used as control³¹ (Fig. 3h, Supplementary Fig. 16). These results were confirmed measuring cell death by annexin staining (Supplementary Fig. 17). To reinforce the terminal differentiation hypothesis, when we treated CRPC-NE organoids with an EZH2 inhibitor in combination with the AR antagonist enzalutamide, no additive effects or synergy were observed (Supplementary Fig. 18). These data suggest that EZH2 inhibition has activity in CRPC-NE and this does not require expression of the AR. However given the high doses required, combination therapies may be required similar to what has been described in other cancer types^{26,32,33}.

High-throughput drug screening. Given a lack of therapeutic options available for patients with CRPC-NE and our observed EZH2 inhibitor single agent activity, we explored the activity of existing drugs and drug combinations by performing a high-throughput drug dose–response screen using a drug library of 129 chemotherapeutics and targeted agents³⁴. We tested the four CRPC-NE organoids as well as two CRPC-Adeno organoids as controls³¹.

As expected, drugs approved for patients with CRPC-Adeno including enzalutamide, an AR-antagonist, and the taxane chemotherapies cabazitaxel and docetaxel^{35,36} were identified as active in CRPC-Adeno organoids based on the drug screen. The drug screening results for CRPC-NE vs CRPC-Adeno organoids nominated a modest number of drugs such as pozotinib (HER) and vandetanib (VEGFR2) more effective in killing CRPC-NE over control CRPC-Adeno tumor cells (Fig. 4a, Supplementary Fig. 19).

High-throughput drug screening also highlighted patient-specific sensitivities (Fig. 4b). For instance, the CRPC-NE organoid OWCM155 exhibited significant sensitivity to the aurora kinase inhibitor alisertib consistent with the corresponding patient’s exceptional response in the Phase 2 trial of alisertib for CRPC-NE (NCT01482962) (Beltran et al. ESMO)³⁷. We confirmed responses to single agents including alisertib and GSK343 in vitro using both cancer and benign prostate cell line and organoids (Supplementary Fig. 20). CRPC-NE organoid OWCM1078 similarly responded well to alisertib (Supplementary Fig. 21). On the other hand, the CRPC-NE organoid OWCM154 did not respond to alisertib in vitro (nor did the patient on the phase 2 clinical trial) but demonstrated response to the MEK inhibitor cobimetinib (Fig. 4c). These data support a potential role of organoid drug screening to predict individual patient responses to therapy. Drug screening also identified drugs predicted by genomic alterations. For instance OWCM155, which harbored *PTEN* deletion by WES and had high basal levels of phospho-AKT, was particularly sensitive to AKT inhibition (AZD5363, afuresertib) (Supplementary Fig. 22).

The CRPC-NE and CRPC-Adeno organoids were tested in a drug combination screen adding sub-lethal doses of the EZH2 inhibitor, GSK503. For OWCM154, one of the top GSK503 combinations that enhanced the effect of the single agent was with alisertib (AURKA) (Fig. 4d, e and Supplementary Fig. 23). As described the OWCM154 organoid (and corresponding patient) was resistant to alisertib as single agent; these data suggest that targeting two pathways implicated as cooperators for CRPC-NE progression¹³ may be an effective approach and may be picked up through an unbiased screen. For the alisertib responder organoids, other drugs nominated as effective in combination with GSK503 included the EGFR class of inhibitors (neratinib, afatinib, erlotinib, and osimertinib). The CRPC-Adeno organoids did not show increased sensitivity to these combinations.

Fig. 1 Development of patient-derived neuroendocrine prostate cancer models. **a** Pie chart of prostate cancer needle biopsies considered for the generation of organoids. No organoid established (light orange) represents no viable cells or no cellular material was found in culture after enzymatical digestion of the tissue. Established organoids ($p < 1$, lighter orange or $1 \geq p > 10$ orange) refers to the presence of clusters of viable cells in culture that became senescent after few passages in culture. Passage ≥ 10 , dark orange, indicates organoids that have been successfully grown in culture, molecularly characterized and used for functional studies and PDOX development. **b** Table of clinical data and biopsy sites of fully characterized CRPC-NE organoids. **c** Schematic view of the models generated from needle biopsies. In the scheme a patient biopsy is processed to generate 3D organoids (ORG). 3D organoids are then used to generate 2D cultures (2D) and also engrafted in an NSG mouse to grow patient-derived organoids xenograft (PDOX) and consequent organoids derived from the PDX (PDOX-ORG). **d** Air Dried Diff-Quick stained smears of organoids from small cell neuroendocrine prostate carcinoma and high grade adenocarcinomas with neuroendocrine features organoids ($\times 40$, scale bar 50 μm). Bright field image analysis ($\times 40$ magnification for 3D, $\times 20$ magnification for 2D, scale bar 100 μm) of 3D and 2D organoids. **e–h** Histology images of native tumor biopsy tissue (Patient) compared with corresponding 3D organoid cultures (ORG), patient-derived organoids xenograft tissue PDOX ($\times 20$ magnification, scale bar 100 μm patient images, $\times 40$, scale bar 50 μm for models). Samples are stained with hematoxylin-and-eosin (H&E) and immunohistochemistry (IHC) for AR, synaptophysin (SYP), chromogranin A (CHGA), and CD56

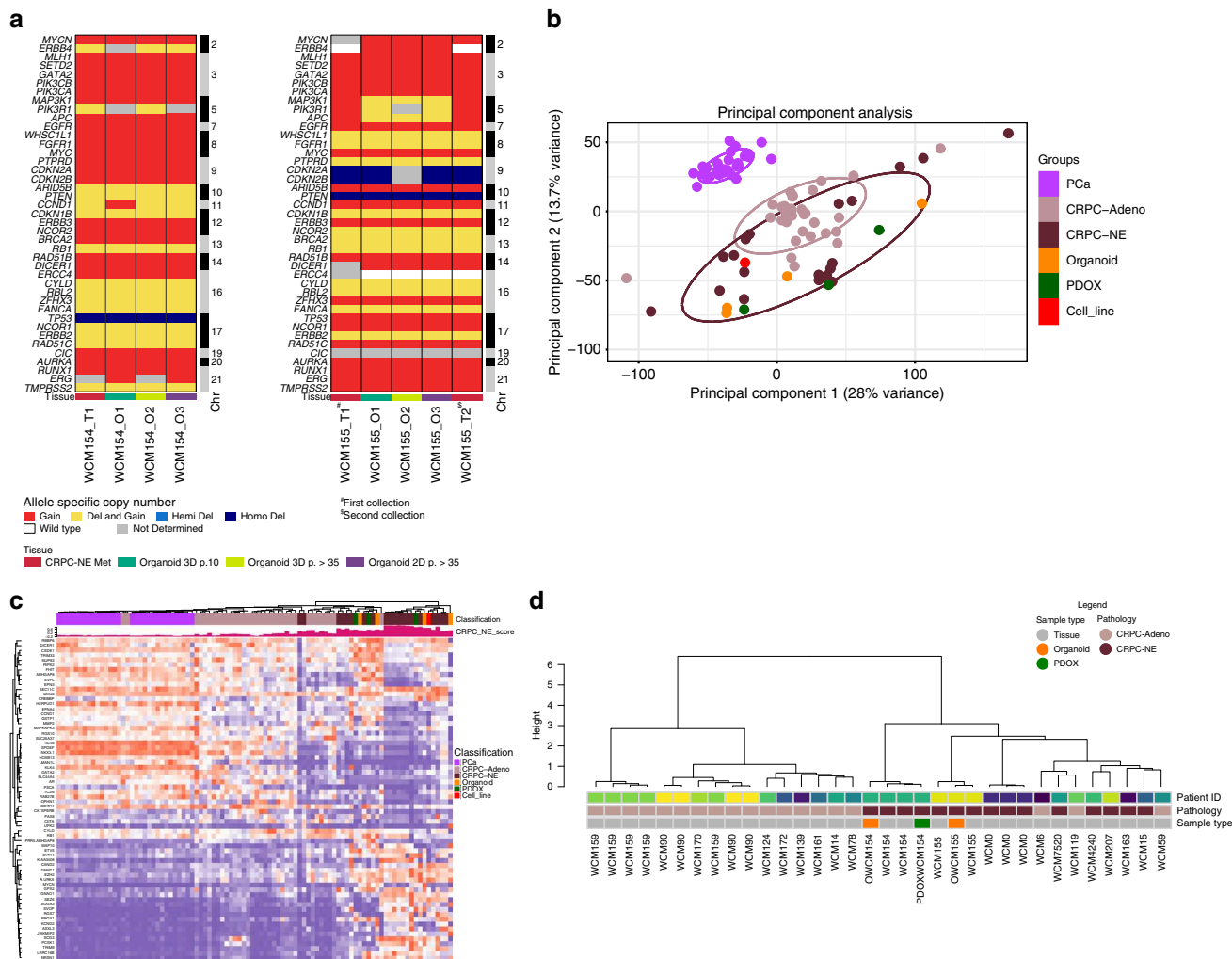


Fig. 2 Molecular characterization of patient-derived neuroendocrine prostate cancer models. **a** Allele-specific copy number of selected genes in organoids at different passages over time. **b** Transcriptomic analysis via principle component analysis including 26 PCA, 33 CRPC-Adeno, 13 CRPC-NE patient samples and CRPC-NE organoids (orange), CRPC-NE PDOX (green), NIC-H660 cell line (red) using RNAseq expression (FPKMS) of ~20k genes. **c** Clustering of study cohort samples using genes involved in CRPC-NE phenotype, based on RNAseq Expression (FPKMS). The cohort of localized prostate adenocarcinoma PCA (pink), CRPC-Adeno (mauve), and CRPC-NE (brown) patients (cohort from Beltran et al.⁶) CRPC-NE organoids (orange), CRPC-NE PDOX (green), NIC-H660 cell line (red). The green barplot on the top of the heatmap represents the CRPC-NE score (range from 0, low to 1, high) calculated according to methodologies described in Beltran et al.⁶ **d** Genome-wide DNA methylation cluster analysis using a cohort of CRPC-Adeno (mauve) and CRPC-NE (brown) patients⁶ together with models generated (organoids in orange and PDOX in green)

Discussion

While there have been significant advances in the treatment of patients with advanced prostate cancer, there is a wide variability in clinical responses to existing drugs. There are few preclinical models that recapitulate the clinical and molecular heterogeneity seen among patients thereby limiting the rational development of molecularly driven treatment strategies. Here we focus on the CRPC-NE phenotype, an emerging and aggressive subtype of advanced prostate cancer that can arise as an androgen-independent mechanism of resistance to AR-directed therapies, due to the lack of approved therapies for patients, limited preclinical models (only one cell line is available through ATCC), and a still preliminary understanding of CRPC-NE biology.

As with other rare cancers, there are few drugs or trials that have been developed for patients with CRPC-NE. Here we show that gaps in our knowledge concerning rare cancers may be addressed through the development of patient-derived preclinical models. Patient organoids retain the molecular features of their

corresponding patient over time and maintain similar responses to drugs in vitro.

Attempts to create prostate cancer organoids from biopsies have also been performed by other laboratories with similarly low overall success rates for indefinite propagation and expansion, perhaps due to the inability of cells to adapt very quickly from tissue to the culture conditions and therefore avoiding senescence. The development of models from biopsies also faces the challenge of scant starting material derived from needle biopsies of metastatic sites (especially bone), limiting the cell-cell interactions that are needed for cells to survive. Of the seven prostate cancer organoids described in Gao et al., one was from a patient with CRPC-NE derived from a malignant pleural effusion³¹. Other systems for organoid expansion have been established as using irradiated mouse cells and a Rho kinase (ROCK) inhibitor as supportive elements for human epithelial cells³⁸ or a matrigel/EGF-based culture system supplemented with androgens³⁹. It has been shown that higher cell density deactivates mTOR pathway and suppress the senescence program⁴⁰. The use of ROCK

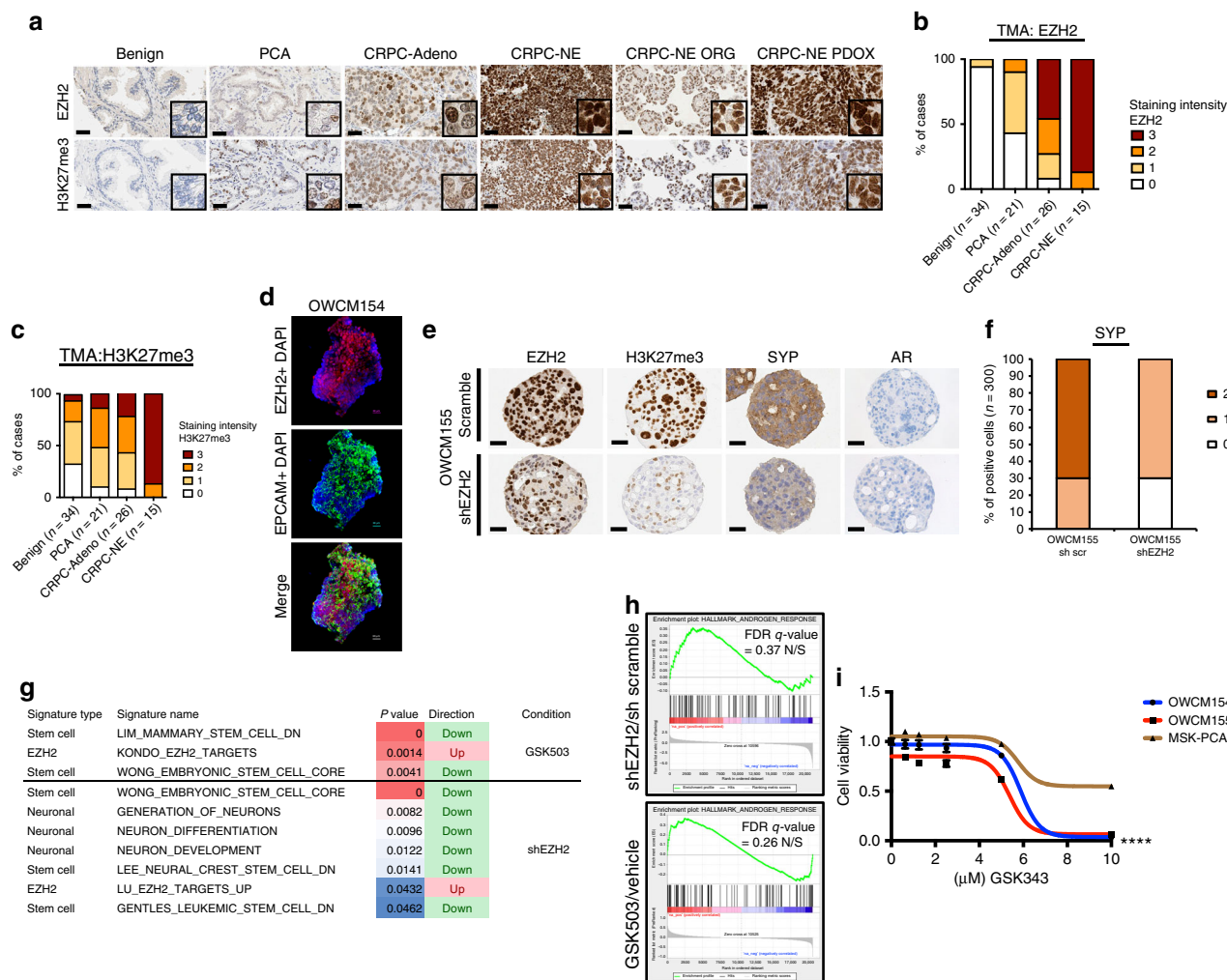


Fig. 3 Manipulation of EZH2 in CRPC-NE models affects neuroendocrine-associated programs and tumor cell viability. **a** Representative EZH2 IHC images of benign prostate, localized prostate cancer (PCA), CRPC-Adeno and CRPC-NE patient tissue, representative CRPC-NE organoids, and corresponding PDOX. Tissue is stained with EZH2 and H3K27me3 antibodies ($\times 20$ magnification, scale bar 100 μm , inset $\times 40$). **b** Bar plots scoring analysis of tissue staining intensity using EZH2 and H3K27me3 antibodies. The cases are represented as % and the number of cases is indicated in the figure. EZH2 and H3K27me3 staining intensity vary from 0 (no to minimal intensity) to 4 (high intensity). **c** Immunofluorescence staining of OWCM154 using EZH2 (Alexa Fluor[®] 555) and APC anti-human EPCAM and 4',6-diamidino-2-phenylindole, dihydrochloride (DAPI) (scale bar 30 μm). **d** Immunohistochemistry of OWCM155 organoids infected with sh scramble or shEZH2. Organoids are stained using EZH2, H3K27me3, synaptophysin (SYP), and AR antibodies. ($\times 40$ objective, scale bar 50 μm). **e** Bar plot scoring analysis of SYP staining intensity in shEZH2 versus sh scramble. Staining intensity is calculated from 0 (no to minimal intensity) to 2 (medium-high intensity). Three hundred cells have been evaluated for the scoring. **f** GSEA table of signatures (EZH2 targets, Neuronal and Stem Cell) that are significantly enriched in organoids treated with shEZH2 compared to scramble and in organoids treated with GSK503 compared to vehicle treatment. These signatures are ranked based on p value (from 0.05 to 0.001) and $\text{FDR} < 0.25$. **g** GSEA enrichment plot of the AR pathway genes in organoids infected with shEZH2 versus scramble or treated with EZH2 inhibitor (GSK503) or vehicle. **h** Cell viability assay (Cell-Title Glo) of OWCM154 (blue), OWCM155 (red), MSK-PCA3 (dark orange) after 11 days of treatment with vehicle or indicated doses of GSK343 ($n = 9$, for each treatment dose, error bars: s.e.m.). Two-way ANOVA test is used, **** $p < 0.0001$ (OWCM154 vs MSK-PCA3 and OWCM155 vs MSK-PCA3)

inhibitors while passaging organoids delays senescence and supports proliferation programs^{41,42} but tissue processing and media optimization are required to make this more suitable for low biopsy input cellular amount. Further, the optimal media conditions to support CRPC-NE vs CRPC-Adeno may be different and whether co-culturing techniques may improve success rates are yet to be elucidated.

Previously described patient-derived xenograft (PDX) models of neuroendocrine prostate cancer have been generated from larger quantities of surgical or autopsy material⁴³ including LuCaP 49⁴⁴, MDA PCA 144 PDX⁴⁵, LTL352 and LTL370⁴⁶, LTL545⁴⁷ and these represent complementary clinically relevant models to study CRPC-NE biology and therapeutic strategies. Adding the organoid development step from smaller input

material including needle biopsies could positively impact the ability to generate patient in vivo models.

We used organoids to assess the functional impact of genes involved in CRPC-NE pathogenesis and highlight the role of EZH2. EZH2 inhibition resulted in a downregulation of neuroendocrine pathway genes and those associated with stem cell and neuronal pathways; however, AR expression or activity did not increase suggesting a later disease state and possibly loss of plasticity and inability of these CRPC-NE organoids to revert back to a more luminal state.

Organoid drug screening generated hypotheses for single-agent and combination therapies confirming the usage of certain drugs in the clinic for CRPC-Adeno and nominating new approaches for CRPC-NE. There are currently no approved therapies for

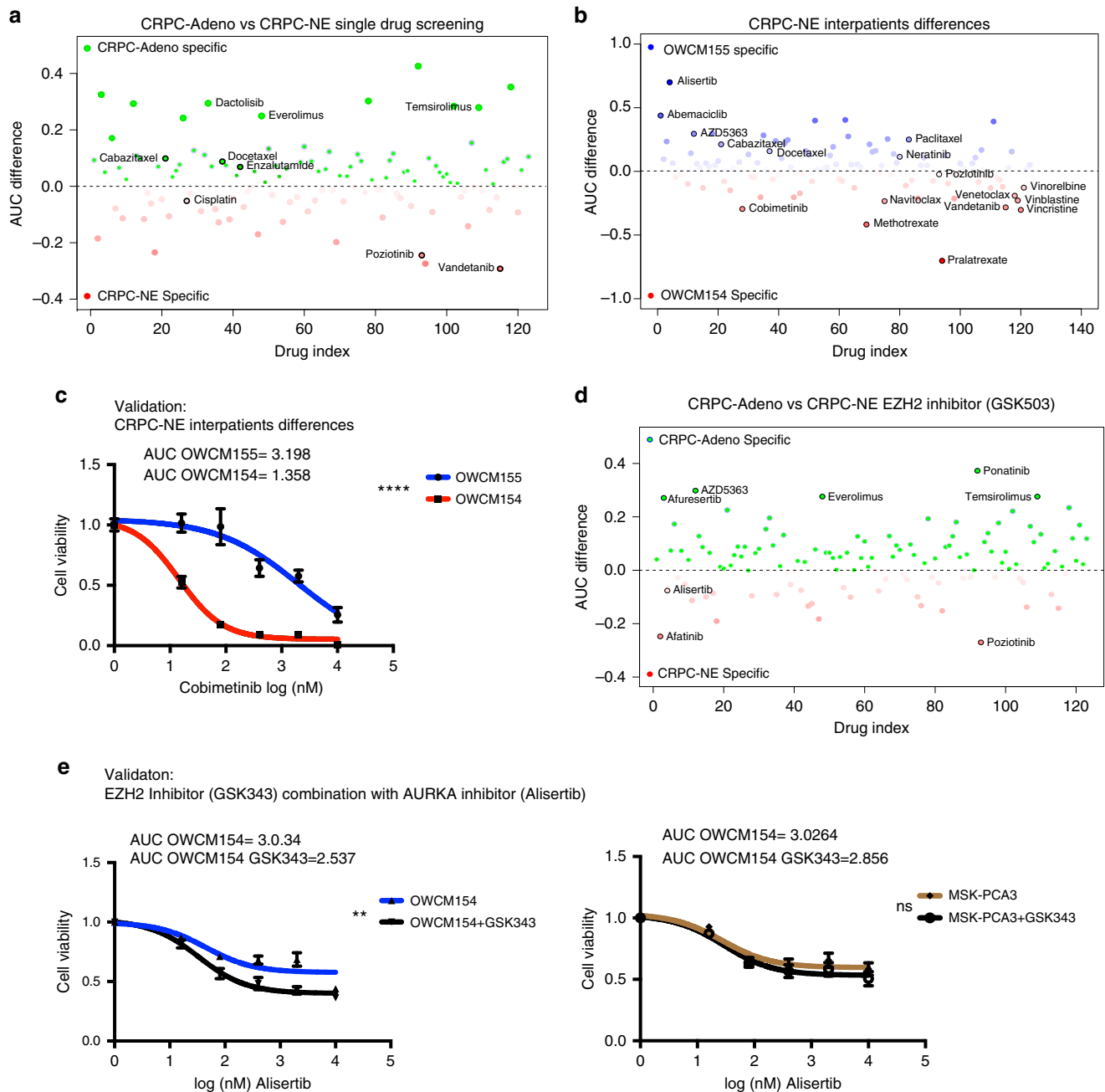


Fig. 4 High-throughput drug screening in organoids identifies novel single agents and combination therapies for CRPC-NE. **a** High-throughput drug screen in CRPC-NE organoids (OWCM154 OWCM155) vs control CRPC-Adeno organoids (MSK-PCA3 MSK-PCA2). The y-axis is the AUC (area under the curve) differential of the mean of the CRPC-NE samples - the mean of the CRPC-Adeno samples. Compounds indicated in red are specific for CRPC-NE while compounds indicated in blue are specific for CRPC-Adeno. Highlighted compounds are clinically relevant and represent a subclass of drugs. **b** High-throughput drug screen single agent analysis of differences in sensitivity within the CRPC-NE samples. **c** Cell viability assay using vehicle or different doses of cobimetinib in OWCM154 and OWCM155 ($n = 9$, for each treatment dose, error bars: s.e.m.), two-way ANOVA test is used. $****p < 0.0001$. **d** High-throughput drug combination screen in CRPC-NE organoids (OWCM154, OWCM155) vs CRPC-Adeno organoids (MSK-PCA3, MSK-PCA2). GSK503 has been added to the drug screening plate at IC30. The y-axis is the AUC differential of the mean of the CRPC-NE samples - the mean of the CRPC-Adeno samples. Compounds indicated in red are more effective for CRPC-NE in combination with GSK503 while compounds indicated in blue are more effective for CRPC-Adeno in combination with GSK503. **e** Cell viability assay for high-throughput drug combination screen validation. Organoids are treated with a fixed dose of GSK343 ($5 \mu\text{M}$) and escalated doses of alisertib. The organoids treated with alisertib in combination with GSK343 are represented with a black line while the organoids treated with alisertib plus DMSO are represented in blue ($n = 9$, for each treatment dose, error bars, s.e.m.). Two-way ANOVA test is used. Alisertib-GSK503 combination in OWCM154 has $p < 0.0027$, (**) while in MSK-PCA3 is not significant (ns)

CRPC-NE, representing a clinical unmet need. We previously reported preferential sensitivity of CRPC-NE to aurora kinase A inhibition, which led to a multicenter phase 2 trial of alisertib for CRPC-NE (NCT01482962). The organoid models OWCM155 and OWCM154 were developed from an exceptional responder and non-responder patient enrolled on the phase 2 clinical trial and demonstrated corresponding responses to alisertib *in vitro*. Drug screening also identified drugs and drug combinations concordant with the genomic background of the tumors, as was the case of *PTEN* loss that conferred response to AKT inhibition. Combination screens using an EZH2 inhibitor as a potential method for priming to other treatments identified novel combinations not yet tested in the clinic for CRPC-NE patients. For instance, EZH2 inhibition combined with the AURKA inhibitor alisertib, both tested as single agents in CRPC-NE, were identified possibly due to their cooperative role in driving N-myc activity in CRPC-NE²⁰. Although additional studies are needed to further understand the biologic implications of several of these findings, these data suggest that CRPC-NE organoids are clinically relevant models to unveil novel targets and therapies, and high-throughput drug screening is a useful tool to generate valid treatment hypotheses for CRPC-NE.

Methods

Cohort description and pathology classification. Fresh tumor biopsy specimens were obtained prospectively through a clinical trial approved by the Weill Cornell Medicine (WCM) Institutional Review Board (IRB) with informed consent (IRB #1305013903). Germline (normal) DNA was obtained from peripheral blood mononuclear cells. All hematoxylin and eosin-stained slides were reviewed by board-certified pathologists (J.M.M. and M.A.R.). Histologic criteria were from the proposed classification of prostate cancer with neuroendocrine differentiation⁹.

Tissue processing and organoid development. Fresh tissue biopsy samples were placed in media DMEM (Invitrogen) with GlutaMAX (1×, Invitrogen), 100 U/ml penicillin, 100 µg/ml streptomycin (Gibco), Primocin 100 µg/ml (InvivoGen), and 10 µmol/l ROCK inhibitor (Selleck Chemical Inc.). Tissue samples were washed in media two times before being placed in a 10 cm petri dish (Falcon) for mechanical dissection. The dissected tissue was then enzymatically digested with 250 U/ml of collagenase IV (Life Technologies) and TrypLE express (Gibco) in a ratio 1:2 with Collagenase IV in a 15 ml conical centrifuge tube (Falcon) incubated in a shaker at 37 °C set to 5 rcf. Incubation time of the specimen was dependent on the amount of collected tissue and ranged from 30 to 90 min, until the majority cell clusters were in suspension. After tissue digestion, DMEM media containing 10% FBS was added to the suspension to inactivate collagenase IV and the mixture was centrifuged at 326 rcf for 4 min. The pellet was then washed with Advanced DMEM (Invitrogen) containing GlutaMAX (1×, Invitrogen), 100 U/ml penicillin, 100 µg/ml streptomycin (Gibco), and HEPES (1 M, Gibco). The pellet was resuspended with prostate-specific culture media composed of Advanced DMEM (Invitrogen) with GlutaMAX (1×, Invitrogen), 100 U/ml penicillin, 100 µg/ml streptomycin (Gibco), Primocin 100 µg/ml (InvivoGen), B27 (Gibco), *N*-Acetylcysteine 1.25 mM (Sigma-Aldrich), Mouse Recombinant EGF 50 ng/ml (Invitrogen), Human Recombinant FGF-10 20 ng/ml (Peprotech), Recombinant Human FGF-basic 1 ng/ml (Peprotech), A-83-01 500 nM (Tocris), SB202190 10 µM (Sigma-Aldrich), Nicotinamide 10 mM (Sigma-Aldrich), (DiHydro) Testosterone 1 nM (Sigma-Aldrich), PGE2 1 µM (R&D Systems), Noggin conditioned media (5%) and R-spondin conditioned media (5%). The final resuspended pellet was combined with growth factor-reduced Matrigel (Corning) in a 1:2 volume ratio. Six droplets of 50 µl cell suspension/Matrigel mixture was pipetted onto each well of a six-well cell suspension culture plate (Sarstedt Ltd.). The plate was placed into a cell culture incubator at 37 °C and 5% CO₂ for 30 min to solidify the droplets before 3 ml of prostate-specific culture media was added to each well. The culture was replenished with fresh media every 3–4 days during organoid growth. Dense cultures with organoids ranging in size from 200 to 500 µm were passaged weekly. During passaging, the organoid droplets were mixed with TrypLE Express (Gibco) and placed in a water bath at 37 °C for a maximum of 5 min. The resulting cell clusters and single cells were washed and replated, following the protocol listed above. Prostate organoids were biobanked using Recovery Cell Culture Freezing Medium (Gibco) at –80 °C. Throughout prostate organoid development, cultures were screened for various *Mycoplasma* strains using the MycoAlert Kit (Lonza) and confirmed negative before being used for experimental assays. The MSK-PCA2 and MSK-PCA3 used for this study as CRPC-Adeno controls were developed and described by Gao et al.³¹.

Patient-derived organoid xenograft development. 1.5 million cells derived from organoids were injected with Matrigel (Corning) 1:1 subcutaneously into NOD scid gamma (NOD.Cg-Prkdc^{scid} Il2rg^{tm1Wjl}/SzJ) male mice (Jackson Laboratories, Bar Harbor, Maine). Mice used for xenografts were 6-to 8-weeks old. Daily light cycles were kept consistent in the animal facility (12 h light and 12 h dark). Cages were changed fully once a week. Tumor volume was measured every week with a caliper. The animals were sacrificed in a CO₂ chamber after 2–4 months of tumor growth. The harvested tumors were partly used for histology, genomic and transcriptomic analysis and partly reimplanted into NOD scid gamma mice. Animal care and experiments were carried out in accordance with IACUC guidelines.

Immunoblot and immunohistochemistry and immunofluorescence. Organoids were lysed in RIPA buffer supplemented with protease inhibitor cocktail and phosphatase inhibitors (Thermo Scientific). In the case of H3K27me3 detection sonication was performed (High, 30" on and 30" off for 5 cycles). The total protein concentration of the soluble extract was determined using the BCA protein assay Kit (Thermo Scientific). Each protein sample (50 µg) was resolved to SDS-PAGE, transferred onto a PVDF membrane (Millipore) and incubated overnight at 4 °C with primary antibodies. Primary antibodies used: Androgen Receptor (1:2000, Abcam, [ER179(2)] ab108341), PTEN (D4.3) XP (1:1000, Cell Signaling 9188S), Actin (1:2000, EMD Millipore clone C4, MAB1501), H3K27me3 (1:1000, Cell Signaling Technology, 9733S), EZH2 (D2C9) XP[®] Rabbit mAb (1:1000, Cell Signaling 5246), Phospho-Rb (Ser780) (1:1000, Cell Signaling, 9307), CDKN2A (1:2000, Abcam, ab108349), Synaptophysin (1:5000, Abcam [YE269], ab32127). Following three washes with TBS-T, the blot was incubated with horseradish peroxidase-conjugated secondary antibody and immune complexes were visualized by enhanced chemiluminescence detection (ECL plus kit, Pierce).

Immunohistochemistry was performed on deparaffinized FFPE sections (organoid, xenograft or patient tissue) using a Bond III automated immunostainer (Leica Microsystems, IL, USA). Heat-mediated antigen retrieval was performed using the Bond Epitope Retrieval solution 1 (ER1) at pH6 or 2 (ER2) at pH9. The following antibodies and conditions were used: EZH2 (clone 11/EZH2, BD Biosciences, CA, USA; ER1, 1:20 dilution), H3K27me3 (C36B11, Cell Signaling, MA, USA; ER1, 1:200 dilution), Synaptophysin (SP11, Thermo Scientific; ER2, 1:100 dilution), Chromogranin A (LK2H10, BioGenex, CA, USA; no antigen retrieval, 1:400 dilution), AR (F39.4.1, BioGenex, CA, USA; ER1, 1:800 dilution with casein), Ki67 (MIB-1, Dako, CA, USA; ER1, 1:50 dilution), PTEN (Cell Signaling, 9559, ER2, 1:100).

Scoring of EZH2 and H3K27me3 was performed on tissue microarrays (85 cases) and whole slides (11 cases). Nuclear staining intensity in tumor tissue was evaluated blindly by a pathologist using a four-tiered scoring system: negative (or present in <5% of tumor nuclei), weak, moderate or strong. If a case showed heterogeneous staining, the intensity score representative of the majority of tumor nuclei of that case was assigned.

Scoring of Synaptophysin was performed blindly by a pathologist analyzing 300 cells on slides sh scramble vs shEZH2 and applying the following scoring system: 0 negative staining, 1 weak, 2 mild staining.

Immunofluorescence was performed on OWCM154 and OWCM155 using the following antibodies EZH2 (D2C9) XP[®] Rabbit mAb (1:250, Cell Signaling 5246), secondary antibody Alexa Fluor[®] 555 (1:1000, ThermoFisher Scientific A27039) and APC anti-human EPCAM (1:250, Biolegend 324208) DAPI. Briefly, organoids were washed with PBS and Paraformaldehyde (PFA 4% in PBS) was added overnight at 4 °C. The following day organoids were incubated with a blocking solution containing 1% Triton[™]-X and 1% FBS in PBS for 60 min at room temperature. Primary and secondary antibodies were added in PBS solution containing 0.5% Triton[™]-X and 0.1% FBS for 1 h respectively at indicated concentration at room temperature⁴⁸. Z-stacks are obtained using a Zeiss confocal microscope (LSM510; Carl Zeiss Microscopy). 3D images are obtained combining the Z-stacks using Imaris software.

Cytology smear. Organoids at early passage were morphologically screened for contamination of benign epithelial cells and fibroblasts. Organoids cells were collected from the Matrigel droplet using an inverted microscope and placed on Super Frost Plus glass slide (VWR MicroSlides # 48311-703). A second glass slide was used to spread the organoids on the entire surface and after air-drying stained with Diff-Quik stain (Siemens Medical Solution USA, INC, Mavren Pa). The stained organoids were reviewed by the study pathologists.

DNA extraction and exome sequencing. DNA extractions from patient tumors, organoids, and PDOXs were performed using DNeasy Blood and Tissue Kit (QIAGEN) and Maxwell 16 Tissue DNA Purification Kit (Promega). Whole-exome capture libraries were constructed after sample-shearing, end repair, and phosphorylation and ligation to barcoded sequencing adaptors. Ligated DNA was size-selected for lengths between 200 and 350 bp and subjected to HaloPlex Exome (Agilent). Sequencing was performed using Illumina HiSeq 2500 (2 × 100 bp). Reads were aligned to GRC37/hg19 reference using Burrows-Wheeler Aligner and processed according to the IPM-Exome-pipeline v0.9.

Copy-number analysis. Concordance between tumor tissues, tumor organoid models, and matching xenografts was assessed using SPIA⁴⁹ genotype distance test. CLONET¹¹ was used to quantify tumor purity and ploidy from WES segmented data and allelic fraction (AF) of germline heterozygous SNP loci. A pair (cnA, cnB) of integer values, representing allele-specific copy number, was assigned to each genomic segment identified by the IPM-Exome-pipeline, as described in Beltran et al.⁶ Quality filters required at least ten informative SNPs and mean coverage of 20 to call allele-specific values of a segment. Post-sequencing manual review of allele-specific calls was performed.

Concordance between two tumor samples was assessed by comparing discretized allele-specific copy number values into five levels (Fig. 2a): homozygous deletion (cnA = 0, cnB = 0), hemizygous deletion (cnA = 1, cnB = 0), wild type (cnA = 1, cnB = 1), gain (cnA ≥ 2, cnB ≥ 1), and reciprocal loss of heterozygosity (cnA > 1, cnB = 0). Reciprocal loss of heterozygosity event captures complex copy number states where one allele is lost, and the other one is gained. Reciprocal loss of heterozygosity was conserved in tumor organoids models and matching xenografts.

RNA extraction sequencing and analysis. mRNA was extracted from organoids and PDXs using RNAsy Mini Kit (QIAGEN) and Maxwell 16 LEV simplyRNA Tissue Kit. Specimens were prepared for RNA sequencing using TruSeq RNA Library Preparation Kit v2 as previously described⁶. RNA integrity was verified using the Agilent Bioanalyzer 2100 (Agilent Technologies). cDNA was synthesized from total RNA using Superscript III (Invitrogen). Each sample was then sequenced with the HiSeq 2500 to generate 2 × 75-bp paired-end reads. All reads were independently aligned with STAR_2.4.0f17 for sequence alignment against the human genome build hg19, downloaded via the UCSCgenomebrowser (<http://hgdownload.soe.ucsc.edu/goldenPath/hg19/bigZips/>), and SAMTOOLS v0.1.19.8 for sorting and indexing reads. Cufflinks (2.0.2)9 was used to get the expression values (FPKMS), and Gencode v19 10 GTF file for annotation. Since the sequenced samples from the published data were processed using different library preps, batch normalization was done using ComBat11 from sva bioconductor package 12. The gene counts from htseq-count13 and DESeq2 Bioconductor package14 were used to identify differentially expressed genes. The hypergeometric test and Gene Set Enrichment Analysis (GSEA)v15 was used to identify enriched signatures using the different pathways collection in the MSigDB database 16. We used GSEA pre-ranked method from GSEA for our purpose. Principal Component Analysis (PCA) was performed using the prcomp function of R “stats” package (<https://cran.r-project.org/>), and visualization was done using ggbiplot package (<https://github.com/vqv/ggbiplot>). A Wald test was applied for mRNA differential analysis, followed by Benjamini–Hochberg correction for multiple hypothesis testing.

AR signaling and integrated CRPC-NE score. For each sample, AR signaling was assessed based on the expression levels of 30 genes⁶. The Integrated Neuroendocrine Prostate Cancer (CRPC-NE) score estimates the likelihood of a test sample to be CRPC-NE and it is computed based on a set of 70 genes⁶. The gene set stems from the integration of differentially deleted/amplified and/or expressed and/or methylated genes in CRPC-NE vs CRPC.

Methylation profiling. Sample preparation, alignment, and enhanced reduced representation bisulfite sequencing (eRRBS) were performed at the WCM Epigenomics Core Facility.⁵⁰ Samples profiled by eRRBS included 19 CRPC-Adeno, 15 CRPC-NE, 2 Organoids, and 1 PDX samples. Only sites covered by at least ten reads were considered for downstream analysis. For each sample, the percentage of methylation per site (beta value) was computed. Ward’s hierarchical clustering of samples was performed by “1-Pearson’s correlation coefficient” as distance measure on the 5% CpG sites showing the highest standard deviation across the cohort.

Cell line infection and drug treatments. NCI-H660 used in this study were purchased from ATCC and maintained according to the manufacturers’ protocols. Cell authentication was performed using STR analysis and cells were routinely tested for *Mycoplasma* contamination and resulted negative. shEZH2 used in these studies was kindly provided by Dr. Beguelin and Dr. Melnick (WCM) with the following sequence: TATGATGGTTAACGGTGA. shEZH2 and sh scramble were used to infect CRPC-NE organoids. In brief, organoid cells were collected and resuspended with infection media containing Y27632 (Selleck Chemical) and Polybrene (Millipore). Organoids cells were then placed in 24 well-plates and centrifuged at 600g at 32 °C for 60min. After centrifugation organoids were incubated at 37 °C overnight and the following day seeded in Matrigel droplets⁵¹. We used pLKO.1-puro vector and infected cells were selected by puromycin treatment (1 µg/ml).

Cell viability assays were performed on 4000 organoid cells treated with increasing doses of GSK343 (Sigma-Aldrich SML0766) and GSK503 (GSK provider) at the indicated concentrations for 6 or 11 days and Neratinib, Alisertib, Afuresertib, Cobimetinib were purchased from SelleckChem and used in cell viability assays for 6 days. Viability was measured with cell viability assay kit according to the manufacturer’s protocol (CellTiter-Glo, Promega). For RNA extraction (Qiagen Kit) and protein lysate, treatments were conducted for 6 days. For viability assays all the data are expressed as mean ± standard error of the mean

(SEM). Multiple sample comparisons were calculated using ANOVA (in GraphPad Prism 6). Differences between values were considered statistically significant at a p value of less than 0.05.

High-throughput drug assay. For high-throughput drug screens, cells were dispensed into 384-well tissue-culture-treated plates at ~30% confluence (500–1200 cells) using a BioTek MultiFlo™. After 24 h, using robotic liquid handling, cells were exposed to 126 unique drugs. Drugs were diluted to a 6-point dose curve incorporating a 3 or 4-fold dilution step in the presence and absence of an IC30 concentration of GSK503. After 6 additional days of incubation, cell viability was assessed using CellTiter-Glo (Promega) and a BioTek Synergy H4 plate reader. All screening plates were subjected to stringent quality control measures, including the Z factor calculation. Raw luminescence units (RLU) were then normalized on a per plate basis to the median values of the negative control: DMSO or PBS, depending on the drug solvent. Dose–response curves were then fit to a 4-parameter logistic model using the R “nplr” package version: 0.1–7. Area Under the Curve (AUC), IC50, and Goodness of Fit (GOF) were calculated for each drug.

AUC values were then compared with the SEngine Precision Medicine internal database of a total of 47 primary tumor samples across multiple tumor types, generating an AUC Z-score that we integrated for the prioritization of future drug investigation. The tumor types included prostate, ovarian, breast, gliomatosis cerebri, myxofibrosarcoma, head and neck, thyroid, liver, CML, endometrial, glioblastoma, colorectal, lung, cholangiosarcoma, uterine carcinosarcoma, and neuroblastoma. This method of statistical analysis allows for the detection of unique sensitivities across multiple samples. For the drug combinations study, the top drug combinations were selected through multiple criteria: AUC fold change, AUC differential, AUC combination Z-score, drug target, novelty, and clinical status of drugs.

Data availability. The RNA-seq and ERRBS data generated during the current study are available through Gene Expression Omnibus (GEO) accession number: GSE112830 with the following sub-series: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE112786>, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE112829>. The whole exome sequencing data related to this study are available through Sequence Read Archive (SRA) with accession number SRP138000. The published human data are available through dbGap:phs000909.v.p1 (http://www.cbiportal.org/study?id=nepc_wcm_2016)⁶.

Received: 25 October 2017 Accepted: 25 April 2018

Published online: 19 June 2018

References

1. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2018. *CA Cancer J. Clin.* **68**, 7–30 (2018).
2. Fizazi, K. et al. Abiraterone plus prednisone in metastatic, castration-sensitive prostate cancer. *N. Engl. J. Med.* **377**, 352–360 (2017).
3. James, N. D. et al. Abiraterone for prostate cancer not previously treated with hormone therapy. *N. Engl. J. Med.* **377**, 338–351 (2017).
4. Bluemn, E. G. et al. Androgen receptor pathway-independent prostate cancer is sustained through FGF signaling. *Cancer Cell* **32**, 474–489.e476 (2017).
5. Wang, H. T. et al. Neuroendocrine prostate cancer (NEPC) progressing from conventional prostatic adenocarcinoma: factors associated with time to development of NEPC and survival from NEPC diagnosis—a systematic review and pooled analysis. *J. Clin. Oncol.* **32**, 3383–3390 (2014).
6. Beltran, H. et al. Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. *Nat. Med.* **22**, 298–305 (2016).
7. Mertz, K. D. et al. Molecular characterization of TMPRSS2-ERG gene fusion in the NCI-H660 prostate cancer cell line: a new perspective for an old model. *Neoplasia* **9**, 200–206 (2007).
8. Pauli, C. et al. An emerging role for cytopathology in precision oncology. *Cancer Cytopathol.* **124**, 167–173 (2016).
9. Epstein, J. I. et al. Proposed morphologic classification of prostate cancer with neuroendocrine differentiation. *Am. J. Surg. Pathol.* **38**, 756–767 (2014).
10. Mosquera, J. M. et al. Concurrent AURKA and MYCN gene amplifications are harbingers of lethal treatment-related neuroendocrine prostate cancer. *Neoplasia* **15**, 1–10 (2013).
11. Prandi, D. et al. Unraveling the clonal hierarchy of somatic genomic aberrations. *Genome Biol.* **15**, 439 (2014).
12. Robinson, D. et al. Integrative clinical genomics of advanced prostate cancer. *Cell* **161**, 1215–1228 (2015).
13. Dardenne, E. et al. N-Myc induces an EZH2-mediated transcriptional program driving neuroendocrine prostate cancer. *Cancer Cell* **30**, 563–577 (2016).
14. Akamatsu, S. et al. The placental gene PEG10 promotes progression of neuroendocrine prostate cancer. *Cell Rep.* **12**, 922–936 (2015).

15. Li, Y. et al. SRRM4 drives neuroendocrine transdifferentiation of prostate adenocarcinoma under androgen receptor pathway inhibition. *Eur. Urol.* **71**, 68–78 (2017).
16. Mu, P. et al. SOX2 promotes lineage plasticity and antiandrogen resistance in TP53- and RB1-deficient prostate cancer. *Science* **355**, 84–88 (2017).
17. Bishop, J. L. et al. The master neural transcription factor BRN2 is an androgen receptor-suppressed driver of neuroendocrine differentiation in prostate cancer. *Cancer Discov.* **7**, 54–71 (2017).
18. Hieronymus, H. et al. Gene expression signature-based chemical genomic prediction identifies a novel class of HSP90 pathway modulators. *Cancer Cell* **10**, 321–330 (2006).
19. Ertel, A. et al. RB-pathway disruption in breast cancer: differential association with disease subtypes, disease-specific prognosis and therapeutic response. *Cell Cycle (Georget., Tex.)* **9**, 4153–4163 (2010).
20. Chatterjee, S. J. et al. Hyperphosphorylation of pRb: a mechanism for RB tumour suppressor pathway inactivation in bladder cancer. *J. Pathol.* **203**, 762–770 (2004).
21. Yang, G., Rajadurai, A. & Tsao, H. Recurrent patterns of dual RB and p53 pathway inactivation in melanoma. *J. Invest. Dermatol.* **125**, 1242–1251 (2005).
22. Beguelin, W. et al. EZH2 is required for germinal center formation and somatic EZH2 mutations promote lymphoid transformation. *Cancer Cell* **23**, 677–692 (2013).
23. Zhang, H. et al. Oncogenic deregulation of EZH2 as an opportunity for targeted therapy in lung cancer. *Cancer Discov.* **6**, 1006–1021 (2016).
24. Suva, M. L. et al. EZH2 is essential for glioblastoma cancer stem cell maintenance. *Cancer Res.* **69**, 9211–9218 (2009).
25. Yan, K. S. et al. EZH2 in cancer progression and potential application in cancer therapy: a friend or foe? *Int. J. Mol. Sci.* **18**, 1172 (2017).
26. Ku, S. Y. et al. Rb1 and Trp53 cooperate to suppress prostate cancer lineage plasticity, metastasis, and antiandrogen resistance. *Science* **355**, 78–83 (2017).
27. Clermont, P. L. et al. Polycomb-mediated silencing in neuroendocrine prostate cancer. *Clin. Epigenetics* **7**, 40 (2015).
28. Kleb, B. et al. Differentially methylated genes and androgen receptor re-expression in small cell prostate carcinomas. *Epigenetics* **11**, 184–193 (2016).
29. Bracken, A. P. et al. EZH2 is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer. *EMBO J.* **22**, 5323–5335 (2003).
30. van Vlerken, L. E. et al. EZH2 is required for breast and pancreatic cancer stem cell maintenance and can be used as a functional cancer stem cell reporter. *Stem Cells Transl. Med.* **2**, 43–52 (2013).
31. Gao, D. et al. Organoid cultures derived from patients with advanced prostate cancer. *Cell* **159**, 176–187 (2014).
32. Kirk, J. S. et al. Top2a identifies and provides epigenetic rationale for novel combination therapeutic strategies for aggressive prostate cancer. *Oncotarget* **6**, 3136–3146 (2015).
33. Fillmore, C. M. et al. EZH2 inhibition sensitizes BRG1 and EGFR mutant lung tumours to TopoII inhibitors. *Nature* **520**, 239–242 (2015).
34. Pauli, C. et al. Personalized in vitro and in vivo cancer models to guide precision medicine. *Cancer Discov.* **7**, 462–477 (2017).
35. Tsao, C. K., Cutting, E., Martin, J. & Oh, W. K. The role of cabazitaxel in the treatment of metastatic castration-resistant prostate cancer. *Ther. Adv. Urol.* **6**, 97–104 (2014).
36. Tran, C. et al. Development of a second-generation antiandrogen for treatment of advanced prostate cancer. *Science* **324**, 787–790 (2009).
37. Beltran, H. et al. A phase 2 study of the aurora kinase A inhibitor alisertib for patients with neuroendocrine prostate cancer (NEPC). *Ann. Oncol.* **27**, 1–36 (2016).
38. Liu, X. et al. Conditional reprogramming and long-term expansion of normal and tumor cells from human biopsies. *Nat. Protoc.* **12**, 439–451 (2017).
39. Chua, C. W. et al. Single luminal epithelial progenitors can generate prostate organoids in culture. *Nat. Cell Biol.* **16**, 951–961 (2014).
40. Leontieva, O. V., Demidenko, Z. N., & Blagosklonny, M. V.. Contact inhibition and high cell density deactivate the mammalian target of rapamycin pathway, thus suppressing the senescence program. *Proc. Natl. Acad. Sci. USA* **111**, 8832–8837 (2014).
41. Drost, J. et al. Organoid culture systems for prostate epithelial and cancer tissue. *Nat. Protoc.* **11**, 347–358 (2016).
42. Karthaus, W. R. et al. Identification of multipotent luminal progenitor cells in human prostate organoid cultures. *Cell* **159**, 163–175 (2014).
43. Lawrence, M. G. et al. A preclinical xenograft model of prostate cancer using human tumors. *Nat. Protoc.* **8**, 836–848 (2013).
44. True, L. D. et al. A neuroendocrine/small cell prostate carcinoma xenograft-LuCaP 49. *Am. J. Pathol.* **161**, 705–715 (2002).
45. Aparicio, A. et al. Neuroendocrine prostate cancer xenografts with large-cell and small-cell features derived from a single patient's tumor: morphological, immunohistochemical, and gene expression profiles. *Prostate* **71**, 846–856 (2011).
46. Lin, D. et al. High fidelity patient-derived xenografts for accelerating prostate cancer discovery and drug development. *Cancer Res.* **74**, 1272–1283 (2014).
47. Beltran, H. et al. Whole-exome sequencing of metastatic cancer and biomarkers of treatment response. *JAMA Oncol.* **1**, 466–474 (2015).
48. Sato, T. et al. Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. *Nature* **459**, 262–265 (2009).
49. Demichelis, F. et al. SNP panel identification assay (SPIA): a genetic-based assay for the identification of cell lines. *Nucleic Acids Res.* **36**, 2446–2456 (2008).
50. Lin, P. C. et al. Epigenomic alterations in localized and advanced prostate cancer. *Neoplasia* **15**, 373–383 (2013).
51. Koo, B. K. et al. Controlled gene expression in primary Lgr5 organoid cultures. *Nat. Methods* **9**, 81–83 (2011).

Acknowledgements

The authors thank Drs. Ari Melnick and Wendy Beguelin at Weill Cornell Medicine for sharing short-harpin EZH2 plasmids. We acknowledge assistance from the Molecular Cytology Core Facility at Memorial Sloan-Kettering Cancer Center (NIH Core Grant P30 CA 008748), the Translational Research Program at WCMC Pathology and Laboratory Medicine, and the WCM CLC Genomics and Epigenomics Core Facility. This study was supported by the Prostate Cancer Foundation (L.P., H.B.), American Italian Cancer Foundation (L.P.), European Research Council ERC-CoG648670 (F.D.), Damon Runyon Cancer Research Foundation CI-67-13 (H.B.), Starr Cancer Consortium 17-A771 (Y.C., M.A.R., H.B.), Department of Defense PCRP PC121341 (H.B.), PC160264 (H.B.), PC131961 (D.M.N., S.T.T., H.B.), and the National Cancer Institute 1U54CA210184-01 (H.B.) and WCM Prostate Cancer SPORE 1 P50 CA211024-01A1 (J.M.M., O.E., H.B.).

Author contributions

L.P. and H.B. designed the study. L.P., M.K., D.G., W.R.K., Y.C., C.L.S., C.P., M.A.R., and H.B. developed the organoid methodology. J.C., V.S., A.V., R.A.R., J.M.M., and M.A.R. performed the histopathological analysis. D.P., M.B. and F.D. performed genomic and epigenomic analyses. R.B. and A.S. analyzed the RNA-sequencing experiments. R.S., R.R. and C.G. performed high-throughput drug screening. L.P., J.H., A.D., and M.S. performed drug screening validations, nanostring analysis and organoid characterization by protein expression. S.T.T., D.M.N. and H.B. provided patient biopsies and patient clinical information. L.D.D., G.I. and L.P. established patient-derived xenografts. O.E., Y.C., T.J.M., and C.C. provided administrative, technical, or material support. L.P. and H.B. wrote the initial draft of the manuscript and all authors contributed to the final version.


Additional information

Supplementary Information accompanies this paper at <https://doi.org/10.1038/s41467-018-04495-z>.

Competing interests: R.S. and R.R. are employees of SEngine Precision Medicine. C.G. has ownership interest (including patents) in SEngine Precision Medicine. The remaining authors declare no competing interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018

Bone Biopsy Protocol for Advanced Prostate Cancer in the Era of Precision Medicine

Verena Sailer, MD^{1,2}; Marc H. Schiffman, MD³; Myriam Kossai, MD^{1,2}; Joanna Cyrta, MD^{1,2}; Shaham Beg, MD^{1,2}; Brian Sullivan, MD³; Bradley B. Pua, MD³; Kyungmouk Steve Lee, MD³; Adam D. Talenfeld, MD³; David M. Nanus, MD^{4,5}; Scott T. Tagawa, MD^{4,5}; Brian D. Robinson, MD^{1,2}; Rema A. Rao, MD^{1,2}; Chantal Pauli, MD^{1,2}; Rohan Bareja, MS^{2,6}; Luis S. Beltran, MD⁷; Alexandros Sigaras, MS²; Kenneth Wa Eng, MS^{2,4,8}; Olivier Elemento, PhD^{1,2,6,8}; Andrea Sboner, PhD^{1,2,6}; Mark A. Rubin, MD^{1,2,4}; Himisha Beltran, MD^{2,4,5}; and Juan Miguel Mosquera, MD, MSc^{1,2}

BACKGROUND: Metastatic biopsies are increasingly being performed in patients with advanced prostate cancer to search for actionable targets and/or to identify emerging resistance mechanisms. Due to a predominance of bone metastases and their sclerotic nature, obtaining sufficient tissue for clinical and genomic studies is challenging. **METHODS:** Patients with prostate cancer bone metastases were enrolled between February 2013 and March 2017 on an institutional review board-approved protocol for prospective image-guided bone biopsy. Bone biopsies and blood clots were collected fresh. Compact bone was subjected to formalin with a decalcifying agent for diagnosis; bone marrow and blood clots were frozen in optimum cutting temperature formulation for next-generation sequencing. Frozen slides were cut from optimum cutting temperature cryomolds and evaluated for tumor histology and purity. Tissue was macrodissected for DNA and RNA extraction, and whole-exome sequencing and RNA sequencing were performed. **RESULTS:** Seventy bone biopsies from 64 patients were performed. Diagnostic material confirming prostate cancer was successful in 60 of 70 cases (85.7%). The median DNA/RNA yield was 25.5 ng/ μ L and 16.2 ng/ μ L, respectively. Whole-exome sequencing was performed successfully in 49 of 60 cases (81.7%), with additional RNA sequencing performed in 20 of 60 cases (33.3%). Recurrent alterations were as expected, including those involving the *AR*, *PTEN*, *TP53*, *BRCA2*, and *SPOP* genes. **CONCLUSIONS:** This prostate cancer bone biopsy protocol ensures a valuable source for high-quality DNA and RNA for tumor sequencing and may be used to detect actionable alterations and resistance mechanisms in patients with bone metastases. *Cancer* 2018;124:1008-15. © 2017 American Cancer Society.

KEYWORDS: biopsy, bone metastases, DNA, next-generation sequencing (NGS), precision medicine, prostate cancer, RNA.

INTRODUCTION

Approximately 90% of patients with metastatic prostate cancer harbor bone metastases, and up to 42.9% of patients have bone-only metastases.^{1,2} With an estimated 26,730 patients expected to die of their disease in 2017 in the United States alone,³ prostate cancer and associated bone metastases represent a large health burden worldwide.

The therapeutic portfolio for patients with metastatic prostate cancer is evolving based on an improved understanding of the molecular framework of metastatic prostate cancer and the identification of potentially actionable targets enriched in advanced disease (eg, homologous recombination genomic alterations and poly [ADP-ribose] polymerase [PARP] inhibitors, mismatch repair/microsatellite instability, and immunotherapy).^{4,5} Biomarker and genomic studies in patients with prostate cancer have been based predominantly on metastatic biopsy protocols embedded within clinical trials and/or developed through consortium efforts at specialized centers, including the Stand Up To Cancer-Prostate Cancer Foundation Dream Teams.⁴⁻⁶

Sclerotic bone lesions have been notoriously challenging to biopsy because the amount of tissue obtained is scant and requires decalcification, and often is inadequate for extensive molecular analyses.^{7,8} Using an unguided approach, a

Corresponding authors: Juan Miguel Mosquera, MD, MSc, Department of Pathology and Laboratory Medicine, Weill Cornell Medicine, 1300 York Ave, Box 69, New York, NY, 10065; jimm9018@med.cornell.edu and Himisha Beltran, MD, Division of Medical Oncology, Weill Cornell Medicine, 413 East 69th St, 14th Fl, New York, NY 10021; hip9004@med.cornell.edu

¹Department of Pathology and Laboratory Medicine, Weill Cornell Medicine, New York, New York; ²Caryl and Israel Englander Institute for Precision Medicine, Weill Cornell Medicine and New York Presbyterian, New York, New York; ³Department of Radiology, Weill Cornell Medicine, New York, New York; ⁴Sandra and Edward Meyer Cancer Center, Weill Cornell Medicine and New York Presbyterian, New York, New York; ⁵Division of Hematology and Medical Oncology, Department of Medicine, Weill Cornell Medicine, New York, New York; ⁶Institute for Computational Biomedicine, Weill Cornell Medicine, New York, New York; ⁷Department of Radiology, NYU Langone Medical Center, New York, New York; ⁸Department of Physiology and Biophysics, Weill Cornell Medicine, New York, New York

The last 2 authors share senior authorship.

The Englander Institute for Precision Medicine clinical team coordinated clinical activities. Rob Kim provided administrative support. Logistic support was provided by Elyze C. Merzier. Art design was provided by Rogerio Paulo da Silva.

Additional supporting information may be found in the online version of this article.

DOI: 10.1002/cncr.31173, **Received:** September 27, 2017; **Revised:** November 7, 2017; **Accepted:** November 10, 2017, **Published online** December 19, 2017 in Wiley Online Library (wileyonlinelibrary.com)

previous study demonstrated that only 25.5% of 184 patients with metastatic prostate cancer had a positive bone biopsy.⁹ Although success rates for image-guided bone biopsies in specialized academic centers have significantly improved in recent years, the broader application of bone biopsy protocols across the clinical community is challenged by a lack of standardized protocols.

The current study provides a detailed description of a successful next-generation biopsy protocol amenable to sclerotic bone lesions of patients with advanced prostate cancer and broadly applicable to the clinical community. We combined clinical and radiographic features, image-guided percutaneous techniques, and standardized pathology processing protocols to optimize the yield and quality of tissues obtained for downstream testing.

MATERIALS AND METHODS

Patient Cohort

Patients with metastatic prostate cancer were prospectively enrolled on an ongoing precision medicine clinical trial at Weill Cornell Medicine from February 2013 until March 2017 and underwent image-guided bone biopsy by interventional radiology.¹⁰ Institutional review board-approved written consent was obtained before metastatic biopsy, genomic testing, and clinical follow-up. All lesions were reviewed prospectively by one interventional radiologist (M.H.S.) and were defined as sclerotic, mixed, or lytic via qualitative comparison of attenuation with adjacent areas of intact cortical bone. A schematic figure of the workflow and specimen images can be found in Figure 1 and Figure 2 (see Supporting Information Table 1).

Computed Tomography-Guided Bone Biopsy

To identify the shortest path to the bone lesion without intervening vital structures, a biopsy plan was developed

based on preprocedure review of standard computed tomography (CT) and bone scan images. The procedure image was reviewed to ensure targeting of metabolic lesions on the nuclear scans, and progressing lesions, if observed, were preferentially biopsied. The plan for each case was to acquire as many core specimens and large blood clots as was safely possible. The shortest path from the skin to the lesion was drawn, attempting to target the periphery of the lesion for the initial cores in sclerotic lesions, or through and through the lesions including the near and far periphery for smaller lesions. Any area was targeted in lytic/mixed lesions. Multiple cores were planned for each case, with the bony yield determining the ultimate number taken. Patients were placed on the CT gantry (Optima CT-580 16 slice wide-bore scanner; GE Healthcare, Little Chalfont, United Kingdom) in supine, prone, or decubitus positions allowing for the selected path and patient comfort. Focal scout images through the target region were performed with 5-mm slices and the patients were prepared in the usual sterile fashion. The skin and periosteum overlying the designated biopsy area were anesthetized with 1% lidocaine (Hospira Inc, Lake Forest, Illinois). Patients received intravenous conscious sedation with midazolam and fentanyl, which was administered and monitored by independent nursing supervision. Biopsies were taken under serial CT guidance (Fig. 3). The following biopsy systems were used for tissue procurement: Arrow OnControl Coaxial Powered Bone Marrow Biopsy System (Teleflex, Morrisville, North Carolina), Temno Evolution Biopsy Device (Bauer, Via del Fosso, Italy), Westcott Fine Needle Aspiration Biopsy (Becton Dickinson, Franklin Lakes, New Jersey), and Madison Comprehensive Bone Biopsy System (Laurane Medical, Westbrook, Connecticut). After each core, aspirates were obtained and allowed to start to clot (Arrow

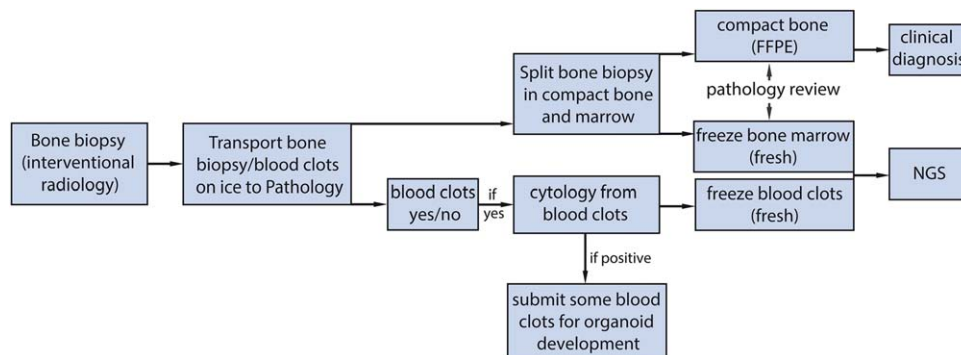


Figure 1. Proposed next-generation bone biopsy protocol in metastatic prostate cancer. Schematic illustration of the biopsy workflow is shown. See additional details in Supporting Information Table 1. FFPE indicates formalin-fixed, paraffin-embedded; NGS, next-generation sequencing.

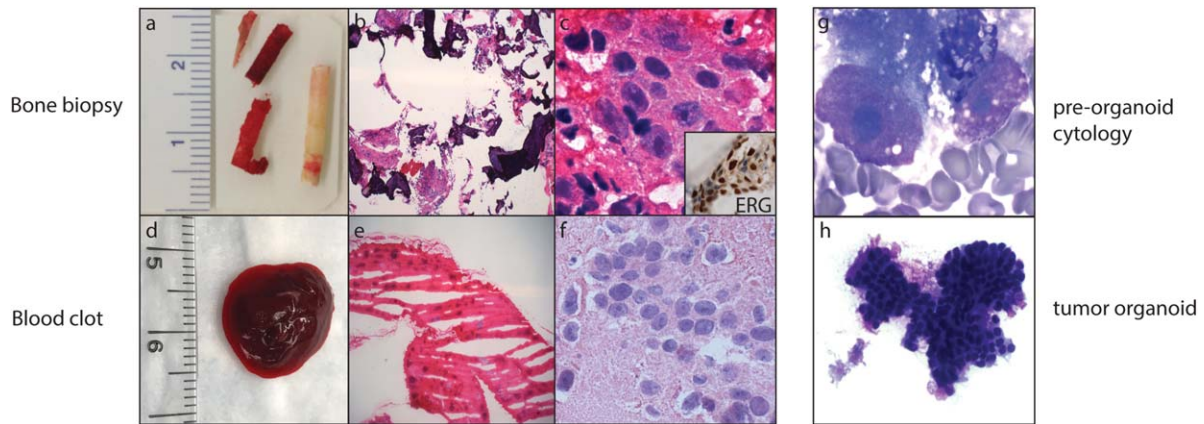


Figure 2. Bone biopsy material used for diagnosis, next-generation sequencing (NGS), and preclinical models. (a) Macroscopic image of a bone biopsy. Compact bone (whitish core) and bone marrow (reddish core) are split upon receipt, and submitted for clinical diagnosis and NGS, respectively. (b and c) Histology of metastatic prostate adenocarcinoma using images from corresponding frozen bone marrow (H & E; low and high magnification). *Inset:* ETS-related gene (ERG)-positive immunohistochemistry in tumor cells. (d) Macroscopic image of a fresh blood clot. (e and f) Histology of metastatic prostate adenocarcinoma using images from a corresponding frozen blood clot (H & E; low and high magnification). (g) Adequacy and tumor cell content are evaluated by cytologic smear examination of the biopsy and/or blood clot (Diff-Quik). (h) Morphologic evaluation of organoid in culture by cytologic smear examination (Diff-Quik).

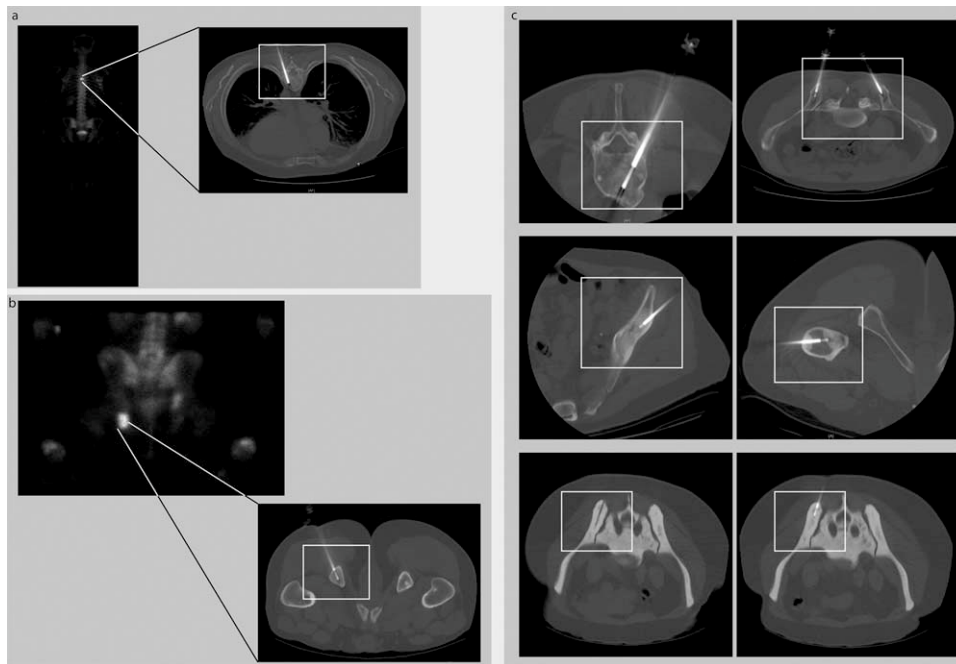


Figure 3. Radiology images of different bone lesions in patients with metastatic prostate cancer who underwent a biopsy using the OnControl drill. (a) T6 vertebral body lesion and (b) lesion of the left ischial tuberosity. (*Left*) Scintigraphic image with clearly visible tracer uptake in both lesions. (*Right*) A corresponding computed tomography (CT) image demonstrating the intralesional biopsy needle. (c) CT images of biopsy procedures in multiple locations in several patients demonstrating safe use of the OnControl drill throughout the bony skeleton. CT-guided biopsy of an extremely dense bone lesion (lower panels: white rectangle, OnControl). The CT images confirm a technically successful biopsy.

OnControl system only), and large clots were placed on a nonadherent pad to be frozen. Bleeding back through the large gauge introducer was allowed, with this material used to gather clot as well. Bone biopsies and aspiration

samples were handed to the assistant, who immediately placed them on top of wet ice. The biopsy needles were removed and sterile dressings were applied. Postprocedure images were routinely obtained to exclude immediate

complications and patients were brought to the recovery area.

Biobanking and Tissue Processing of Bone Biopsies

Bone cores and blood clots were immediately transported on wet ice to the pathology department and tissue was allocated on site for clinical diagnosis, whole-exome sequencing (WES), RNA sequencing, and organoid development (Fig. 1) (see Supporting Information Table 1). Tissue was submitted for organoid development only if the concurrent cytology smear was positive for tumor (Fig. 2).¹¹ If it was negative, tissue destined for organoid development instead was frozen for subsequent next-generation sequencing (NGS) if sufficient tumor cells were present at the time of review of frozen slides. Cores were separated into bone marrow and cortical bone specimens. The latter were submitted for clinical diagnosis following standard decalcification. Both bone marrow and blood clots were snap frozen in optimum cutting temperature compound (Tissue-Tek; Sakura Finetek USA Inc, Torrance, California). Frozen sections were cut from the cryomolds and stained with hematoxylin and eosin (H & E). Clinical (ie, cortical bone) and research/precision medicine (ie, frozen bone marrow and blood clots) H & E-stained slides were reviewed by a board-certified pathologist. The frozen H & E slides also were included during the standard-of-care review process. A standard-of-care report was issued for the cortical bone (clinical) specimen. If the clinical specimen was negative for carcinoma, the tissue on the cryomolds was converted into formalin-fixed, paraffin-embedded tissue (FFPE) for clinical purposes. H & E-stained frozen section slides of bone marrow and/or blood clots always were reviewed for tumor adequacy and tumor quantification before sequencing, and compared with the corresponding clinical material (ie, decalcified cortical bone). Both FFPE tissue slides (surgical pathology clinical slides) and frozen section slides (precision medicine clinical trial slides) were scanned at high definition (svs file format) using the Aperio AT2 system (Leica Biosystems, Nussloch, Germany) into e-Slide manager, a password-protected online platform hosted by the Information, Technologies and Service (ITS) department at Weill Cornell Medicine. These images are integrated in the WES report (for a WES/RNA report example and standard operating procedures, see the Supporting Information).

High-density tumor areas were annotated on the H & E-stained glass slide by the pathologist, and tumor cell content was estimated. A published morphologic classification of metastatic prostate cancer was applied to

annotate histology.¹² Tumor areas on cryomolds were sampled with a 1.5-mm biopsy punch using the annotated H & E slide as guide. DNA and RNA extraction was performed as previously described.¹⁰ The DNA concentration was measured using the Qubit Fluorometric Quantitation system in ng/ μ L (ThermoFisher Scientific, Waltham, Massachusetts) and NanoDrop (NanoDrop Technologies, Wilmington, Delaware).¹³ NanoDrop also was used for RNA concentration and nucleic acids purity (wavelength absorption at 260/280 nm). Germline DNA from blood or buccal swabs served as the control.¹⁰ Subsequent WES/RNA sequencing and computational analysis were performed as described in detail previously.¹⁰ Our custom computational pipeline includes clonality analysis and sample matching check by CLoNeT (ID Genomics, Seattle, Washington) and single-nucleotide polymorphism panel identification assay, respectively.^{14,15}

Statistical Analysis

Comparison between groups was performed using the Fisher exact test, Student *t* test, and one-way analysis of variance. Data collection and analysis was performed with SPSS statistical software (version 24.0; IBM Corporation, Armonk, New York). *P* values < .05 were considered to be statistically significant.

RESULTS

Demographic and Radiologic Characteristics

The median patient age at time of consent was 71 years (range, 51-92 years). Seventy bone biopsies from 64 patients were performed; of these, 49 biopsies (70%) yielded both bone cores and blood clots simultaneously. Bone biopsies were obtained in 63 of 70 cases and blood clots were acquired in 56 of 70 cases. The number of bone biopsy cores ranged from 1 core up to 8 cores. Blood clots only were obtained in 7 of 70 cases (Table 1, Supporting Information Table 2). Of the 70 bone biopsies, 61 (87.1%) were from sclerotic (blastic) lesions, 8 (11.4%) were from lytic lesions, and 1 lesion (1.4%) was mixed blastic/lytic. Biopsy sites were the iliac bone in 38 of 70 cases (54.3%), lumbar vertebra in 10 cases (14.3%), thoracic vertebra in 7 cases (10%), sacrum in 4 cases (5.7%), and femur in 3 cases (4.3%). Two bone biopsies each (altogether 8.6%) were procured from the pubic bone, rib, and ischial bone, and 1 biopsy each (1.5%) was taken from the sternum and humerus (Table 1). In 7 of these biopsies, material was obtained from 2 anatomically distinct sites during the same procedure. In all 7 cases, 1 biopsy was obtained from the iliac bone. The second biopsy was procured from the opposite iliac bone in 3 cases, from the sacrum in 2

TABLE 1. Clinical and Pathological Characteristics of 70 Bone Metastases From 64 Patients

Characteristic	No (%)
No. of procedures	70 (100)
Age, y	
Mean	70.3 ± 8.7
Median	71
Radiologic appearance of bone metastasis	
Blastic	61 (87.1)
Lytic	8 (11.4)
Mixed	1 (1.5)
Location	
Iliac bone	38 (54.3)
Lumbar vertebra	10 (14.3)
Thoracic vertebra	7 (10.0)
Sacrum	4 (5.7)
Femur	3 (4.3)
Pubic bone	2 (2.9)
Rib	2 (2.9)
Ischial bone	2 (2.9)
Sternum	1 (1.4)
Humerus	1 (1.4)
Biopsy diagnostic	
Yes	60 (85.7)
No	10 (14.3)
Histologic subtype ^a	
Adenocarcinoma, usual	55 (91.7)
Adenocarcinoma, with neuroendocrine differentiation	3 (5.0)
Small cell carcinoma	2 (3.3)

^a Percentages based on 60 diagnostic biopsies.

cases, and from sternum and lumbar vertebra in 1 case each. All 7 procedures with 2 biopsy sites (total of 14 biopsies) yielded diagnostic material.

In 60 of 70 bone biopsies (85.7%), tissue was procured using the Arrow OnControl Powered Bone Marrow Biopsy System. The Temno Evolution Biopsy Device (18-gauge), the Madison Comprehensive Bone Biopsy System, and the Westcott Fine Needle Aspiration Biopsy (22-gauge) were used in the remaining 5 cases (7.1%), 4 cases (5.7%), and 1 case (1.4%), respectively. No complications (eg, clinically apparent hematoma or infection) were reported. One patient was found to have 2 sclerotic lesions in the right iliac bone. We performed an additional sodium fluoride (¹⁸F-NaF) positron emission tomography (PET)-magnetic resonance imaging (MRI) to identify the most metabolic active lesions.¹⁶ Based on the results, the biopsy plan subsequently was adjusted to target L5 vertebra and anterior superior iliac spine (Fig. 4).

Biopsy Processing and Histopathology

Sixty of 70 tissue samples (85.7%) contained enough tumor tissue to establish a diagnosis, and the other 10 samples (14.3%) were nondiagnostic. Sufficient FFPE tissue was available with which to perform routine immunohistochemistry in 33 of 60 cases (55.0%). Pathology

classification of positive biopsies demonstrated that 55 of 60 cases (91.7%) contained usual prostate adenocarcinoma histology (Table 1).¹²

Metrics for NGS

Tumor tissue was sufficient for sequencing in 49 of the 60 diagnostic cases (81.7%). The tumor content was estimated by histopathology evaluation (data not shown) and by using an algorithm to assess tumor DNA purity and cancer cell ploidy (CLONeT),¹⁴ and ranged between 10% and 93.9% (mean CLONeT tumor purity in bone biopsies: 49.6%; mean CLONeT tumor purity in blood clots: 55.9%). No differences in DNA yield, tumor cell content, coverage, and capture efficiency was observed between anatomic locations. Data regarding the DNA concentration from the bone/blood clot samples were available for 46 of 60 samples (76.7%). The median DNA concentration overall was 25.5 ng/μL (range, 0.53-164.40 ng/μL), was 41.4 ng/μL for DNA from blood clots, and was 32.3 ng/μL for DNA from bone biopsies ($P = .40$). The median 260/280 ratio was 1.81 for DNA and 1.95 for RNA, and thus was sufficient for subsequent NGS. WES, which required a minimum of 225 ng DNA, was successful in 49 of 60 diagnostic cases (81.7%).¹⁰ The mean input was 232.72 ng of DNA for WES and 323.13 ng of RNA for RNA sequencing. WES was successfully performed from blood clots in 26 of 56 cases (46.4%) and from bone biopsies in 23 of 63 cases (36.5%). The mean capture efficiency was 83.6% and the mean coverage was 94.7% (see Supporting Information Table 2). Recurrent alterations were as expected in metastatic prostate cancer and included mutation or amplification of the *AR* gene (4.0% mutation and 32.7% amplification), deletion of *PTEN* gene (10.2%), mutation or deletion of the *TP53* gene (14.3% mutation and 2% deletion), mutation or deletion of *BRCA2* (10.2% mutation and 4.0% deletion), and mutation of the *SPOP* gene (4%).⁵

In 37 of 60 diagnostic cases (61.7%), the amount of tissue was insufficient to perform concurrent DNA and RNA extraction, and WES was prioritized. There was enough tissue to also run RNA sequencing in 23 of 60 diagnostic cases (requiring a minimum of 100 ng of RNA), among which it was performed successfully in 20 cases (87.0%). Data regarding the RNA concentration from the bone/blood clot samples were available for 11 samples. The median RNA concentration overall was 16.2 ng/μL. The remaining 3 samples (13.0%) failed RNA sequencing due to poor per base sequence quality (see Supporting Information Table 2 for quality control metrics and additional results).

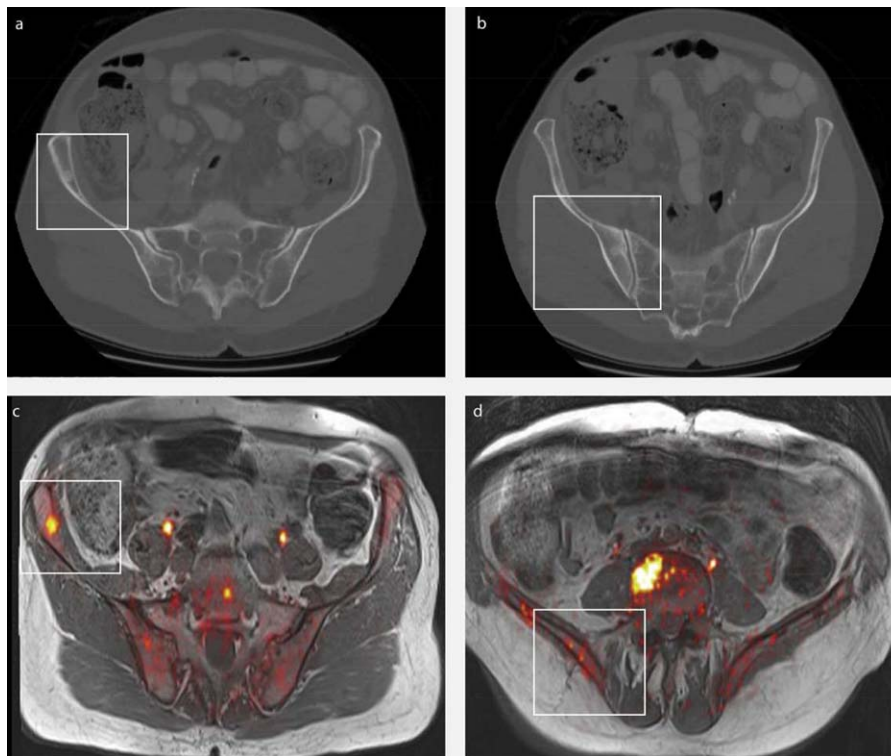


Figure 4. Addition of positron emission tomography (PET)-magnetic resonance imaging (MRI) to help identify which lesion to biopsy. (a and b) A 73-year-old patient with 2 bone lesions in the right iliac bone (white rectangle) shown on computed tomography. ^{18}F -sodium fluoride (^{18}F -NaF) PET-MRI clearly identified (c) tracer enrichment in the anterior lesion, (d) with no enrichment noted in the posterior lesion. The biopsy plan subsequently was changed to target the anterior lesion and the L5 vertebra.

DISCUSSION

Patients with advanced prostate cancer commonly harbor sclerotic bone metastases, which are challenging sites to biopsy and the tumor tissue obtained often is limited. We developed a protocol to successfully procure high-quality nucleic acids from metastatic tumors for sequencing, specifically from bone biopsy specimens. We focused on optimizing techniques that limit patient risk by limiting procedure time and radiation exposure and maximizing tissue samples such as the use of blood clots for successful NGS. We also wanted to develop a protocol that did not rely on time-consuming techniques such as laser capture microdissection because these are not always available. Instead, we sampled cryomolds with a biopsy needle using the annotated H & E slide as a guide; this technique resulted in tumor purity that was sufficient for subsequent NGS in the vast majority of cases.

Percutaneous bone biopsies have a low complication rate, with a range of 0% to 7.4% reported in the literature.¹⁷ Common complications include muscular hematomas and bleeding and, rarely, biopsy device fracture, nerve injury, and cerebrospinal fluid leak.¹⁸ No

complications were noted in the current series, which correlates with other published series using a drill-assisted biopsy system.¹⁹ Fracture of a biopsy device in the thoracic spine has been reported with the use of a 16-gauge Bonopty device (AprioMed, Uppsala, Sweden).¹⁸ In a series of 162 consecutive patients in which the Bonopty device was used, sclerotic lesions were significantly associated with a lower diagnostic yield.²⁰ This difference was not observed in the current study cohort: 85.2% and 87.7% of sclerotic and lytic lesions, respectively, were found to be diagnostic ($P = .187$). This also is a significantly higher rate of diagnostic biopsies than previously reported for CT-guided biopsies. Table 2 shows a comparison of the current study results with other published studies of bone biopsies in metastatic prostate cancer.^{8,21-23} In a series of 39 biopsies from patients with metastatic prostate cancer, approximately 77% were found to have at least 1 positive biopsy core. The authors found that a skin-to-bone distance < 6.1 cm was associated with a higher tumor yield on biopsy.²¹ In contrast, in the current series, we were able to successfully sequence 88.5% of cases (23 of 26 cases) in which the skin-to-bone

TABLE 2. Comparison of Results of the Current Study With Other Published Studies of Bone Biopsies in Patients With Metastatic Prostate Cancer

No. of Procedures	Image Guidance?	Diagnostic Yield	Sufficiency for Molecular Studies	Study	PMID
80	Yes	69%	64%	Holmes et al. ²³	28549709
54	Yes	67%	39%	Spritzer 2013 ⁸	23925271
115	No	65.2%	Not performed	Lorente 2016 ²²	27246360
39	Yes	77%	Not performed	McKay 2014 ²¹	25091040
70	Yes	85.7%	81.7%	Current study	NA

Abbreviations: NA, not applicable; PMID, PubMed unique identifier.

distance was ≥ 6.1 cm. We observed that the drill-assisted biopsy system improved the diagnostic yield, and thus there was no need to stratify patients further according to published radiologic and clinical criteria for selection because the biopsy was likely to be diagnostic.²² In 1 patient, additional ¹⁸F-NaF PET-MRI was performed to evaluate 2 lesions in the iliac bone and the biopsy site subsequently was adjusted.¹⁶ PET-MRI has been shown to be more sensitive than PET-CT in detecting bone metastases.²⁴ It might be a valuable additional tool for identifying those lesions promising the greatest yield of tumor tissue.

To the best of our knowledge, the current study is the first to describe the use of blood clots for NGS, which can be obtained through the drill-assisted biopsy systems. Although cortical bone lesions may have to be decalcified, blood clots can be frozen and further used without additional processing.

Inadequate biopsies may subject the patient to risk without benefit. To our knowledge to date, little has been published assessing the quality of bone biopsies for NGS. One study found a sequencing failure rate of 5 of 21 FFPE bone biopsies, although the reason for this result is uncertain due to a lack of information regarding preanalytical variables such as decalcification.⁷ In the current study, the median DNA yield was sufficient for WES (requiring > 225 ng) in the majority of cases and also would have been sufficient for other commonly used targeted gene panels (requiring 10-300 ng).²⁵

In 7 patients, concurrent biopsies from 2 anatomically distinct sites were performed with an excellent diagnostic yield of 100% and subsequent successful WES in 5 cases (71.4%). Although performing double biopsies increased the procedural time slightly, no complications were noted. This approach could be beneficial to increase the overall amount of tissue. The additional radiation exposure is negligible and could avoid additional biopsy procedures should the initial attempt fails to be diagnostic.

Although liquid biopsies such as circulating tumor DNA are attractive as noninvasive methods for molecular biomarker detection, metastatic tumor biopsies do provide additional information that is not currently well captured by liquid approaches. For example, our approach is to ensure that bone biopsies, and blood clots when available, are handled in a way that allows for morphology assessment and also preserves RNA. Spritzer et al reported a 39% yield for RNA using a 15-gauge Bonopt system. In that study, all 5 iliac bone biopsies in sites other than those adjacent to the sacroiliac joint were insufficient for RNA isolation.⁸ In the current study, 6 of 7 sclerotic lesions and 2 of 2 of lytic lesions in such sites were found to be adequate for sequencing.

Conclusions

Historically, the goal of cancer biopsies has aligned with the concept of adequacy, or obtaining just enough sample to allow pathologists to confirm a diagnosis. It now is recognized that antineoplastic treatment changes the molecular landscape of metastatic tumors.^{5,26,27} Our experience highlights a necessary shift in paradigm in an the era of precision medicine from adequate to sufficient biopsies to address both diagnostic and molecular needs. The results presented in the current study demonstrate that this can be performed safely without significant complications.

FUNDING SUPPORT

Supported by National Institutes of Health grants R01 CA116337 (to Mark A. Rubin and Himisha Beltran) and 5U01 CA111275-09 (to Mark A. Rubin and Juan Miguel Mosquera), Department of Defense grant PCRP PC121341 (to Himisha Beltran), the Starr Cancer Consortium (to Mark A. Rubin and Himisha Beltran), and the Nuovo Soldati Foundation (to Joanna Cyrta). This work also was supported by the Translational Research Program at Weill Cornell Medicine Pathology and Laboratory Medicine.

CONFLICT OF INTEREST DISCLOSURES

The authors made no disclosures.

AUTHOR CONTRIBUTIONS

Verena Sailer: Conceptualization, tissue processing and histopathological review, methodology, formal analysis, data curation, writing-original draft, and writing-review and editing. **Marc H. Schiffman:** Biopsy procedures, data curation, and writing-original draft. **Myriam Kossai:** Tissue processing and histopathological review and data curation. **Joanna Cyrra:** Tissue processing and histopathological review and data curation. **Shaham Beg:** Tissue processing and histopathological review and data curation. **Brian Sullivan:** Biopsy procedures. **Bradley B. Pua:** Biopsy procedures. **Kyungmouk Steve Lee:** Biopsy procedures. **Adam D. Talenfeld:** Biopsy procedures. **David M. Nanus:** Investigation and patient consent. **Scott T. Tagawa:** Investigation and patient consent. **Brian D. Robinson:** Tissue processing and histopathological review. **Rema A. Rao:** Tissue processing and histopathological review. **Chantal Pauli:** Tissue processing and histopathological review and data curation. **Rohan Bareja:** Data analysis. **Luis S. Beltran:** Radiographic procedure and analysis. **Alexandros Sigaras:** Data analysis. **Kenneth Wa Eng:** Data analysis. **Olivier Elemento:** Data analysis. **Andrea Sboner:** Data analysis. **Mark A. Rubin:** Conceptualization, funding acquisition, and writing-review and editing. **Himisha Beltran:** Conceptualization, investigation, funding acquisition, patient consent, and writing-original draft. **Juan Miguel Mosquera:** Conceptualization, supervision, funding acquisition, and writing-original draft.

REFERENCES

- Halabi S, Kelly WK, Ma H, et al. Meta-analysis evaluating the impact of site of metastasis on overall survival in men with castration-resistant prostate cancer. *J Clin Oncol*. 2016;34:1652-1659.
- Bubendorf L, Schopfer A, Wagner U, et al. Metastatic patterns of prostate cancer: an autopsy study of 1,589 patients. *Hum Pathol*. 2000;31:578-583.
- Siegel RL, Miller KD, Jemal A. Cancer statistics, 2017. *CA Cancer J Clin*. 2017;67:7-30.
- Mateo J, Carreira S, Sandhu S, et al. DNA-repair defects and olaparib in metastatic prostate cancer. *N Engl J Med*. 2015;373:1697-1708.
- Robinson D, Van Allen EM, Wu YM, et al. Integrative clinical genomics of advanced prostate cancer. *Cell*. 2015;161:1215-1228.
- Van Allen EM, Foye A, Wagle N, et al. Successful whole-exome sequencing from a prostate cancer bone metastasis biopsy. *Prostate Cancer Prostatic Dis*. 2014;17:23-27.
- Zheng G, Lin MT, Lokhandwala PM, et al. Clinical mutational profiling of bone metastases of lung and colon carcinoma and malignant melanoma using next-generation sequencing. *Cancer*. 2016;124:744-753.
- Spritzer CE, Afonso PD, Vinson EN, et al. Bone marrow biopsy: RNA isolation with expression profiling in men with metastatic castration-resistant prostate cancer—factors affecting diagnostic success. *Radiology*. 2013;269:816-823.
- Ross RW, Halabi S, Ou SS, et al; Cancer and Leukemia Group B. Predictors of prostate cancer tissue acquisition by an undirected core bone marrow biopsy in metastatic castration-resistant prostate cancer—a Cancer and Leukemia Group B study. *Clin Cancer Res*. 2005;11:8109-8113.
- Beltran H, Eng K, Mosquera JM, et al. Whole-exome sequencing of metastatic cancer and biomarkers of treatment response. *JAMA Oncol*. 2015;1:466-474.
- Sailer V, Pauli C, Merzier EC, et al. On-site Cytology for Development of Patient-Derived Three-dimensional Organoid Cultures - A Pilot Study. *Anticancer Res*. 2017;37:1569-1573.
- Epstein JI, Amin MB, Beltran H, et al. Proposed morphologic classification of prostate cancer with neuroendocrine differentiation. *Am J Surg Pathol*. 2014;38:756-767.
- Mardis E, McCombie WR. Library quantification: fluorometric quantitation of double-stranded or single-stranded DNA samples using the Qubit system. *Cold Spring Harb Protoc*. 2017;2017: pdb.prot094730.
- Prandi D, Baca SC, Romanel A, et al. Unraveling the clonal hierarchy of somatic genomic aberrations. *Genome Biol*. 2014;15:439.
- Demichelis F, Greulich H, Macoska JA, et al. SNP panel identification assay (SPIA): a genetic-based assay for the identification of cell lines. *Nucleic Acids Res*. 2008;36:2446-2456.
- Rakheja R, Chandarana H, Ponzio F, et al. Fluorodeoxyglucose positron emission tomography/magnetic resonance imaging: current status, future aspects. *PET Clin*. 2014;9:237-252.
- Hau A, Kim I, Kattapuram S, et al. Accuracy of CT-guided biopsies in 359 patients with musculoskeletal lesions. *Skeletal Radiol*. 2002; 31:349-353.
- Shaikh H, Thawani J, Pukenas B. Needle-in-needle technique for percutaneous retrieval of a fractured biopsy needle during CT-guided biopsy of the thoracic spine. *Interv Neuroradiol*. 2014;20: 646-649.
- Wallace AN, McWilliams SR, Wallace A, et al. Drill-assisted biopsy of the axial and appendicular skeleton: safety, technical success, and diagnostic efficacy. *J Vasc Interv Radiol*. 2016;27:1618-1622.
- Li Y, Du Y, Luo TY, et al. Factors influencing diagnostic yield of CT-guided percutaneous core needle biopsy for bone lesions. *Clin Radiol*. 2014;69:e43-e47.
- McKay RR, Zukotynski KA, Werner L, et al. Imaging, procedural and clinical variables associated with tumor yield on bone biopsy in metastatic castration-resistant prostate cancer. *Prostate Cancer Prostatic Dis*. 2014;17:325-331.
- Lorente D, Omlin A, Zafeiriou Z, et al. Castration-resistant prostate cancer tissue acquisition from bone metastases for molecular analyses. *Clin Genitourin Cancer*. 2016;14:485-493.
- Holmes MG, Foss E, Joseph G, et al. CT-Guided Bone Biopsies in Metastatic Castration-Resistant Prostate Cancer: Factors Predictive of Maximum Tumor Yield. *J Vasc Interv Radiol*. 2017;28:1073-1081.
- Catalano OA, Nicolai E, Rosen BR, et al. Comparison of CE-FDG-PET/CT with CE-FDG-PET/MR in the evaluation of osseous metastases in breast cancer patients. *Br J Cancer*. 2015;112:1452-1460.
- Chen H, Luthra R, Goswami RS, Singh RR, Roy-Chowdhuri S. Analysis of pre-analytic factors affecting the success of clinical next-generation sequencing of solid organ malignancies. *Cancers (Basel)*. 2015;7:1699-1715.
- Pao W, Miller VA, Politi KA, et al. Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain. *PLoS Med*. 2005;2:e73.
- Cancer Genome Atlas Research Network. The molecular taxonomy of primary prostate cancer. *Cell*. 2015;163:1011-1025.

Titre : Le rôle ambigu du complexe SWI/SNF dans les cancers, illustré à travers deux exemples : les carcinomes à petites cellules de l'ovaire de type hypercalcémiant (SCCOHT) et le cancer de la prostate.

Mots clés : SWI/SNF, épigénétique, SMARCA4, SCCOHT, cancer de la prostate, carcinome neuroendocrine.

Résumé : Le complexe de remodelage de la chromatine SWI/SNF est un régulateur épigénétique majeur impliqué dans le développement embryonnaire et dans la différenciation cellulaire. De plus, les gènes qui encodent les sous-unités de SWI/SNF sont altérés dans au moins 20% de cancers. Bien que le complexe SWI/SNF soit le plus souvent considéré comme suppresseur des tumeurs, il existe des preuves croissantes que le rôle de SWI/SNF dans le cancer peut dépendre du type de tissu et du contexte.

Dans la première partie de cette dissertation, nous présentons la caractérisation moléculaire d'une cohorte indépendante de carcinomes à petites cellules de l'ovaire de type hypercalcémiant (SCCOHT), comme exemple d'un cancer sous-tendu par des altérations perte-de-fonction de la sous unité catalytique de SWI/SNF, SMARCA4.

Dans la deuxième partie, nous explorons le rôle du SWI/SNF dans le cancer de la prostate (CP), y compris ses formes les plus agressives : le CP résistant à la castration et le carcinome neuroendocrine. Alors que les mutations des gènes de SWI/SNF sont très rares dans le CP, nous montrons que l'expression de certaines sous-unités peut être dérégulée et qu'une haute expression de SMARCA4 est associée à des CP agressifs. De plus, nous montrons que plusieurs lignées cellulaires de CP dépendent de SWI/SNF pour leur croissance.

Au total, ces deux exemples supportent l'hypothèse que SWI/SNF peut jouer des rôles différents dans le cancer en fonction du type tumoral.

Title: A pleiotropic role of the SWI/SNF complex in cancer - insights from two tumor types: small cell carcinoma of the ovary, hypercalcemic type and prostatic carcinoma.

Keywords: SWI/SNF, epigenetics, SMARCA4, SCCOHT, prostate cancer, neuroendocrine carcinoma

Abstract: The SWI/SNF chromatin remodeling complex is a major epigenetic regulator involved in embryonic development and in cell differentiation. In addition, genes encoding components of SWI/SNF are altered in at least 20% of cancers. Even though the SWI/SNF complex is usually regarded as a tumor suppressor, there is increasing evidence that the role of SWI/SNF in cancer may be tissue type- and context-dependent.

In the first part of this dissertation, we present the molecular characterization of an independent cohort of small cell carcinomas of the ovary, hypercalcemic type (SCCOHT), as an example of a malignancy driven by loss-of-function alterations of the catalytic subunit of SWI/SNF, SMARCA4.

In the second part, we explore the role of SWI/SNF in prostate cancer (PCa), including its most aggressive forms: castration-resistant prostate cancer and neuroendocrine prostate cancer. We show that while SWI/SNF mutations are exceedingly rare in PCa, the expression of several SWI/SNF subunits can be deregulated and that high SMARCA4 expression is associated with aggressive PCa. In addition, we show that many PCa cell lines are dependent on SWI/SNF for their growth.

Taken together, these two examples further support the hypothesis that SWI/SNF can play different roles in cancer, depending on the tumor type.