Investigating the Role of ARID1A Inactivation in Colon Cancer Pathogenesis

Investigating the Role of ARID1A Inactivation in Colon Cancer Pathogenesis

A dissertation presented

by

Radhika Mathur

to

The Division of Medical Sciences

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Biological and Biomedical Sciences

Harvard University

Cambridge, Massachusetts

April 2017

Dissertation Advisors: Drs. Charles W. M. Roberts and Stuart H. Orkin Radhika Mathur

Investigating the Role of ARID1A Inactivation in Colon Cancer Pathogenesis

Abstract

The genomics-era search for cancer-causing mutations has revealed epigenomic regulators, in particular SWI/SNF chromatin remodeling complexes, as major targets of mutation. Genes encoding SWI/SNF subunits are collectively mutated in 20% of all human cancers, showing a broad pattern of mutations across solid epithelial, brain, and hematological malignancies. SWI/SNF complexes are evolutionarily conserved regulators of transcription with ATP-dependent chromatin remodeling activity. However, the precise mechanisms by which they function in mammalian cells have been unclear, as have their roles in malignancy.

Here, we investigate the tumor suppressor role of ARID1A, the SWI/SNF subunit that is most frequently mutated in human cancer. We demonstrate that inducible ARID1A inactivation in mice drives the formation of invasive colon adenocarcinoma with remarkable resemblance to the corresponding human cancer. Tumor formation does not require cooperating mutations in genes associated with human colon cancer. Tumors also do not show aberrant Wnt signaling, an initiating event in genetic models of colon cancer pathogenesis. Rather, ARID1A inactivation antagonizes tumorigenesis driven by aberrant Wnt signaling.

Our investigation reveals that ARID1A targets SWI/SNF complexes to enhancers, where they function in control of enhancer activity. Upon ARID1A inactivation, SWI/SNF binding is lost at the majority of enhancers, which subsequently lose activity, showing reduced levels of H3K27acetylation and downregulation of nearest genes. Residual complexes containing ARID1B preserve SWI/SNF function at a subset of enhancers, but defects in SWI/SNF targeting and control of enhancer activity cause extensive dysregulation of gene expression.

These results implicate enhancer-mediated gene regulation as a principal tumor suppressor function of ARID1A in the colon epithelium, with broad relevance to other SWI/SNF-mutant cancers. ARID1B has been identified as a synthetic lethal vulnerability in *ARID1A*-mutant human cancers; these results suggest that defective SWI/SNF control of enhancer activity drives tumorigenesis and also confers vulnerabilities that might be targeted for therapy. These findings represent an advance in colon cancer modeling, establishing a novel pathway to colon tumorigenesis and a new mouse model that recapitulates features of the human disease – aggressive local tissue invasion, long latency, exclusive origin to the colon rather than the small intestine – that are absent in current, widely utilized models.

Table of Contents

Acknowledgements

Thank you to my thesis advisor, Dr. Charles Roberts, for allowing me the opportunity and resources to conduct this work, and for guiding me expertly every step of my journey toward becoming a scientist – whether from Boston or from Tennessee. Thank you, Dr. Stuart Orkin and Dr. Ramesh Shivdasani, for being the reservoirs of wisdom and mentorship that I have called upon throughout this journey. Thank you to my rotation advisors, Dr. David Frank and Dr. Loren Walensky, for giving me a running start, and my dissertation advisory committee members, Dr. Brad Bernstein, Dr. George Daley, and Dr. Rick Young, for motivating me to strive ever higher.

I am ever grateful for the support and lessons from Dr. Xiaofeng Wang, Dr. Boris Wilson, and Dr. Adrianna San Roman in the laboratory. I'm grateful to have shared in the worlds of bioinformatics with Dr. Burak Alver and Dr. Peter Park, and of histopathology with Dr. Tony Agoston. This project would not have reached its full potential without their contributions and expertise. Thank you to the laboratory technicians – Jeff Haswell, Haley Manchester, Weishan Wang, Kelly Sullivan, Ev Tzvetkov, Tam Pham, Emma Troisi, Ben Hubbell-Engler – for the incredible work that you do. Thank you Elaine Oberlick and Brittany Michel for your unwavering support in your many roles as labmates, friends, and occasional cat-sitters.

Thank you to my family for never letting me feel alone in this journey. Thank you to my parents for always believing in me, teaching me to believe in myself, clearing every barrier that you could, and trying to clear even more. Thank you to my amazing little sister for being the best friend that I have needed these past years. Thank you to my grandparents for being sources of endless inspiration and motivation. To the relatives and friends around the world who have supported me in this journey – thank you for being the best cheerleaders. I have needed you, and I am so lucky to have had you. I cannot wait to celebrate with you at the finish line.

Chapter 1:

Introduction

Discovery and characterization of SWI/SNF complexes

In 1984, two independent yeast mutagenesis screens identified SWI and SNF genes as required for transcriptional activation of genes involved in mating-type switching (SWI, Stern et al. 1984) and sucrose fermentation (sucrose non-fermenting, SNF, Neigeborn et al. 1984). SWI and SNF genes were found to encode components of a multi-subunit complex required for transcriptional activation of a large number of inducible genes in yeast (Peterson and Herskowitz, 1992). Subsequent studies in yeast characterized SWI/SNF complexes as chromatin remodelers, linking their role in transcriptional activation with their ability to mobilize nucleosomes using ATP-hydrolysis and to facilitate the binding of activators to their targets in nucleosomal DNA (Hirschhorn et al. 1992, Laurent et al. 1993, Cairns et al. 1994, Côté 1994).

In 1998, SWI/SNF complexes were independently identified in a screen in *Drosophila melanogaster* as regulators of homeotic genes, which control body segment identity and are transcribed in different subsets of cells of the developing embryo. Polycomb group proteins had previously been identified as chromatin-based silencers of homeotic genes not transcribed in a given cell; the *brm* gene was identified a suppressor of the Polycomb-mutant phenotype (Kennison and Tamkun 1988, Kennison 1995). *Brm* shared extensive sequence similarity with the ATPase subunit of yeast SWI/SNF (Tamkun et al. 1992), and was essential in *Drosophila* for embryogenesis and for all stages of development (Elfring et al. 1998).

In humans, SWI/SNF complexes were identified by sequence homology with SWI/SNF ATPase subunits in yeast and *Drosophila* (Khavari et al. 1993). Chromatin remodeling activity was conserved in human SWI/SNF complexes, as demonstrated by their ability to disrupt nucleosomes and facilitate activator binding on chromatin templates *in vitro* (Imbalzano et al. 1994, Kwon et al. 1994). Immunoprecipitation revealed extensive heterogeneity in subunit

composition of SWI/SNF complexes between tissue types and stages of development, and also within individual cells (Wang et. al 1996a, Wang et. al 1996b). Reconstitution of only four SWI/SNF subunits – an ATPase (SMARCA4 or SMARCA2) and three conserved subunits (SMARCB1, SMARCC1, SMARCC2) – was sufficient for chromatin remodeling *in vitro* (Phelan et al. 1999). The remaining subunits were suggested to function in the regulation of chromatin remodeling activity *in vivo* (Kingston and Narlikar 1999).

Roles for mammalian SWI/SNF complexes *in vivo* were described first in hormone-inducible signaling, as stimulation of the glucocorticoid receptor (GR) targeted SWI/SNF activity to the GR response element, triggering nucleosome disruption and activating transcription (Ostlund Farrants et al. 1997). Roles for SWI/SNF complexes were also suggested in development, as SWI/SNF complexes interacted directly with lineage-specifying factors, such as with C/EBPβ in hematopoiesis to activate genes of the myeloid lineage (Kowenz-Leutz and Leutz 1999). The importance of SWI/SNF complexes in mammalian development was established when SWI/SNF complexes were found to be required for MyoD-mediated induction of muscle differentiation genes in fibroblasts (la Serna et al. 2001), and for differentiation of a variety of other cell types including adipocytes (Pedersen et al. 2001), lymphocytes (Gebuhr et al. 2003), hepatocytes (Gresh et al. 2005), and neurons (Matsumoto et al. 2006). Investigation of SWI/SNF function in neuronal development revealed that a switch in subunit composition occurred during the transition of proliferating neural progenitors to post-mitotic neurons, and that this switch was essential for neuronal differentiation to occur (Lessard et al. 2007). A crucial link between SWI/SNF subunit composition and function was further indicated by the study of SWI/SNF complexes in embryonic stem cells, where SWI/SNF complexes had specialized

subunit assemblies tailored for interaction with Oct4 and Sox2 in the regulation of pluripotency genes (Ho et al. 2009).

It is now well established that mammalian SWI/SNF complexes form by combinatorial assembly with a large number of subunits, and that diverse subunit assemblies provide dynamic gene regulation across cell types and stages of development (**Figure 1-1**, Wu et al. 2009, Wilson and Roberts 2011, Wang et al. 2014). Mammalian SWI/SNF complexes contain one of two functionally distinct ATPase subunits (SMARCA4 and SMARCA2), a set of conserved core subunits, and variant subunits that are often encoded by multi-gene families and expressed in a lineage-restricted manner. For example, the BAF60 gene family encodes three variants of the BAF60 subunit and BAF60c is selectively incorporated into SWI/SNF complexes in the embryonic heart (Lickert 2004). However, much remains to be elucidated of the contributions of individual subunits to SWI/SNF function and of the mechanisms by which SWI/SNF complexes assemble and regulate transcription in mammalian cells.

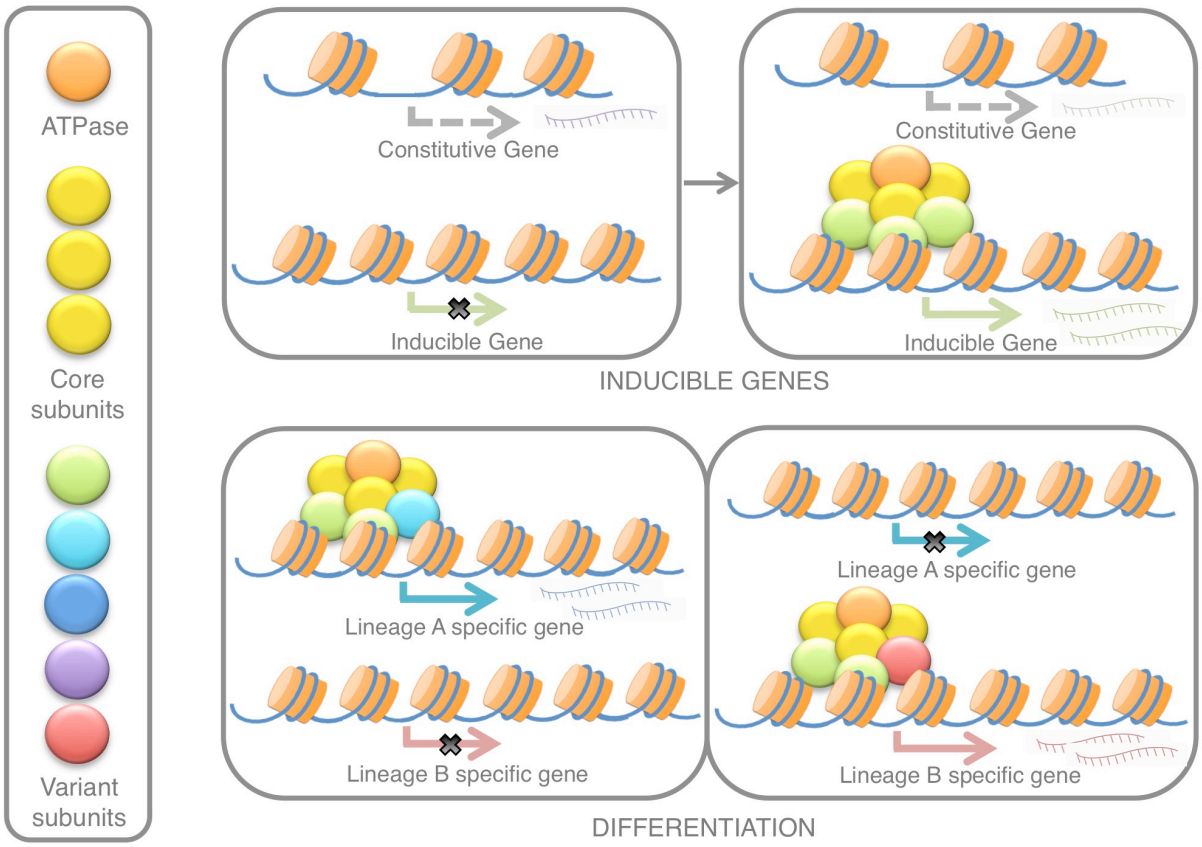


Figure 1-1: Transcriptional regulation by SWI/SNF complexes in mammalian cells. SWI/SNF complexes are multi-subunit protein complexes that form via combinatorial assembly. They consist of a catalytic ATPase subunit, a set of conserved core subunits, and several variant subunits that contribute to targeting and specificity. SWI/SNF complexes utilize ATP hydrolysis to mobilize nucleosomes and alter the accessibility of nucleosomal DNA to transcriptional and co-regulatory machinery. SWI/SNF complexes are crucial for regulating the transcription of inducible genes and for regulating the transcription of lineage-specific genes during differentiation.

SWI/SNF complexes in human cancer

Cancer is generally described as a genetic disease caused by a sequential accumulation of mutations, each of which confers upon cells a selective growth advantage and other "hallmark capabilities" of cancer (Hanahan and Weinberg 2000, Vogelstein et al. 2013). Technological advances of the genomics era have facilitated the search for cancer-causing mutations, with large-scale efforts initiated to comprehensively characterize mutational landscapes of human cancers ("The Cancer Genome Atlas" 2017, "International Cancer Genome Consortium" 2017). Unexpectedly, these genomic studies have identified epigenomic regulators as major targets of mutation in human cancer (Garraway and Lander 2013). Subunits of SWI/SNF chromatin remodeling complexes, in particular, are altered with a high frequency – 20% of all human cancers – and broad mutational pattern resembling the most frequently mutated tumor suppressor, P53 (Shain et al. 2013, Kadoch et al. 2013, Masliah-Planchon et al. 2015).

SWI/SNF complexes were first linked to cancer in 1998, when inactivation of the SMARCB1 subunit was identified as the hallmark characteristic of malignant rhabdoid tumor, a rare and highly aggressive cancer of early childhood (Versteege 1998). *Smarcb1* inactivation in mice drives a highly penetrant cancer phenotype with a median onset of only 11 weeks (Roberts et al. 2002, Roberts and Orkin 2004), unprecedented for inactivation of any tumor suppressor gene including *P53* (which drives a variety of neoplasms with mean onset of 20 weeks, Donehower et al. 1992). Malignant rhabdoid tumors were found to be diploid and genomically stable (McKenna et al. 2008); whole-exome sequencing has placed these amongst the tumor types with the lowest rate of somatic mutation, with some tumor samples showing no mutations in protein-coding genes other than *SMARCB1* (Lee et al. 2012, Lawrence et al. 2013).

In vivo and *in vitro* studies of SMARCB1 inactivation have revealed that SWI/SNF function is perturbed rather than abrogated in malignant cells. Inactivation of SMARCB1 was lethal in normal cellular contexts (Roberts et al. 2002), unexpected for loss of a tumor suppressor. In the context of SMARCB1-deficient cancer, inactivation of the catalytic subunit SMARCA4 was lethal, indicating presence of residual SWI/SNF function (Wang et al. 2009). These results were extended to the broader spectrum of SWI/SNF-mutant cancers when synthetic lethality-based screening approaches identified mutually exclusive SWI/SNF subunits as top vulnerabilities in SWI/SNF-mutant cancers (ARID1B in *ARID1A*-mutant cancers, Helming et al. 2014a, and SMARCA2 in *SMARCA4*-mutant cancers, Wilson et al. 2014, Hoffman et al. 2014). Residual complexes were thus implicated as potential means of therapeutically targeting SWI/SNF-mutant cancers (Helming et al. 2014b), although it remained unclear as to whether and how residual complexes might actively contribute to the mechanism of oncogenesis.

Several other lines of investigation suggest potential mechanisms underlying SWI/SNF mutations in cancer. SWI/SNF complexes interact directly with transcription factors that are classical oncogenes (for example, MYC, Cheng et al. 1999) and tumor suppressor genes (for example, P53, Lee et al. 2002), and regulate transcriptional programs downstream of signaling pathways commonly deregulated in cancer (for example, Gli-Hedgehog, Jagani et al. 2010, and Wnt, Mora-Blanco et al. 2013). Beyond transcriptional regulation, SWI/SNF complexes also have roles in other cellular processes requiring chromatin-remodeling activity such as DNA repair (Hara et al. 2002, Gong et al. 2006, Park et al. 2006) and decatenation (Dykhuizen 2013, Miller 2017). Deregulation of classical cancer pathways and defects in DNA repair/decatenation provide direct routes to oncogenesis. However, they do not explain certain observations from genome sequencing of SWI/SNF-mutant cancers. These include (1) the failure of mutual

exclusivity analyses to identify key genes that mediate tumorigenesis (Shain et al. 2013); (2) exquisite context specificity in mutational patterns of individual SWI/SNF subunits, with *PBRM1*, for example, mutated with high frequency in renal clear cell carcinoma, but infrequently in other tumor types (Masliah-Planchon et al. 2015); and (3) apparent absence of genomic instability in rare, but highly aggressive cancers driven by inactivation of SWI/SNF subunits (malignant rhabdoid tumor, Lee et al. 2012, and small cell carcinoma of the ovary, hypercalcemic type, Jelinic et al. 2014).

Another line of investigation has focused on a possible role for Polycomb group complexes in SWI/SNF-mutant cancers, due to the genetic link between SWI/SNF and Polycomb complexes in *Drosophila* (Kennison 1995). A functionally antagonistic relationship between these two complexes was also identified in malignant rhabdoid tumor (Kia et al. 2008) and in mouse models of SMARCB1 inactivation, where inactivation of Polycomb group complexes blocked onset of SMARCB1-deficient tumors (Wilson et al. 2010). The catalytic subunit of Polycomb group complexes, the EZH2 histone methyltransferase, was thus proposed as a potential therapeutic target in *SMARCB1*-mutant cancer. Therapeutic development of EZH2 inhibitors has since progressed at a remarkable pace, showing early success in clinical trials for patients with *SMARCB1*-mutant cancer (Knutson et al. 2013, Kim and Roberts 2016) and promise in preclinical studies with other SWI/SNF-mutant cancers (Fillmore et al. 2015, Bitler et al. 2015). However, as with residual SWI/SNF complexes, it has been unclear as to whether and how Polycomb activity might actively contribute to the mechanism of oncogenesis underlying SWI/SNF-mutant cancers. This question has opened an active area of investigation, with recruitment-based assays now developed to more specifically interrogate the relationship

between SWI/SNF and Polycomb group complexes in cancer (Stanton et al. 2017, Kadoch et al. 2017).

ARID1A and ARID1B

The *ARID1A* and *ARID1B* genes encode large (250-kda) subunits of SWI/SNF chromatin remodeling complexes that are highly evolutionarily conserved, present as single genes *Swi1* in yeast and *Osa* in *Drosophila* (Wu and Roberts 2013). *ARID1A* and *ARID1B* are 60% identical across the length of the gene and encode homologous, mutually exclusive subunits that occupy the same position in SWI/SNF complexes (Wang et al. 2004, **Figure 1-2**). They are named for the "ARID" AT-rich DNA interaction domain, which binds DNA in a sequence non-specific manner *in vitro* (Wilsker et al. 2004). They have been suggested to function as targeting subunits of SWI/SNF complexes, although their precise functions are unknown (Nie et al. 2000, Wilson and Roberts 2011). Functional distinction between ARID1A and ARID1B has been demonstrated in a pre-osteoblast cell line model, where ARID1A-depleted cells fail to undergo cell-cycle arrest upon induction, while ARID1B-depleted cells are normal (Nagl et al. 2005).

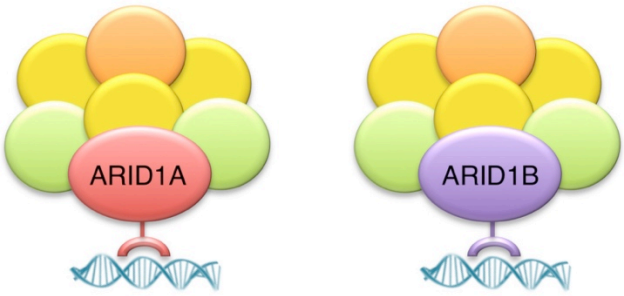


Figure 1-2: ARID1A and ARID1B-containing SWI/SNF complexes. ARID1A and ARID1B are mutually exclusive SWI/SNF subunits with sequence non-specific DNA binding activity.

ARID1A is the subunit of SWI/SNF complexes that is most frequently mutated in human cancer (Shain et al. 2013, Kadoch et al. 2013, Masliah-Planchon et al. 2015). The vast majority of cancer-associated mutations in *ARID1A* are inactivating (>97%), with nonsense or frameshift mutations detected across the length of the gene (Wu and Roberts 2013). *ARID1A* is recurrently mutated in a broad array of tumor types including 45.2% of endometrioid and clear-cell ovarian, 18.7% of gastric, 18.6% of bladder, 13.7% of hepatocellular, 9.4% of colorectal, 11.5% of melanoma, 8.2% of lung, 3.6% of pancreatic and 2.5% of breast cancers (Kadoch et al. 2013). In contrast, *ARID1B* is infrequently mutated, except in childhood neuroblastomas (Masliah-Planchon et al. 2015). Synthetic lethality-based screening has identified *ARID1B* as the number one dependency in *ARID1A*-mutant cancers (Helming et al. 2014a, Helming et al. 2014b). While this has established an essential role for ARID1B-containing residual complexes, the mechanism underlying synthetic lethality has been poorly understood.

Molecular genetics of colon cancer

Colon cancer is one of the three leading causes of cancer-related death in the United States (American Cancer Society 2017). Prior to cancer genome sequencing, genes involved in colon tumorigenesis were discovered through molecular genetic studies of colon cancer predisposition syndromes such as familial adenomatous polyposis (FAP) and hereditary nonpolyposis colon cancer (HNPCC, Taketo and Edelmann 2009). Colon cancer initiation and progression are proposed to follow a multi-stage genetic model, where a driver event is required for tumor initiation and each stage of tumor progression (Fearon and Vogelstein 1990, Kinzler and Vogelstein 1996, Vogelstein et al. 2013, **Figure 1-3**). Driver events are usually genetic

mutations in oncogenes and tumor suppressor genes, each favored at a distinct stage of tumorigenesis. Driver events usually affect pathways of particular importance in the disease, which in colon cancer include Wnt, TGF-β, EGFR, PI3K, and p53. Five to seven driver events are estimated to be required for progression to invasive cancer.

Colon cancers arise from adenomatous polyps, which are precursor lesions characterized by dysplastic morphology and altered differentiation of epithelial cells (Fearon 2011). Inactivation of APC, a major component of the Wnt signaling pathway, occurs in 70-80% of these polyps and is thought to be a critical early event. APC inactivation mimics constitutive activation of Wnt ligand-mediated signaling, as it allows β-catenin to translocate to the nucleus and activate transcription of its target genes. Mutations affecting TGF-β, EGFR, PI3K, and p53 also occur at high frequencies at particular stages of colon cancer development.

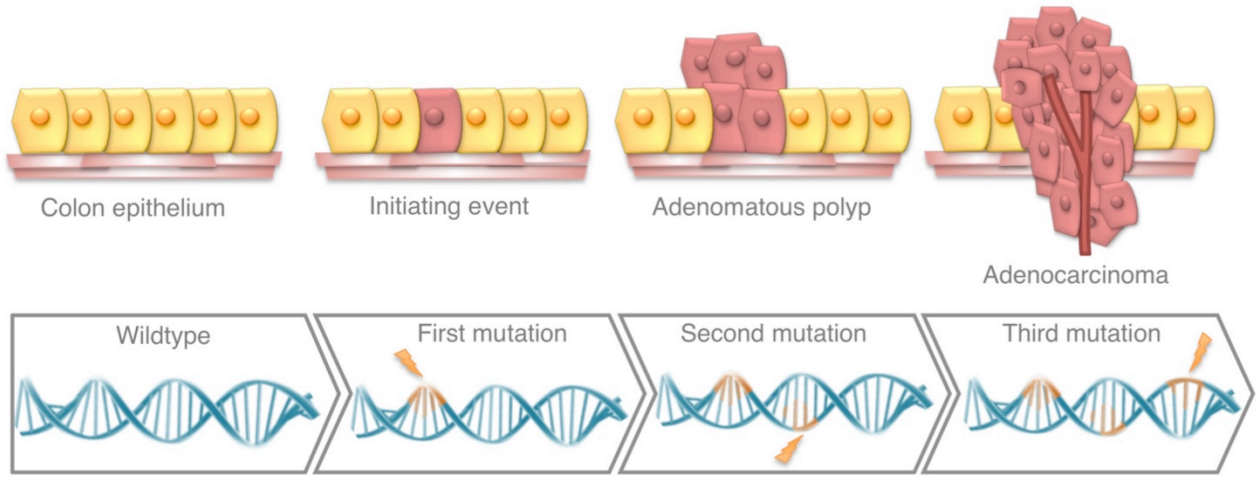


Figure 1-3: Multi-stage genetic model of colon cancer pathogenesis. Tumor initiation and progression occur in distinct stages; stage transitions are consequences of mutations in driver genes.

Sequencing efforts led by The Cancer Genome Atlas Consortium to comprehensively characterize genomic changes in human colon cancer have identified 24 significantly mutated genes (Cancer Genome Atlas Network, 2012). *ARID1A* is the third most significantly mutated gene across human colon cancers, following *APC* and *P53,* showing a disproportionally high number of inactivating (frameshift and nonsense) mutations. *ARID1A* mutations occur in 37% of colon cancers that exhibit microsatellite instability (MSI) and 5% of others. An independent sequencing study has validated this finding, showing *ARID1A* mutations in 39% of MSI-type colon cancers (Cajuso et al. 2014). MSI tumors are associated with defects in mismatch repair, which causes a large number of small insertions and deletions at repetitive areas in the genome; a large number of mutations in MSI tumors are found at a G mononucleotide tract in the coding region of *ARID1A*, although mutations are also found across the length of the gene. Several immunohistochemical studies have also evaluated ARID1A protein levels in patient tumor samples, identifying loss of ARID1A protein and validating the specific association of *ARID1A* mutations with colon cancers of the MSI type (Chou et al. 2014, Ye et al. 2014, Wei et al. 2014).

Recent stratification of human colon cancers by the Colorectal Cancer Subtyping Consortium (CRCSC, Guinney et al. 2015) has defined four consensus molecular subtypes (CMSs) with distinguishing features: CMS1 (microsatellite instability immune, 14%), hypermutated, microsatellite unstable and strong immune activation; CMS2 (canonical, 37%), epithelial, marked WNT and MYC signaling activation; CMS3 (metabolic, 13%), epithelial and evident metabolic dysregulation; and CMS4 (mesenchymal, 23%), prominent transforming growth factor–activation, stromal invasion and angiogenesis. Mutations in *ARID1A* are highly enriched in CMS1 cancers ($p = 7.10 \times 10^{-10}$); notably, mutations in *APC* are under-enriched in this class of cancers relative to others.

Enhancers in the mammalian genome

Enhancers were originally detected as genetic elements that increased transcription from promoters on the same molecule of DNA, able to act over large distances in a manner that had little precedent in prokaryotes (Khoury and Gruss 1983, Dynan 1989). As evaluation of primary nucleotide sequences failed to reveal a consistent pattern, enhancers were operationally classified by distance from the site of transcription initiation and biological demonstration of increased transcription. As advancements in chromatin immunoprecipitation (ChIP)-based methods enabled epigenomic profiling (Kim and Ren 2006), enhancers were found to be associated with distinct and predictive chromatin signatures. Chromatin signatures could distinguish enhancers from promoters and were predictive of activity, thus enabling functional classification and genome-wide mapping (Heintzman 2007, Visel 2009).

Active enhancers are devoid of nucleosomes such that DNA is accessible, and nucleosomes in the vicinity typically contain histones with characteristic modifications – histone H3 lysine 4 monomethylation (H3K4me1) and H3K27 acetylation (H3K27ac, Shlyueva et al. 2014). Chromatin signature-based enhancer mapping in different cells and tissue types has implicated enhancers as the most variable class of regulatory element, of primary importance in cell-type-specific gene expression (Heintzman 2009, Hnisz et al. 2013). Understanding enhancer function has since become an area of renewed interest, with broad implications suspected in development and in disease (Long et al. 2016, Sur and Taipale 2016, Hnisz et al. 2017). As enhancers contain transcription factor binding sites, co-activator roles for SWI/SNF complexes at enhancers have been explored and are described in various processes including interferon-mediated gene induction (Ni et al. 2008), lineage commitment (Alexander et al. 2015), and differentiation (Hu et al. 2011, Yu et al. 2013, Bossen et al. 2015).

Chapter 2:

ARID1A inactivation drives invasive colon adenocarcinoma in mice

Abstract

We investigate the tumor suppressor role of ARID1A *in vivo* utilizing an MX1-Cre *Arid1a*$^{fl/fl}$ mouse model, in which ARID1A is inactivated in a sporadic, interferon-responsive manner across various tissues. We find that these mice develop invasive colon adenocarcinoma with remarkable resemblance to human colon cancer. Villin-Cre^{ERT2} *Arid1a*$^{fl/fl}$ mice, with *Arid1a* excision restricted to the intestinal epithelium, also develop invasive colon adenocarcinoma with gross and histologic features indistinguishable from MX1-Cre *Arid1a*$^{fl/fl}$ mice. *ARID1A* is the third most significantly mutated gene in human colon cancer, with mutations enriched in cancers of the microsatellite-instable type (MSI, 37-39% frequency). Tumors in *Arid1a*$^{fl/fl}$ mice show mucinous differentiation and lymphocytic infiltration, features associated particularly with MSI-type colon cancer. These mice establish ARID1A as a tumor suppressor in the colonic epithelium, and provide a novel system for further investigation, with high relevance to the human disease. Tumorigenesis in widely utilized mouse models of intestinal cancer occurs predominantly in the small intestine, with tumors developing in a short period of time and advancing past the adenoma stage only on rare occasion. Tumorigenesis in *Arid1a*$^{fl/fl}$ mice more accurately reflects human disease, showing exclusivity in origin to the colon rather than the small intestine, longer latency, and aggressive invasion of local tissues. *Arid1a*$^{fl/fl}$ mice thus establish a sound model for investigation of various factors that might contribute to colon cancer initiation, progression, and response to therapy.

Contributions

The following individuals contributed to the work described in this chapter:

Radhika Mathur[1,2], Adrianna K. San Roman[1,3], Boris G. Wilson[2], Agoston T. Agoston[4], Ramesh A. Shivdasani[3,5], and Charles W. M. Roberts[2,6]

[1]Program in Biological & Biomedical Sciences, Harvard Medical School, Boston MA, 02215, USA

[2]Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, MA 02215, USA

[3]Department of Medical Oncology and Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, Boston, MA 02215, USA

[4]Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA

[5]Departments of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA

[6]Department of Oncology, St. Jude Children's Research Hospital, Memphis, TN 38105, USA

R.M., B.G.W., R.A.S, and C.W.M.R conceived experiments and study design. Experiments were performed by R.M., A.K.S.R., and B.G.W. Histopathological analysis was conducted by A.T.A. All authors contributed to data analysis and interpretation. R.M., R.A.S., and C.W.M.R. wrote the manuscript with input from all authors.

Acknowledgements

We thank S.H. Orkin for mentorship, and all members of the Roberts and Orkin labs for discussion; Jeffrey Haswell, Haley Manchester, Weishan Wang, Kelly Sullivan, Ev Tzvetkov, Tam Pham, Emma Troisi for assistance with management of the mouse colony; and Sylvie Robine for providing Villin-Cre^{ER-T2} transgenic mice. We also thank the Dana-Farber/Harvard Cancer Center cores in Rodent Histopathology and Specialized Histopathology, which are supported in part by a NCI Cancer Center Support Grant P30CA06516.

This work was supported by US National Institutes of Health grants R01CA172152 (C.W.M.R.) and R01DK081113 (R.A.S), by a Claudia Adams Barr grant (C.W.M.R.), and by an Innovation Award from Alex's Lemonade Stand (C.W.M.R). R.M. and A.K.S.R. were supported by the US National Institutes of Health predoctoral fellowships (1F31CA199994 and 1F31CA180784). The Cure AT/RT Now foundation, the Avalanna Fund, the Garrett B. Smith Foundation, Miles for Mary (C.W.M.R.), and the Lind Family (R.A.S.) provided additional support.

Citation of published work

All data presented in this chapter are published in the article:

Mathur, R., Alver, B.H., San Roman, A.K., Wilson, B.G., Wang, X., Agoston, A.T., Park, P.J., Shivdasani, R.A., and Roberts, C.W.M. (2017) ARID1A loss impairs enhancer-mediated gene regulation and drives colon cancer in mice. *Nature Genetics*, *49*(2), 296–302. http://doi.org/10.1038/ng.3744.

ARID1A is the subunit of SWI/SNF chromatin remodeling complexes that is most frequently mutated in human cancer (Shain and Pollock 2013, Kadoch et al. 2013). As inactivating *ARID1A* mutations occur in a broad spectrum of human cancers, we generated a mouse model to identify tissues where ARID1A might function as a tumor suppressor *in vivo*. We utilized the MX1 interferon-responsive promoter (Kuhn 1995) to inactivate ARID1A in a sporadic manner in cells across many different tissues of *Arid1a*$^{fl/fl}$ mice (Gao et al. 2008). MX1-Cre *Arid1a*$^{fl/fl}$ mice injected with synthetic interferon Poly I:C achieved sporadic inactivation of both *Arid1a* alleles across many tissues (**Figure 2-1A)**. Mice were initially healthy following induction of Cre activity, but developed emaciation and rectal prolapse, requiring euthanasia at a median of 296 days **(Figure 2-1B, Table 2-1**). Upon dissection, we identified nodular and polypoid tumors in the colons of several of these mice. Tumors were often present at multiple non-contiguous sites of the colon, including the cecum and the rectum **(Figure 2-1C)**, but were not identified in the small intestine. Tumor histology was consistent with invasive colon adenocarcinoma **(Figure 2-2A)**, a malignant neoplasm derived from glandular colonic epithelium (Hamilton et al. 2010). ARID1A protein was lost in few (<10%) crypts of the colon epithelium, as detected by immunohistochemistry (IHC), but was consistently absent in all tumor cells **(Figure 2-2B)**. Tumors were marked by prominent mucinous differentiation **(Figure 2-2C)** and the presence of tumor-infiltrating lymphocytes **(Figure 2-2D)**, features associated particularly with human colon cancers of the MSI type (Greenson et al. 2009). As *ARID1A* is the third-most significantly mutated gene in human colon cancer, with the highest frequency of mutations in cancers of the MSI type (37-39%, Cancer Genome Atlas Network 2012, Cajuso 2014), these findings are highly relevant to human disease.

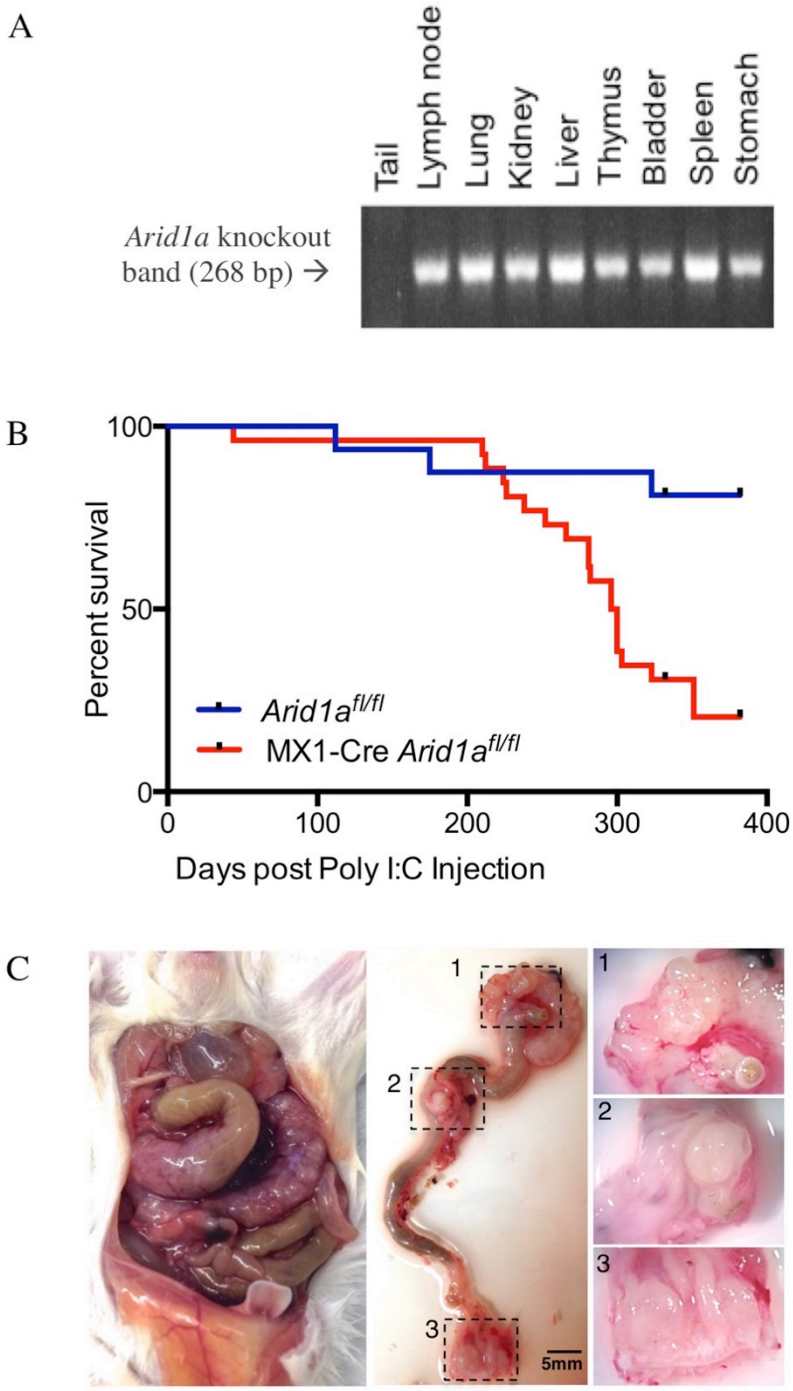


Figure 2-1: Characterization of MX1-Cre *Arid1a*$^{fl/fl}$ mice. (A) Excision of *Arid1a* in many tissues detected by PCR 1 week following injection of Poly I:C; (B) Survival of MX1-Cre *Arid1a*$^{fl/fl}$ mice (n=26) and control *Arid1a*$^{fl/fl}$ mice (n=16) following injection of Poly I:C; (C) MX1-Cre *Arid1a*$^{fl/fl}$ mouse with tumors in the (1) cecum, (2) mid-colon, and (3) rectum.

Table 2-1: Survival time (days), reason of death, and results of histopathological analysis for all MX1-Cre *Arid1a*$^{fl/fl}$ mice.

Mouse ID	Sex	Mx-Cre	*Arid1a*	Survival	Reason for Death	Colon Histopathological Analysis
4583/4812	M	+	fl/fl	44	Infection	Severe infection
4851	M	+	fl/fl	210	Rectal prolapse	Invasive colon adenocarcinoma
4669	F	+	fl/fl	212	Rectal prolapse/Emaciation	(Not analyzed)
4979	M	+	fl/fl	224	Found dead	(Not analyzed)
4587	M	+	fl/fl	226	Emaciation	Invasive colon adenocarcinoma
4589/4814	F	+	fl/fl	238	Emaciation	(Not analyzed)
4848/5396	M	+	fl/fl	252	Rectal prolapse/Severe rectal bleeding	Invasive colon adenocarcinoma
4667	M	+	fl/fl	266	Rectal prolapse/Severe rectal bleeding	Invasive colon adenocarcinoma
4806	F	+	fl/fl	281	Emaciation	Invasive colon adenocarcinoma
4981	F	+	fl/fl	281	Severe rectal bleeding/Abnormal Gait	Colon carcinoma in situ
4594	F	+	fl/fl	282	Emaciation	Invasive colon adenocarcinoma
4855	F	+	fl/fl	296	Rectal prolapse/Emaciation	Adenoma, low-grade dysplasia
4916/5401	F	+	fl/fl	296	Emaciation	Lymphoma
4849	M	+	fl/fl	300	Rectal prolapse/Emaciation	Invasive colon adenocarcinoma
4853	F	+	fl/fl	300	Emaciation	Invasive colon adenocarcinoma
4980	F	+	fl/fl	300	Emaciation	Invasive colon adenocarcinoma
4914/5399	M	+	fl/fl	303	Emaciation	Invasive colon adenocarcinoma
4588/5395	M	+	fl/fl	323	Rectal Prolapse/Hunched	(Not analyzed)
4591	M	+	fl/fl	351	Rectal prolapse/Hunched/Abnormal Gait	Invasive colon adenocarcinoma
4763	F	+	fl/fl	323+	N/A	N/A
4807/5182	F	+	fl/fl	323+	N/A	N/A
4808	F	+	fl/fl	323+	N/A	N/A
4809	F	+	fl/fl	323+	N/A	N/A
4915	M	+	fl/fl	323+	N/A	N/A
4584	F	+	fl/fl	373+	N/A	N/A
4585	F	+	fl/fl	373+	N/A	N/A

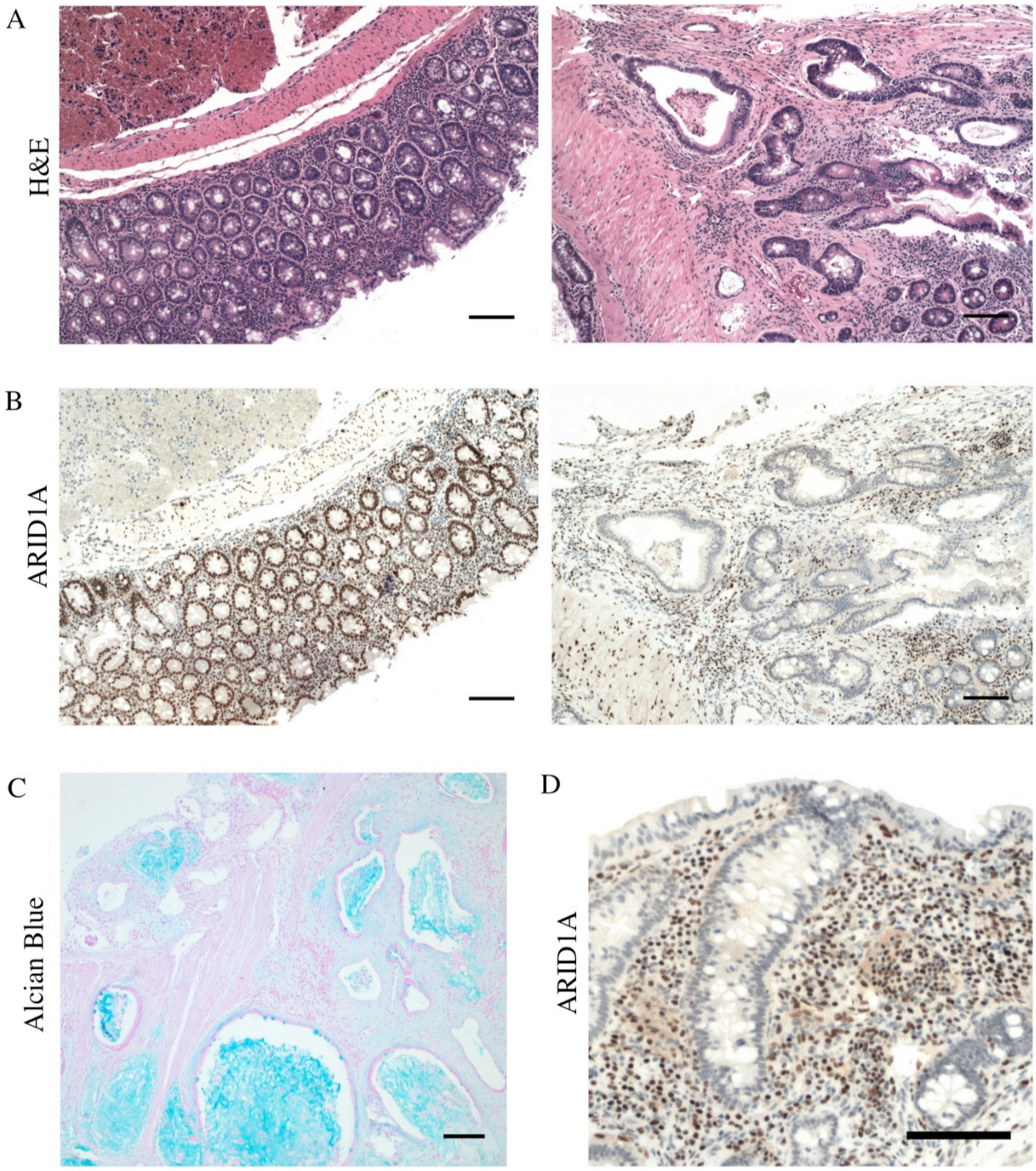


Figure 2-2: Histopathological examination of MX1-Cre *Arid1a*$^{fl/fl}$ mice. (A) H&E staining on normal colon epithelium (left) and tumor (right) tissue sections; (B) ARID1A immunohistochemistry (IHC) on above tissue sections; (C) Alcian blue staining; (D) ARID1A IHC on tumor section showing lymphocytic infiltrate. Scale bars in A-D, 100um.

To determine if colon tumorigenesis was driven by epithelial cell-intrinsic ARID1A deficiency, we utilized the Villin-Cre$^{ER\text{-}T2}$ Tamoxifen-inducible promoter to inactivate ARID1A specifically in intestinal epithelial cells (Marjou et al. 2004). IHC showed that loss of ARID1A protein was restricted to the intestinal epithelium of Villin-Cre$^{ER\text{-}T2}$ *Arid1a*$^{fl/fl}$ mice, where it occurred with nearly complete efficiency **(Figure 2-3A).** Villin-Cre$^{ER\text{-}T2}$ *Arid1a*$^{fl/fl}$ mice developed colon adenocarcinomas with indistinguishable features from those in MX1-Cre *Arid1a*$^{fl/fl}$ mice **(Figure 2-3B)**. As with MX1-Cre *Arid1a*$^{fl/fl}$ mice, tumors showed invasion into the submucosa **(Figure 2-4A)**, ARID1A deficiency **(Figure 2-4B)**, prominent mucinous differentiation **(Figure 2-4C)**, and presence of tumor-infiltrating lymphocytes **(Figure 2-4D).** These mice thus establish ARID1A as a *bona fide* tumor suppressor in the mouse colonic epithelium.

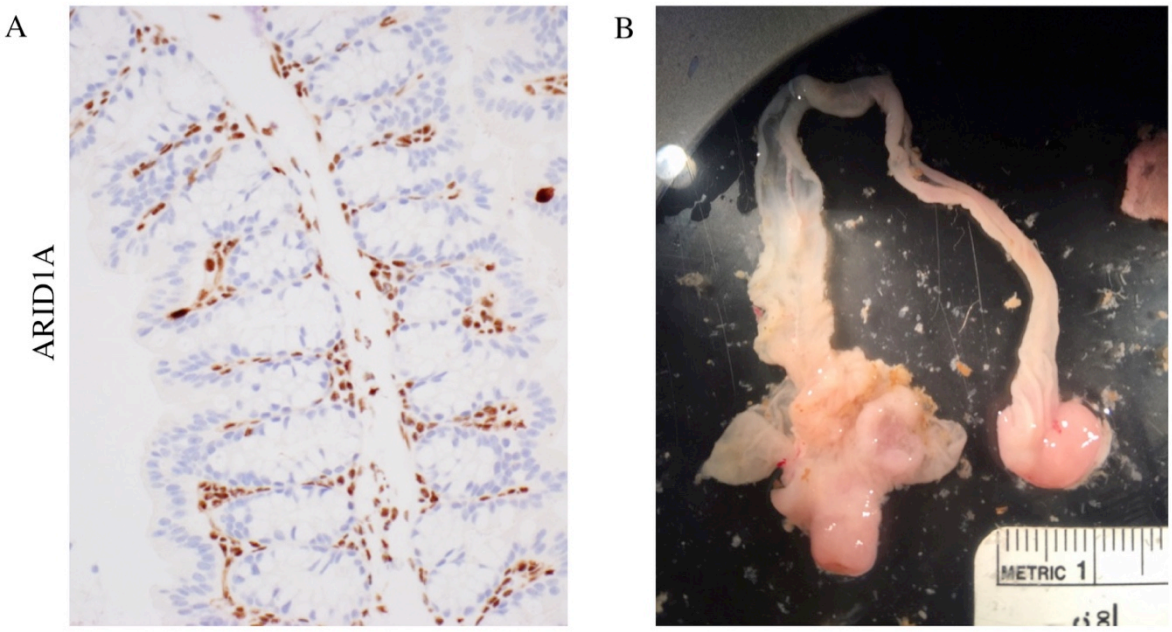


Figure 2-3: Characterization of Villin-Cre$^{ER\text{-}T2}$ *Arid1a*$^{fl/fl}$ mice. (A) ARID1A IHC of normal colon showing ARID1A loss in intestinal epithelial cells, 40X; (B) Colons with grossly visible tumors in cecum and rectum.

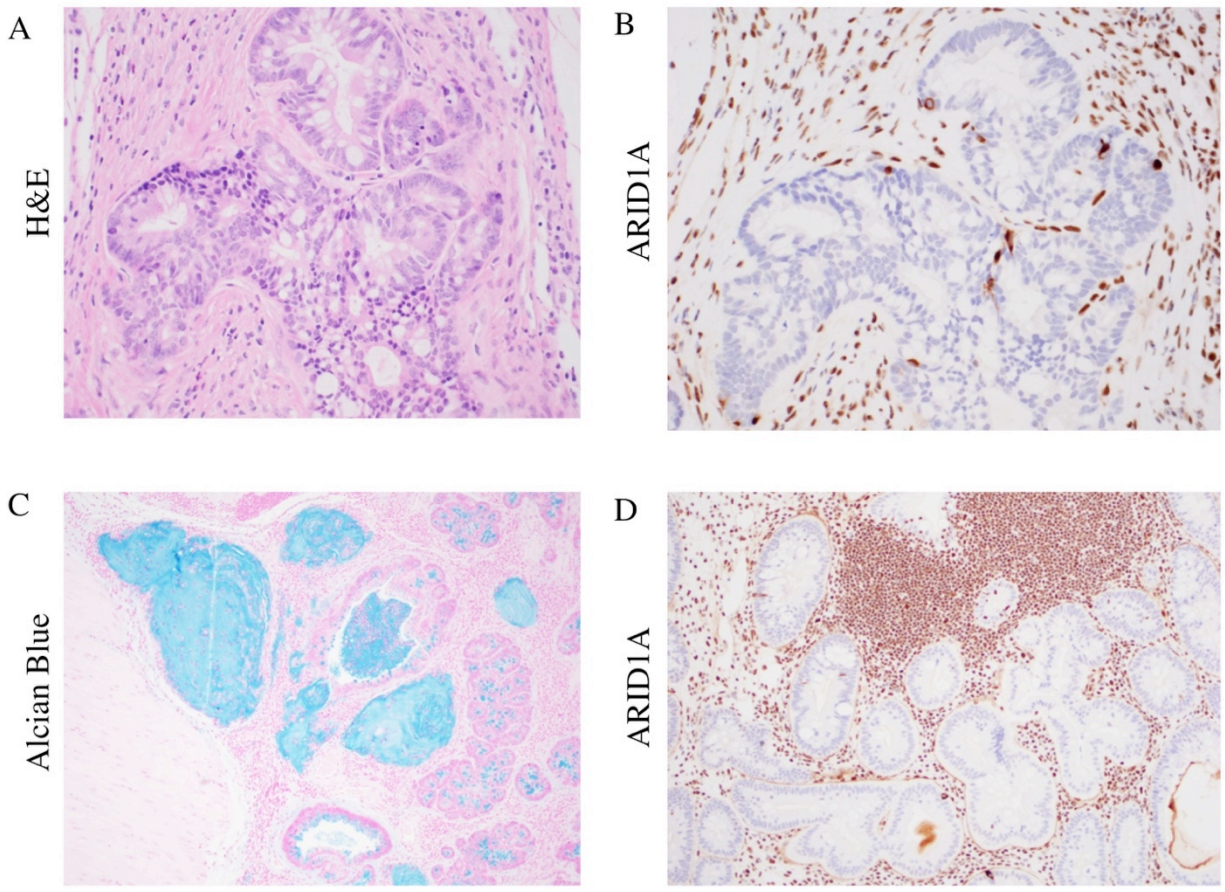


Figure 2-4: Histopathological examination of Villin-Cre^{ER-T2} *Arid1a*$^{fl/fl}$ mice. (A) H&E staining of tumor section showing invasive colon adenocarcinoma, 40X; (B) ARID1A IHC of section in (A); (C) Alcian Blue staining; (D) ARID1A IHC showing presence of lymphocytic infiltrate, 20X.

Arid1a$^{fl/fl}$ mice thus establish a new model for investigation of the mechanism underlying tumor suppression by ARID1A. This model is directly relevant to human disease, showing histological features associated with the particular type of human colon cancer enriched for mutations in *ARID1A*. Genes involved in colon tumorigenesis, such as *APC*, were discovered through studies of hereditary cancer predisposition syndromes; mice with mutations in these

genes are commonly used to model intestinal tumorigenesis (Taketo and Edelmann 2009). However, there are critical drawbacks in relating these models to human disease (McCart et al. 2008). Tumor formation in these mice is favored in the small intestine, while human disease occurs predominantly in the colon. Tumors in these mice also show little to no invasion into the submucosa, another key feature of the human disease. Tumors in *Arid1a*$^{fl/fl}$ mice, in contrast, show exclusive origin to the colon rather than the small intestine and are marked by aggressive local tissue invasion. *Arid1a*$^{fl/fl}$ mice thus provide a substantial improvement in mouse models for intestinal cancer, with broad utility in the study of tumor initiation, progression to invasive cancer, and response to therapy.

Methods

Mouse colony

All experiments were performed with strict adherence to our IACUC-approved Animal Experimentation Protocol #12-017 and guidelines of the Dana-Farber Animal Resource Facility (ARF). Mice were monitored for health by the ARF veterinarian staff and euthanized upon instruction.

MX1-Cre Arid1a$^{fl/fl}$ mice

MX1-Cre mice purchased from Jackson Labs (Stock number 003556) were bred to *Arid1a*$^{fl/fl}$ mice obtained from Dr. Zhong Wang (Gao et al. 2008). 6-8 week old MX1-Cre *Arid1a*$^{fl/fl}$ and littermate control *Arid1a*$^{fl/fl}$ mice were administered poly I:C (polyinosinic-polycytidylic acid, Invivogen tlrl-pic) via intraperitoneal injection at 25ug/g every other day for 7 days. Excision of *Arid1a* was evaluated in mice 1-week post injection by PCR of DNA harvested from mouse tissues using primers flanking the floxed exon (5'-GTAATGGGAAAGCGACTACTGGAG-3' and 5'-TGTTCATTTTTGTGGCGGGAG-3'). Whole mouse necroscopies were conducted on the first cohort of mice at the Dana-Farber/Harvard Cancer Center Rodent Pathology core. Sample size for survival analysis was calculated using estimated effect size from this cohort of mice; no animals were excluded. Formalin-fixed intestines were processed, sectioned, and stained at the DFCI Specialized Histopathology Core.

Villin-Cre^{ER-T2} Arid1a$^{fl/fl}$ mice

5-7 week old Villin-Cre^{ER-T2} *Arid1a*$^{fl/fl}$ mice and littermate control *Arid1a*$^{fl/fl}$ mice were administered 1mg Tamoxifen (T5648 SIGMA) dissolved in sunflower oil (S5007 SIGMA) via

intraperitoneal injection for 5 consecutive days, and were monitored for health. Formalin-fixed intestines were processed, sectioned, and stained at the DFCI Specialized Histopathology Core.

Chapter 3:

ARID1A inactivation drives colon tumorigenesis independent of Wnt signaling

Abstract

Colon cancer initiation and progression are generally thought to occur in distinct stages, where each stage transition is a consequence of a genetic mutation or other driver event. As we identified late-stage invasive colon cancers in *Arid1a*$^{fl/fl}$ mice, we sought to identify events that cooperate with ARID1A inactivation in driving tumorigenesis. We performed whole-exome sequencing, but did not identify cooperating mutations in genes associated with human colon cancer. Tumors also did not show aberrant activation of Wnt signaling, an initiating event in established genetic models of colon cancer. To directly investigate cooperation between ARID1A inactivation and Wnt signaling, we inducibly inactivated ARID1A in a Wnt-driven model of intestinal cancer. Remarkably, we found that ARID1A inactivation blocked tumor onset in these mice; the few tumors that did form had retained expression of ARID1A. These results reveal that ARID1A inactivation drives tumorigenesis by a mechanism that is independent of Wnt signaling and distinct from established genetic models of colon cancer. This is consistent with several studies of human colon cancer, which show under-enrichment of *APC* mutations and lack of nuclear β-catenin staining in the type of colon cancer enriched for mutations in *ARID1A*. These results also indicate that the function of ARID1A is required to facilitate tumorigenesis driven by Wnt signaling. This is consistent with a previously defined relationship between SWI/SNF complexes in regulating transcription of Wnt target genes and highlights the importance of context specificity in tumor suppression by SWI/SNF complexes.

Contributions

The following individuals contributed to the work described in this chapter:

Radhika Mathur[1,2], Adrianna K. San Roman[1,3], Agoston T. Agoston[4], Ramesh A. Shivdasani[3,5], and Charles W. M. Roberts[2,6]

[1]Program in Biological & Biomedical Sciences, Harvard Medical School, Boston MA, 02215, USA

[2]Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, MA 02215, USA

[3]Department of Medical Oncology and Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, Boston, MA 02215, USA

[4]Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA

[5]Departments of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA

[6]Department of Oncology, St. Jude Children's Research Hospital, Memphis, TN 38105, USA

R.M., R.A.S, and C.W.M.R conceived experiments and study design. Experiments were performed by R.M. and A.K.S.R. Histopathological analysis was conducted by A.T.A. All authors contributed to data analysis and interpretation. R.M., R.A.S., and C.W.M.R. wrote the manuscript with input from all authors.

Acknowledgements

We thank S.H. Orkin for mentorship, and all members of the Roberts and Orkin labs for discussion. We thank Jeffrey Haswell, Haley Manchester, Weishan Wang, Kelly Sullivan, Ev Tzvetkov, Tam Pham, Emma Troisi for assistance with management of the mouse colony. We thank Sylvie Robine for providing Villin-Cre^{ER-T2} transgenic mice. We also thank the Dana-Farber/Harvard Cancer Center cores in Rodent Histopathology and Specialized Histopathology, which are supported in part by a NCI Cancer Center Support Grant P30CA06516.

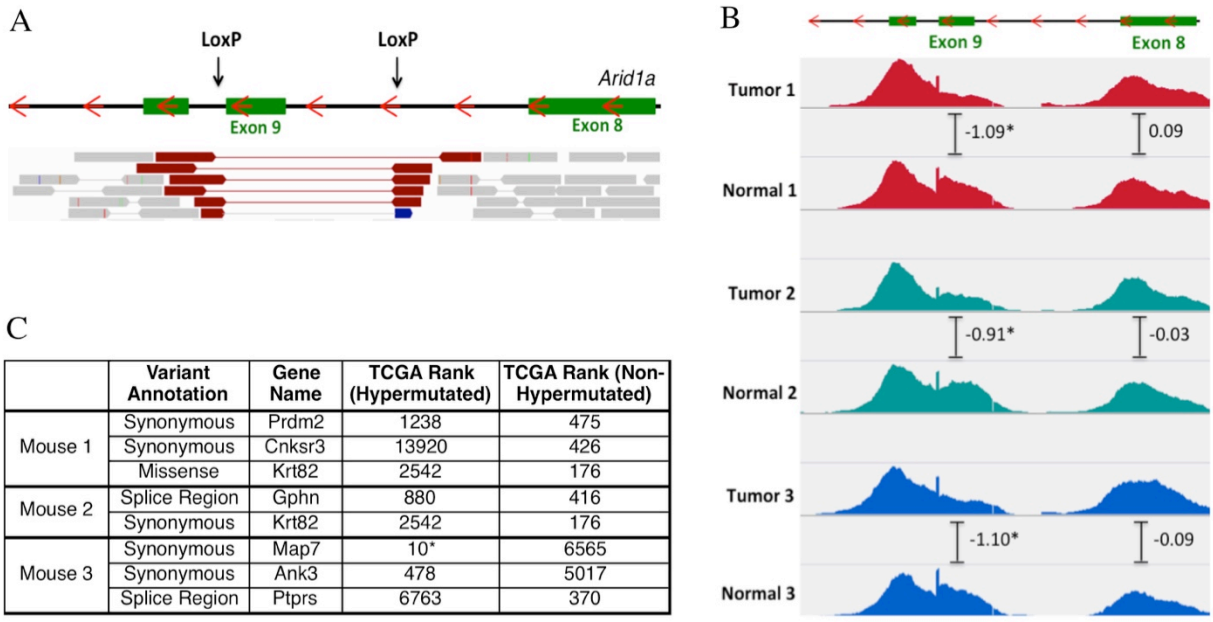
This work was supported by US National Institutes of Health grants R01CA172152 (C.W.M.R.) and R01DK081113 (R.A.S), by a Claudia Adams Barr grant (C.W.M.R.), and by an Innovation Award from Alex's Lemonade Stand (C.W.M.R). R.M. and A.K.S.R. were supported by the US National Institutes of Health predoctoral fellowships (1F31CA199994 and 1F31CA180784). The Cure AT/RT Now foundation, the Avalanna Fund, the Garrett B. Smith Foundation, Miles for Mary (C.W.M.R.), and the Lind Family (R.A.S.) provided additional support.

Citation of published work

All data presented in this chapter are published in the article:

Mathur, R., Alver, B.H., San Roman, A.K., Wilson, B.G., Wang, X., Agoston, A.T., Park, P.J., Shivdasani, R.A., and Roberts, C.W.M. (2017) ARID1A loss impairs enhancer-mediated gene regulation and drives colon cancer in mice. *Nature Genetics*, *49*(2), 296–302. http://doi.org/10.1038/ng.3744.

To identify potential cooperating events in tumorigenesis driven by ARID1A deficiency, we obtained whole-exome sequences of DNA isolated from tumor and matched normal tissue from three MX1-Cre *Arid1a*$^{fl/fl}$ mice. Variant analysis of the exome data confirmed excision of the floxed *Arid1a* exon (**Figure 3-1A, B**), but identified few non-synonymous mutations, none of which were in genes recurrently mutated in human colorectal cancer (as identified by The Cancer Genome Atlas, TCGA, Cancer Genome Atlas Network 2012, **Figure 3-1C**).



C

	Variant Annotation	Gene Name	TCGA Rank (Hypermutated)	TCGA Rank (Non-Hypermutated)
Mouse 1	Synonymous	Prdm2	1238	475
	Synonymous	Cnksr3	13920	426
	Missense	Krt82	2542	176
Mouse 2	Splice Region	Gphn	880	416
	Synonymous	Krt82	2542	176
Mouse 3	Synonymous	Map7	10*	6565
	Synonymous	Ank3	478	5017
	Splice Region	Ptprs	6763	370

Figure 3-1: Whole exome sequencing of tumors from MX1-Cre *Arid1a*$^{fl/fl}$ mice. (A) Alignment of paired-end exome sequencing reads at the *Arid1a* gene; (B) Quantification of exome-sequencing reads of matched tumor and normal tissues at *Arid1a* exon 9 (floxed) and *Arid1a* exon 8 (control); read coverage tracks are shown with $\log_2$ ratio (tumor reads/normal reads) with significant copy number variations denoted by *; (C) Genetic variants identified in each tumor relative to genes mutated in human colon cancer for hypermutated or non-hypermutated subtypes (Cancer Genome Atlas Network 2012); only genes ranking within top 500 in at least one category are shown; significantly mutated genes denoted by *.

APC inactivation is considered an early initiating event in genetic models of human colon cancer (Fearon and Vogelstein 1990, Fearon 2011). In the absence of Wnt ligand, APC normally ubiquitinates β-catenin and targets it for degradation. APC inactivation allows β-catenin to translocate to the nucleus in the absence of Wnt ligand, triggering constitutive activation of Wnt target genes. To determine if Wnt signaling was aberrantly activated in colon tumors from *Arid1a*$^{fl/fl}$ mice, we examined β-catenin localization by IHC. We found that β-catenin localized exclusively outside of the nucleus, indicating intact function of APC **(Figure 3-2)**.

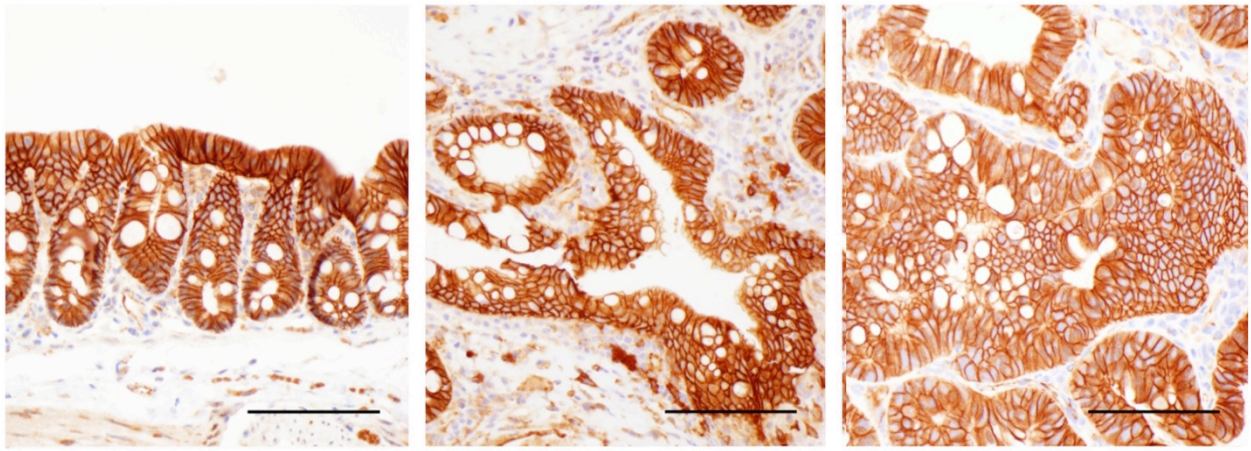


Figure 3-2: Tumors in *Arid1a*$^{fl/fl}$ mice do not show nuclear localization of β-catenin. β-catenin IHC is shown for wildtype mouse colon and tumor tissue sections from MX1-Cre *Arid1a*$^{fl/fl}$ and Villin-Cre^{ER-T2} *Arid1a*$^{fl/fl}$ mice. Scale bars, 100um.

To characterize further the relationship between ARID1A and Wnt signaling in colon tumorigenesis, we investigated the consequences of inducibly inactivating ARID1A in a mouse model of intestinal cancer driven by aberrant Wnt signaling. We obtained mice carrying the

germline Apc^{Min} mutation and generated Apc^{Min}:$Arid1a^{KO}$ (Apc^{Min}: Villin-Cre^{ER-T2} $Arid1a^{fl/fl}$) mice. As expected, Apc^{Min} mice developed a large number of non-invasive tumors, predominantly in the small intestine rather than the colon (McCart et al. 2008). Remarkably, we found that significantly fewer tumors developed in the intestines of Apc^{Min}:$Arid1a^{KO}$ mice (**Table 3-1**). Thus, ARID1A inactivation does not cooperate with aberrant Wnt signaling in driving tumorigenesis; rather, ARID1A inactivation antagonizes tumorigenesis driven by aberrant Wnt signaling.

Table 3-1: Tumor counts for small intestine and colon in Apc^{Min} and Apc^{Min}: $Arid1a^{KO}$ mice.

Genotype	Mouse ID	Tumor count		
		Small Intestine	Colon	Total
Apc^{Min}	5479	44	0	44
	5487	49	2	51
	5273	51	1	52
	5454	53	1	54
	5463	56	2	58
	5254/5394	63	6	69
	5480	79	7	86
	5469	100	3	103
	5459	103	4	107
	5462	102	5	107
	5336	149	7	156
	5255	158	4	162
	Mean	***83.9***	***3.5***	***87.5***
Apc^{Min}: $Arid1a^{KO}$	5460	2	0	2
	5456	4	0	4
	5450	7	0	7
	5272	8	1	9
	5467	7	2	9
	5468	10	2	12
	5457	14	3	17
	5446	18	1	19
	5253	15	13	28
	5458	30	2	32
	5464	54	5	59
	5260	69	3	72
	Mean	***19.8***	***2.7***	***22.5***

$p < 0.0001$ in unpaired two-tailed T-test for total intestinal tumor counts.

Histology showed that the few tumors that did arise in *Apc*^{Min}*:Arid1a*^{KO} mice were non-invasive adenomas, indistinguishable from those in *Apc*^{Min} mice **(Figure 3-3)**. These tumors had selectively retained expression of ARID1A and showed nuclear localization of β-catenin. These results reveal that ARID1A is required for tumorigenesis driven by aberrant Wnt signaling.

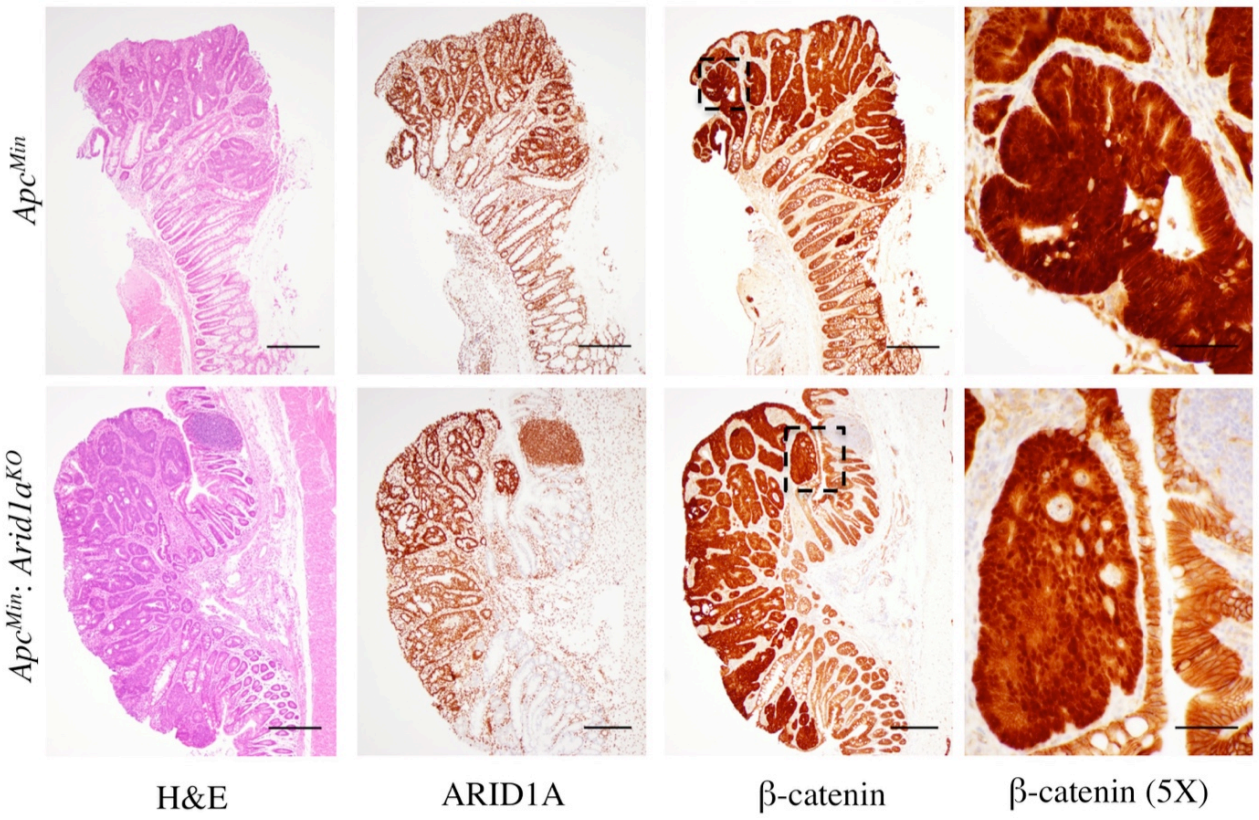


Figure 3-3: ARID1A is selectively retained in tumors driven by deregulated Wnt signaling. Colon adenomas in *Apc*^{Min} mice and *Apc*^{Min}*:Arid1a*^{KO} mice with H&E staining and immunohistochemistry for ARID1A and β-catenin, with magnification shown for marked tumor regions. Scale bars 250um (50um in magnification).

Collectively, these results indicate that ARID1A inactivation drives colon cancer via a novel pathway distinct from established multi-stage genetic models of colon cancer pathogenesis. The apparent absence of cooperating genetic mutations in invasive colon cancers in *Arid1a*$^{fl/fl}$ mice is consistent with an epigenetic mechanism of oncogenesis underlying these cancers. Certain highly aggressive human cancers driven by inactivation of SWI/SNF subunits have stable genomes, lacking cooperating mutations in cancer-associated genes (malignant rhabdoid tumor, Lee et al. 2012, and small cell carcinoma of the ovary, hypercalcemic type, Jelinic et al. 2014). While *ARID1A*-mutant human colon cancers are not characterized by stable genomes (rather by MSI), these results demonstrate that tumor formation driven by ARID1A inactivation does not require the cooperation of other mutations in cancer-associated genes.

Importantly, these results are consistent with data from human colon cancers. Despite the characterization of APC inactivation as a "gatekeeper" of human colon cancer (Fearon and Vogelstein 1990, Fearon 2011), normal patterns of β-catenin staining have long been noted in the subset of tumors with MSI (Jass et al. 1999). While *ARID1A* mutations are enriched in human colon cancers that show microsatellite instability (MSI), mutations in classical colon cancer genes including *APC, P53,* and *KRAS* are under-enriched in these cancers (Cancer Genome Atlas Network 2012). Recent molecular classification of human colorectal cancers by the Colorectal Cancer Subtyping Consortium (CRCSC) also shows that *APC* mutations occur with relatively low frequency in the subtype (consensus molecular subtype, CMS 1: MSI, Immune) enriched for mutations in *ARID1A* (Guinney et al. 2015). Our findings establish ARID1A as a critical tumor suppressor in this subtype of cancer and suggest that the mechanism by which these cancers develop is independent of established genetic models.

Our results further demonstrate that ARID1A is required to facilitate tumorigenesis driven by aberrant Wnt signaling – tumor onset in *Apc* mutant mice is blocked by ARID1A inactivation, and the few tumors that do form retain expression of ARID1A. Our laboratory has previously identified a role for SWI/SNF complexes in transcriptional activation of β-catenin targets (Mora-Blanco et. al 2013); however, SWI/SNF complexes were not anticipated to have a crucial role in mediating tumorigenesis driven by aberrant Wnt signaling. While therapeutic targeting of ARID1A in Wnt-driven colon cancers merits further exploration, these results also highlight the crucial importance of context specificity in consideration of normal functions of SWI/SNF complexes and their extensive roles in malignancy.

Methods

Whole exome sequencing

DNA from flash-frozen matched tumor and tail tissue from MX1-Cre *Arid1a*$^{fl/fl}$ mice was purified with the DNEasy Blood & Tissue Kit (Qiagen 69504) following histological confirmation of invasive adenocarcinoma. Samples were further processed and analyzed at the Dana-Farber Center for Cancer Computational Biology (CCCB). Target enrichment was performed using the SureSelectXT Mouse All Exon (Agilent Technologies) bait library. Samples were sequenced on the Next-Seq500 (Illumina) system using the PE-150 flowcell. Following sequencing and demultiplexing, sequencing reads were trimmed such that the lowest quartile of the phred-scaled Q-score was greater than 28; typical read lengths were approximately 140bp. The paired reads were aligned to the Ensembl GRCm38.75 genome using BWA-mem (Li and Durbin 2010, Li 2013) using default parameters. Following initial alignment, reads were further processed using GATK best-practices for WES data, including marking of duplicates, de novo realignment near putative indels, and base-quality score recalibration. Variant calling was performed with VarScan2 software (v2.4.1, Koboldt 2012), due to the paired (tumor/normal) design of the experiment. Somatic, germline, and loss-of-heterozygosity events are reported. Default parameters were used, requiring a minimum coverage depth of 8 reads for variant calls. In addition to the set of VarScan2 calls, MuTect software (Cibulskis 2013) was used to generate a second set of somatic point mutations. VarScan2 was also used for generation of putative copy-number variations between the matched samples. Following the recommendation of the documentation, CNV calls were filtered and finely smoothed/segmented using Bioconductor's DNACopy package, which implements the CBS algorithm. Additionally, a

second set of CNV calls was generated using CONTRA software (Li et al. 2012) with default parameters and specifying the targeted regions from the exome capture process.

Apc^{Min} and Apc^{Min}:$Arid1a^{KO}$ mice

Apc^{Min} mice purchased from Jackson labs (C57BL/6J-Apc-min/J; stock number 002020) were bred to Villin-$Cre^{ER\text{-}T2}$ *$Arid1a^{fl/fl}$* mice to generate *Apc^{Min}*:Villin-$Cre^{ER\text{-}T2}$ *$Arid1a^{fl/fl}$ (Apc^{Min}:$Arid1a^{KO}$)* mice and littermate control *Apc^{Min} $Arid1a^{fl/fl}$ (Apc^{Min})* mice. Mice aged 5-6 weeks were injected intraperitoneally with 1mg Tamoxifen (T5648 SIGMA) dissolved in sunflower oil (S5007 SIGMA) for 5 consecutive days. Age-matched *Apc^{Min}* and *Apc^{Min}:$Arid1a^{KO}$* mice were euthanized at 5-6 months. Blinded counts of intestinal tumors were obtained under a dissecting microscope. A two-tailed *t*-test was used to evaluate statistical significance, F-test to compare variances. Intestines were formalin-fixed and processed for histology and immunohistochemistry at the DFCI Specialized Histopathology Core. All experiments were performed with strict adherence to our IACUC-approved Animal Experimentation Protocol #12-017 and guidelines of the Dana-Farber Animal Resource Facility (ARF).

Chapter 4:

ARID1A inactivation impairs SWI/SNF targeting and control of enhancer activity

Abstract

ARID1A is implicated as a targeting subunit of SWI/SNF complexes as it has DNA-binding activity. Here, we seek to determine if altered targeting of SWI/SNF complexes may underlie tumor suppression by ARID1A. We establish a human cancer cell model to study the consequences of ARID1A inactivation on SWI/SNF targeting, noting morphological changes consistent with loss of epithelial cell character. We find that SWI/SNF binding occurs predominantly at active enhancers marked by H3K27ac and H3K4me1. We find the majority of enhancers lose SWI/SNF binding upon ARID1A inactivation, and subsequently lose activity, showing reduced levels of H3K27ac and downregulation of nearest genes. These enhancers are broadly implicated in control of developmental gene expression programs mediated by transcription factors. These results elucidate a crucial role for SWI/SNF complexes in control of enhancer activity, which has broad relevance for the normal functions of SWI/SNF complexes and their roles in malignancy. We validate these findings *in vivo* in the ARID1A-deficient mouse colonic epithelium, finding that defective SWI/SNF targeting and control of enhancer activity cause extensive dysregulation of gene expression. These results thus establish enhancer-mediated gene regulation as a principal tumor suppressor function of ARID1A.

Contributions

The following individuals contributed to the work described in this chapter:

Radhika Mathur[1,2], Burak Han Alver[3], Xiaofeng Wang[2], Peter J. Park[3,6], Ramesh A. Shivdasani[4,5], and Charles W. M. Roberts[2,6]

[1]Program in Biological & Biomedical Sciences, Harvard Medical School, Boston MA, 02215, USA

[2]Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, MA 02215, USA

[3]Department of Biomedical Informatics, Harvard Medical School, Boston, MA 02115, USA

[4]Department of Medical Oncology and Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, Boston, MA 02215, USA

[5]Departments of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA

[6]Department of Oncology, St. Jude Children's Research Hospital, Memphis, TN 38105, USA

R.M. and C.W.M.R conceived experiments and study design. Experiments were performed by R.M. and X.W. Computational and statistical analyses were performed by R.M. and B.H.A., with guidance from P.J.P. All authors contributed to data analysis and interpretation. R.M., R.A.S., and C.W.M.R. wrote the manuscript with input from all authors.

Acknowledgements

We thank S.H. Orkin for mentorship, and all members of the Roberts and Orkin labs for discussion. This work was supported by US National Institutes of Health grants R01CA172152 (C.W.M.R.) and R01DK081113 (R.A.S), by a Claudia Adams Barr grant (C.W.M.R.), and by an Innovation Award from Alex's Lemonade Stand (C.W.M.R). R.M. supported by the US National Institutes of Health predoctoral fellowships (1F31CA199994). X.W. was supported by the Pathway to Independence Award from the US National Institutes of Health (K99CA197640). The Cure AT/RT Now foundation, the Avalanna Fund, the Garrett B. Smith Foundation, Miles for Mary (C.W.M.R.), and the Lind Family (R.A.S.) provided additional support.

Citation of published work

All data presented in this chapter are published in the article:

Mathur, R., Alver, B.H., San Roman, A.K., Wilson, B.G., Wang, X., Agoston, A.T., Park, P.J., Shivdasani, R.A., and Roberts, C.W.M. (2017) ARID1A loss impairs enhancer-mediated gene regulation and drives colon cancer in mice. *Nature Genetics*, *49*(2), 296–302. http://doi.org/10.1038/ng.3744.

Accession codes

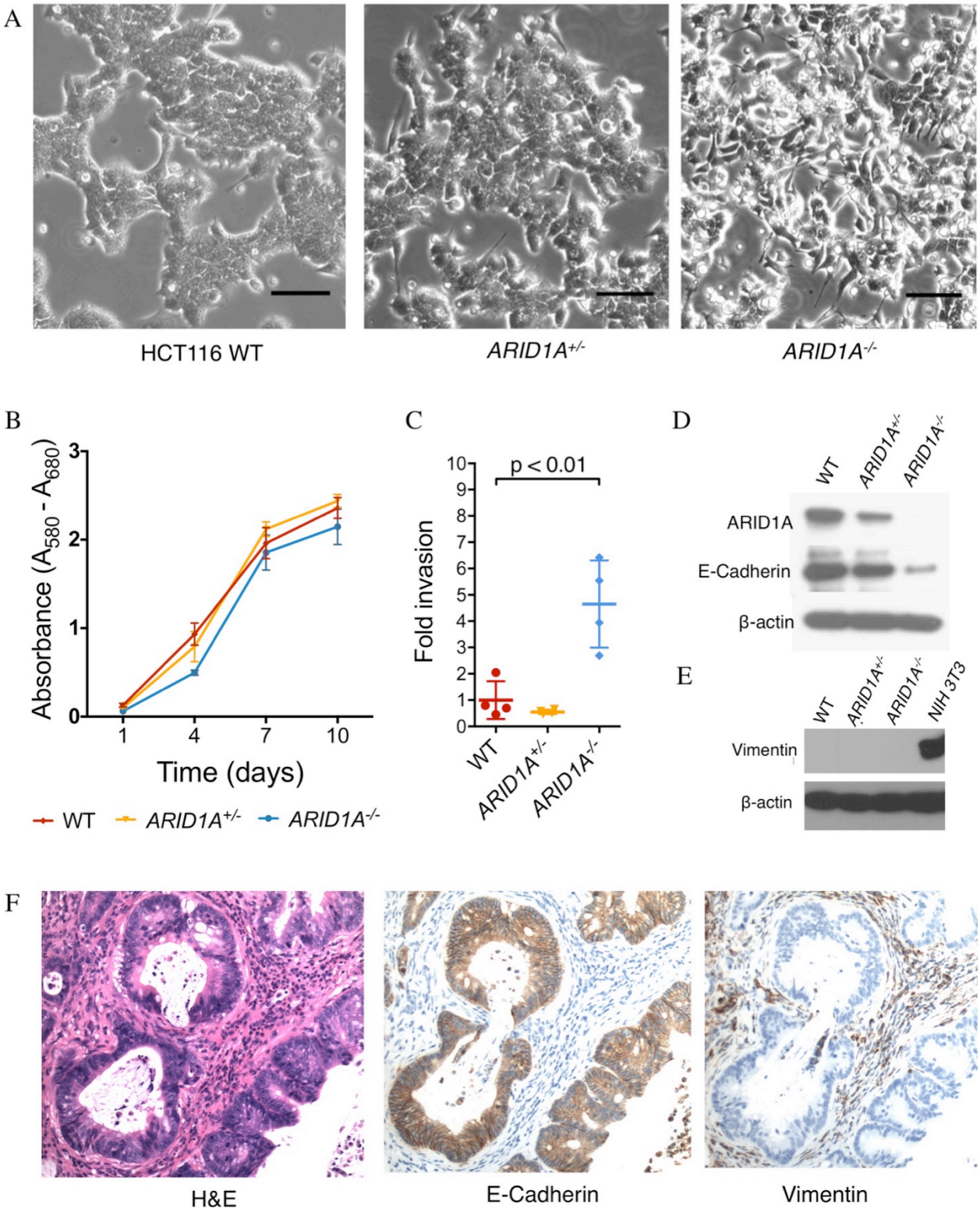
All sequencing data are deposited in the GEO database (http://www.ncbi.nlm.nih.gov/geo/) with accession number GSE71514.

To investigate the mechanism underlying tumor suppression by ARID1A, we sought to define the consequences of ARID1A inactivation on SWI/SNF function. ARID1A is implicated as a targeting subunit of SWI/SNF complexes as it contains the "ARID" AT-rich DNA interaction domain, which is capable of binding DNA in a sequence non-specific manner (Wilsker et al. 2004, Nie et al. 2000, Wilson and Roberts 2011). We therefore sought to determine whether ARID1A inactivation might alter the targeting of SWI/SNF complexes to chromatin, and hence, affect their function.

For this study, we utilized the HCT116 MSI human colon cancer cell line and isogenic lines with mono- (*ARID1A*$^{+/-}$) or bi- (*ARID1A*$^{-/-}$) allelic deletion of ARID1A. We noted that while parental *ARID1A*-wildtype (WT) cells grew in clustered colonies with tight cell-cell adhesion, *ARID1A*$^{-/-}$ cells spread across the culture dish with elongated, spindle-shaped morphologies and frequent filopodia **(Figure 4-1A)**. *ARID1A*$^{-/-}$ cells proliferated normally **(Figure 4-1B)**, but showed increased invasiveness **(Figure 4-1C)** and reduced expression of the cell adhesion protein E-Cadherin **(Figure 4-1D)**. We sought to determine if loss of epithelial cell characteristics was a consequence of epithelial-mesenchymal transition (EMT), but did not identify a molecular signature consistent with this regulated cell fate transition in *ARID1A*$^{-/-}$ cells **(Figure 4-1E)** or in tumors from *Arid1a*$^{fl/fl}$ mice **(Figure 4-1F)**. Notably, ARID1A depletion causes similar defects in cells derived from human gastric and hepatocellular carcinoma, which also frequently carry inactivating mutations in *ARID1A* (Yan 2014, He 2015).

Figure 4-1: ARID1A inactivation causes loss of epithelial cell character. (A) Live cell morphology of HCT116 *ARID1A* WT, *ARID1A*$^{+/-}$, and *ARID1A*$^{-/-}$ cells in culture; (B) Proliferation measured by MTT assay; (C) Invasion measured by Matrigel-chamber based assay; (D) Protein levels of ARID1A, E-Cadherin, and β-actin; (E) Protein levels of Vimentin and β-actin (NIH 3T3 fibroblast cells included for positive control); (F) Immunohistochemical staining of E-Cadherin and Vimentin in MX1-Cre *Arid1a*$^{fl/fl}$ tumor section.

Figure 4-1 (Continued)



To identify SWI/SNF binding sites in HCT116 WT and *ARID1A*$^{-/-}$ cells, we performed ChIP-Seq for two core subunits – SMARCA4 and SMARCC1. We also profiled histone modifications associated with gene regulatory elements – H3K4me3, H3K4me1, and H3K27ac. These clustered with corresponding HCT116 ChIP-Seq profiles generated by the ENCODE (Encycopedia of DNA Elements) Consortium **(Figure 4-2A**, ENCODE Project Consortium 2012). SMARCA4 and SMARCC1 binding were highly correlated and we considered sites enriched for both subunits as SWI/SNF binding sites. We found that in *ARID1A*$^{-/-}$ cells, a large majority of SWI/SNF binding sites (79.2%) showed loss of SMARCA4 and SMARCC1 binding **(Figure 4-2B)**, indicating a role for ARID1A in targeting SWI/SNF complexes to these sites.

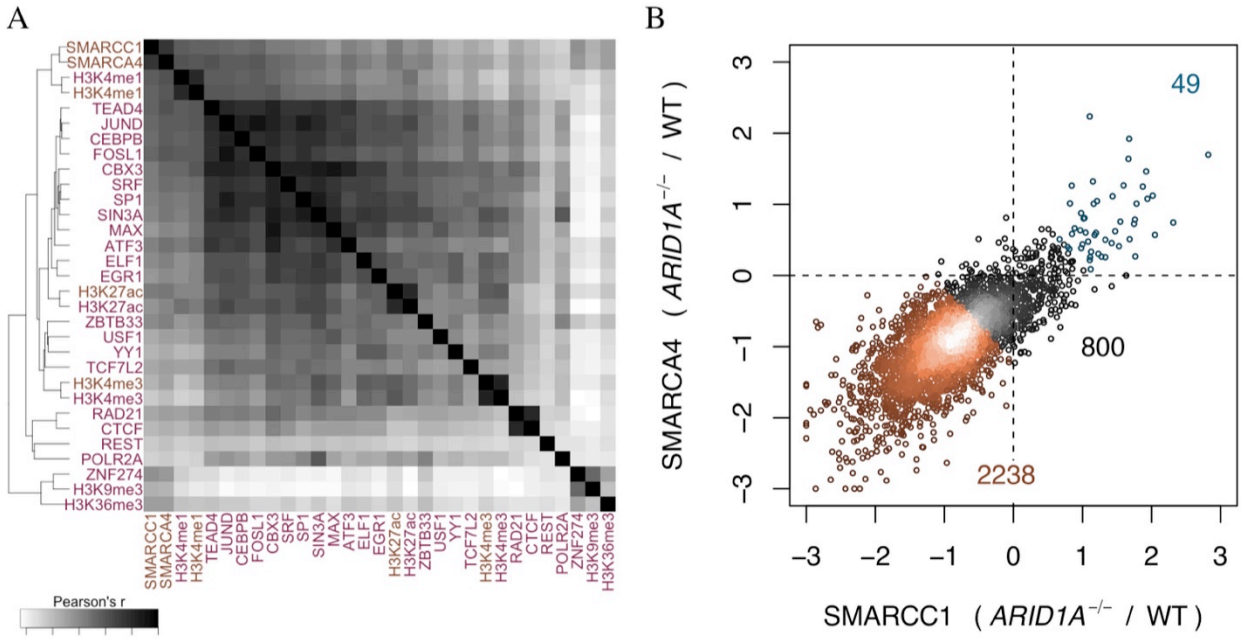


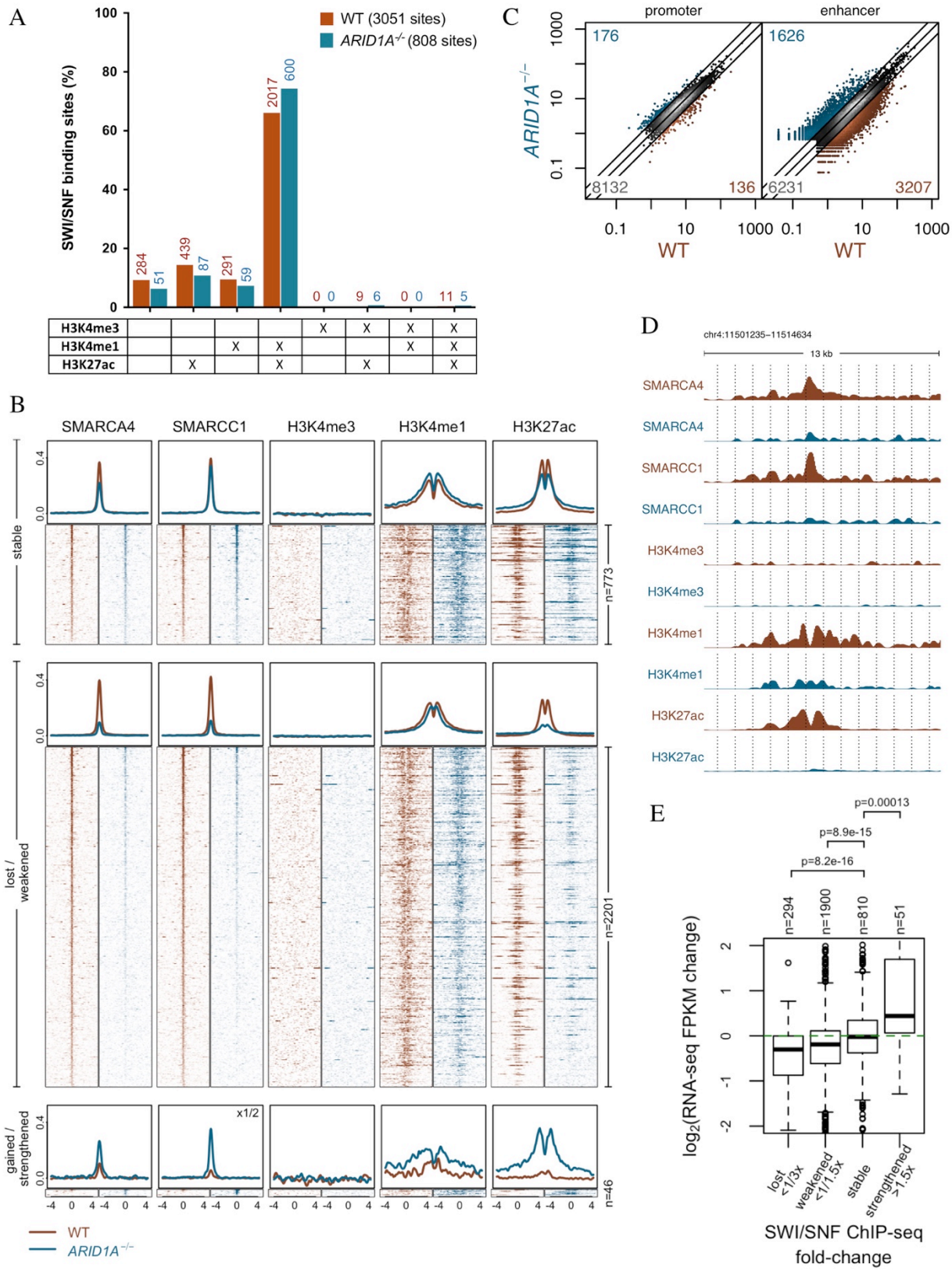
Figure 4-2: ARID1A inactivation impairs SWI/SNF targeting to chromatin. (A) Heatmap showing genome-wide correlations between all ChIP-Seq profiles in WT HCT116 cells (ENCODE datasets are colored pink); (B) Fold change (log_2) in SMARCA4 and SMARCC1 ChIP-Seq signals at SWI/SNF binding sites in *ARID1A*$^{-/-}$ cells relative to WT.

We found that SWI/SNF binding in both WT and *ARID1A*$^{-/-}$ cells was highly enriched at active enhancers, marked with H3K4me1 and H3K27ac **(Figure 4-3A)**. In contrast, SWI/SNF binding was limited at active promoters, defined as transcription start site (TSS)-overlapping regions enriched for H3K4me3. In TSS-distal regions, H3K27ac was diminished in *ARID1A*$^{-/-}$ cells at sites that lost SWI/SNF binding, and was increased at the few sites where SWI/SNF binding was gained **(Figure 4-3B)**. Genome-wide examination revealed widespread changes in H3K27ac at enhancers, while H3K27ac at promoters showed little change, consistent with a specific effect of ARID1A inactivation at enhancers **(Figure 4-3C)**. ARID1A inactivation changed not only the level of histone modifications at enhancers, but also their locations **(Figure 4-3D).**

Because H3K27ac distinguishes active from poised/inactive enhancers (Creyghton et al. 2010, Shlyueva et al. 2014), we asked whether altered SWI/SNF targeting in *ARID1A*$^{-/-}$ cells affected enhancer activity. Indeed, changes in SWI/SNF binding at TSS-distal sites correlated with changes in mRNA levels of nearest genes as quantified by RNA-Seq **(Figure 4-3E).** Mono-allelic ARID1A inactivation also affected enhancer activity, with *ARID1A*$^{+/-}$ cells showing intermediate changes in H3K27ac at enhancers and in transcription of the nearest genes **(Figure 4-4A-C).** Together, these findings indicate that ARID1A inactivation specifically impairs enhancer configuration and activity, with marked consequences on gene expression.

Figure 4-3: ARID1A inactivation impairs SWI/SNF control of enhancer activity. (A) Distribution of SWI/SNF binding sites in HCT116 WT and *ARID1A*$^{-/-}$ cells relative to histone modifications (distribution between WT and *ARID1A*$^{-/-}$ cells is not significant in a paired t-test); (B) ChIP-seq profiles of SMARCA4, SMARCC1, H3K4me3, H3K4me1, and H3K27ac in WT and *ARID1A*$^{-/-}$ cells around all TSS-distal SWI/SNF binding sites (labels on the right of the figure indicate number of sites in each category; labels in top right corners indicate any alterations made in scaling of Y-axis); (C) H3K27ac levels in WT and *ARID1A*$^{-/-}$ cells at TSS-proximal (promoter) and TSS-distal (enhancer) enrichment regions (the numbers in the three corners denote numbers of activated (>2x), inactivated (<1/2x), and stable sites); (D) ChIP-seq tracks of SMARCA4, SMARCC1, H3K4me3, H3K4me1, and H3K27ac in WT and *ARID1A*$^{-/-}$ cells; (E) Fold changes ($\log_2$) of gene expression between WT and *ARID1A*$^{-/-}$ cells for genes nearest to TSS-distal SWI/SNF binding sites split based on *ARID1A*$^{-/-}$ / WT ChIP-Seq signal.

Figure 4-3 (Continued)



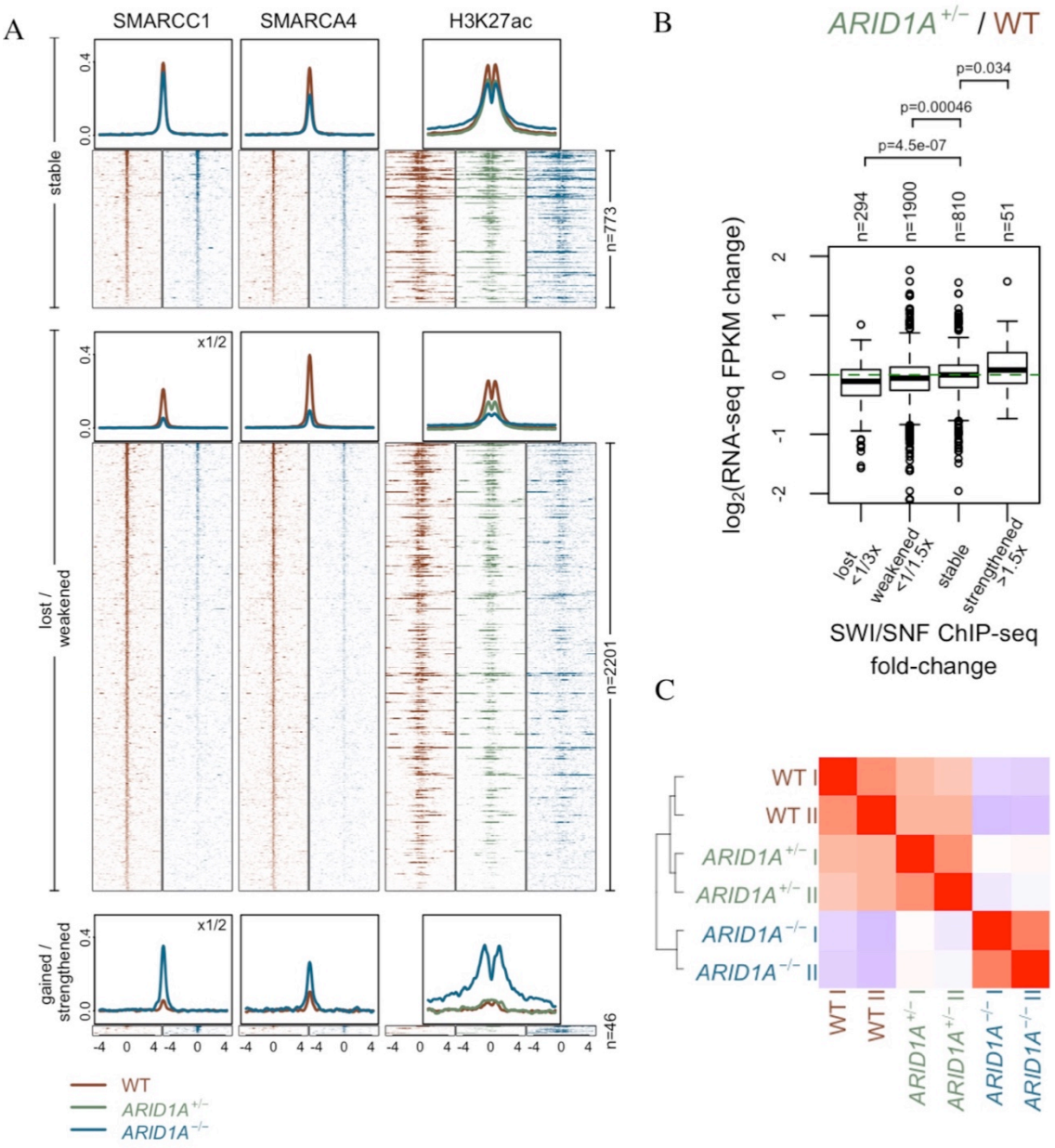
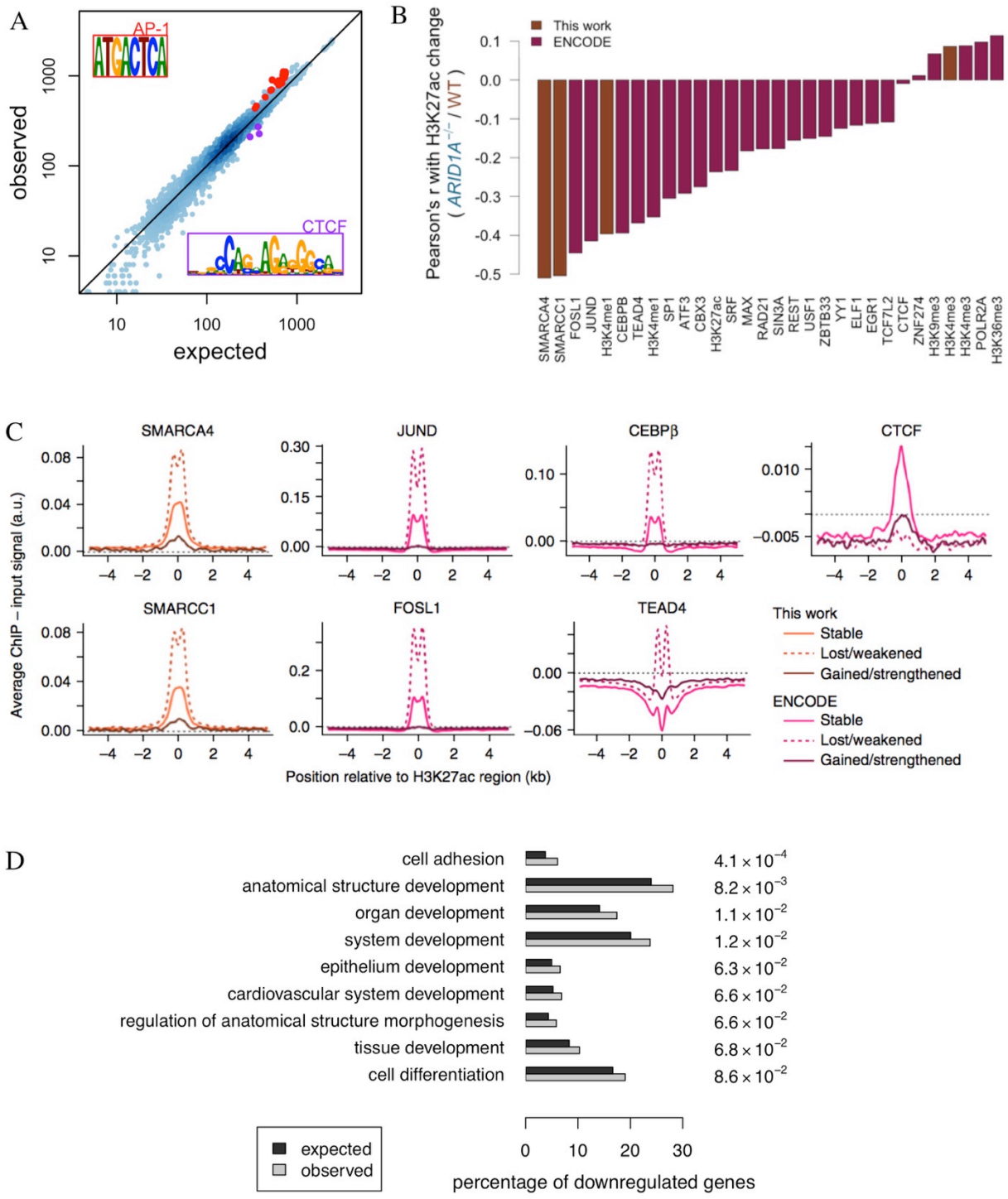


Figure 4-4: Mono-allelic ARID1A inactivation causes partial loss of enhancer activity. (A) H3K27ac ChIP-Seq profile in *ARID1A*$^{+/-}$ cells at TSS-distal SWI/SNF binding sites (same sites as in Figure 4-4B); (B) Fold changes (log_2) of gene expression between WT and *ARID1A*$^{+/-}$ cells for genes nearest to TSS-distal SWI/SNF binding sites split based on *ARID1A*$^{-/-}$ / WT ChIP-Seq signal (genes as Figure 4-4E); (C) Pearson's correlation among RNA-Seq samples based on FPKM values for two replicates each of WT, *ARID1A*$^{+/-}$, and *ARID1A*$^{-/-}$ cells.

To determine if control of enhancer activity by SWI/SNF complexes is coordinated with transcription factors (TFs), we analyzed sequence motifs at enhancers sensitive to ARID1A loss. The CTCF motif was relatively depleted in regions of H3K27ac loss, implying CTCF-bound insulator regions may resist modulation, while the AP1 (JUND/FOSL1) motif was most enriched **(Figure 4-5A)**. ChIP-Seq profiles for HCT116 cells, including those generated by ENCODE, revealed further that H3K27ac loss was most strongly associated with sites bound in WT cells by SWI/SNF complexes and/or TFs including AP1, CEBPB, and TEAD4 **(Figure 4-5B)**. Among factors assessed in HCT116 cells, binding of these TFs correlated most strongly with SWI/SNF occupancy **(Figure 4-2A)** and was higher at enhancers that lost activity than at enhancers that were unaffected by ARID1A inactivation **(Figure 4-5C).** Gene Ontology analysis of nearest genes implicated enhancers that lost activity as regulators of cell adhesion, development, differentiation, and morphogenesis **(Figure 4-5D)**. Collectively, these results establish a broad role for ARID1A in regulating active enhancers in HCT116 cells, converging on dominant TFs that provide a large fraction of tissue-specific gene regulation.

Figure 4-5: Characterization of enhancers that lose activity upon ARID1A inactivation. (A) Observed vs. expected TF motif instances at TSS-distal H3K27ac regions (enhancers) with reduced H3K27ac signal based on enrichment regions with stable H3K27ac signal. Motifs highly similar to AP1 and CTCF motifs are highlighted; (B) Correlation between H3K27ac signal change (*ARID1A*$^{-/-}$ / WT) and WT ChIP-Seq signal levels of different factors profiled in this work and by the ENCODE Project; (C) ChIP-Seq profiles for SMARCA4, SMARCC1, JUND, FOSL1, and CTCF in WT HCT116 cells centered around TSS-distal H3K27ac regions (enhancers) that remain stable, show lost/weakened H3K27ac, or show gained/strengthened H3K27ac in *ARID1A*$^{-/-}$ cells relative to WT; (D) Genes enriched near with TSS-distal H3K27ac regions (enhancers) with downregulated H3K27ac between HCT116 WT and *ARID1A*$^{-/-}$ cells relative to all TSS-distal H3K27ac regions.

Figure 4-5 (Continued)

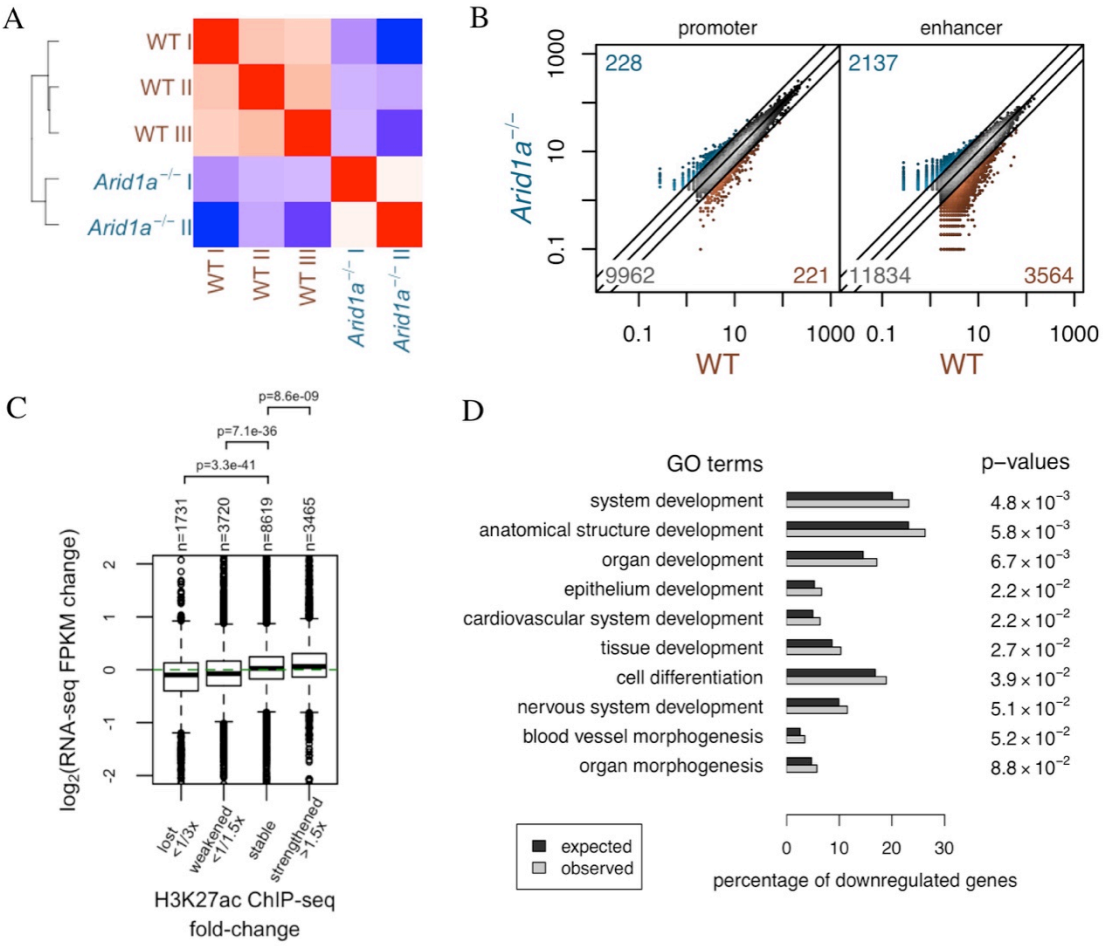


To determine if ARID1A inactivation impairs enhancer activity *in vivo*, we examined gene expression and H3K27ac in the colonic epithelium of wildtype and Villin-Cre^{ER-T2} *Arid1a*$^{fl/fl}$ mice. More than 1,000 genes were dysregulated >2-fold in *Arid1a*$^{-/-}$ colonic epithelium (FDR<0.05) (**Figure 4-5A**). Again, whereas promoter activation states were changed little, the effects on H3K27ac at enhancers were significant and were correlated with changes in mRNA levels of the nearest genes **(Figure 4-5B-C).** Developmental enhancers were identified as most sensitive to ARID1A loss, showing many Gene Ontology terms shared with HCT116 cells **(Figure 4-5D).** For example, the Gasdermins were markedly downregulated **(Figure 4-5E-F)**. Although it is difficult to attribute tumorigenesis to discrete target genes, our findings implicate broad control of enhancer activity as the SWI/SNF function crucially impaired by ARID1A deficiency.

While SWI/SNF function is described at promoters as well as enhancers (Euskirchen et al. 2011, Tolstorukov et al. 2013), our findings implicate enhancers as the principal sites at which SWI/SNF complexes function to regulate gene activation. When ARID1A is absent, SWI/SNF binding is lost from thousands of enhancers that subsequently lose activity, showing reduced H3K27ac levels and expression of nearest genes. ARID1A loss impairs SWI/SNF control of enhancers bound by dominant transcription factors that activate gene expression programs critical for development and differentiation. Enhancers are major determinants of cell type specificity in gene expression (Bulger and Groudine 2011). As SWI/SNF complexes regulate gene expression across lineages and developmental states, defective SWI/SNF control of enhancer activity may underlie not only the oncogenic drive of *ARID1A*-mutant colon cancers but also that of other human malignancies driven by alterations in ARID1A and other SWI/SNF subunits.

Figure 4-6: ARID1A inactivation impairs enhancer-mediated gene regulation *in vivo*. (A) Pearson's correlation among RNA-Seq samples based on FPKM values for mouse colon epithelium dissociated from individual wildtype mice (WT, n=3) and Villin-Cre^{ER-T2} *Arid1a*$^{fl/fl}$ mice (*Arid1a*$^{-/-}$, n=2); (B) H3K27ac levels at TSS-proximal (promoter) and TSS-distal (enhancer) enrichment regions for colon epithelium from wildtype (WT) and Villin-Cre^{ER-T2} *Arid1a*$^{fl/fl}$ (*Arid1a*$^{-/-}$) mice (numbers in the three corners denote numbers of activated (>2x), inactivated (<1/2x) and stable sites); (C) Fold changes ($\log_2$) of gene expression between WT and *Arid1a*$^{-/-}$ mouse colon epithelium for genes nearest to TSS-distal H3K27ac regions (enhancers) split based on *Arid1a*$^{-/-}$ / WT ChIP-Seq signal; (D) Genes enriched near with TSS-distal H3K27ac regions (enhancers) with downregulated H3K27ac between WT and *Arid1a*$^{-/-}$ mouse colonic epithelium relative to all TSS-distal H3K27ac regions; (E) H3K7ac ChIP-Seq tracks, super-enhancer (SE) calls, and RNA-Seq tracks at *Gsdmc* locus in WT and *Arid1a*$^{-/-}$ mouse colon epithelium; (F) RNA-Seq FKPM values for individual WT and *Arid1a*$^{-/-}$ mice for *Gsdmc, Gsdmc2, Gsdmc3,* and *Gsdmc4*.

Figure 4-6 (Continued)



F

Gene	WT I	WT II	WT III	*Arid1a*$^{-/-}$ I	*Arid1a*$^{-/-}$ II	fold-change
Gsdmc	21.9	16.5	20.9	1.33	0.442	0.0491
Gsdmc2	1140	835	797	33.2	14.7	0.0255
Gsdmc3	232	211	234	13	5.16	0.0414
Gsdmc4	1520	996	1110	104	27	0.0533

Methods

Cell culture

The HCT116 cell line and derivative *ARID1A*$^{+/-}$ and *ARID1A*$^{-/-}$ isogenic cell lines were purchased from Horizon Discovery (HD 104-031 and HD 104-049) and cultured in RPMI 1640 supplemented with 10% FBS as per instructions. These cell lines were negative for mycoplasma and all other infectious agents evaluated under the Mouse/Rat Comprehensive CLEAR Panel (Charles River Research Animal Diagnostic Services). Whole cell extracts of isogenic HCT116 cell lines were used in Western blots for E-Cadherin (Cell Signaling Technology: 24E10), ARID1A (Cell Signaling Technology: 12354), and ACTIN (Cell Signaling Technology: 5125). For invasion assay, 25,000 serum-starved cells were added to the top of each BioCoat Matrigel Invasion Chamber (BD Biosciences #354480) with chemoattractant (RPMI, 10% FBS) at the bottom of wells. Chambers were incubated for 22 hours at 37°C, 5% CO_2. Non-invading cells were removed from upper surface of the membrane by scrubbing with cotton tipped swabs. Blind counts of invaded cells were obtained following Crystal Violet staining of live cells. This assay was performed with 4 replicates for each cell type. Two-tailed *t*-tests were used to determine statistical significance, F tests to compare variances.

Dissociation of mouse colon epithelial cells

5-7 week old Villin-Cre^{ER-T2} *Arid1a*$^{fl/fl}$ mice and littermate control *Arid1a*$^{fl/fl}$ mice were administered 1mg Tamoxifen (T5648 SIGMA) dissolved in sunflower oil (S5007 SIGMA) via intraperitoneal injection for 5 consecutive days. Three Villin-Cre^{ER-T2} *Arid1a*$^{fl/fl}$ mice and three

control *Arid1a*$^{fl/fl}$ mice were euthanized 8 weeks post-injection for dissociation of colon epithelial cells. Mouse colons were dissected, flushed, and splayed. Colon pieces were rinsed in PBS, incubated in 5mM EDTA, and shaken for 2 minutes to release dissociated epithelial cells.

Sample preparation for RNA-Seq:

Trizol Reagent (Life Technologies 15596-026) was used to isolate RNA from harvested HCT116 isogenic cell lines and dissociated mouse colon epithelial cells. RNA was further purified with the RNeasy Mini Kit (Qiagen) and Turbo DNA-Free Kit (Ambion; murine cells only). Sequencing libraries were generated using Tru-Seq Technology (Illumina) for the Illumina Hi-seq Genome Analyzer.

RNA-Seq Processing

The sequenced reads from each sample were aligned to the human/mouse genome+transcriptome assemblies GRCh37.72/NCBIM37.67 using TopHat (Kim et al. 2013) v2.0.8 with default parameters except turning off novel junction search ('-G <gtf> --no-novel-juncs' options). The transcriptome was self-merged to allow processing with cufflinks (Trapnell et al. 2013) v2.1.1 tool cuffdiff, "cuffcompare -s hg19.fa -CG -r GRCh37.72.gtf GRCh37.72.gtf" (and similarly for mouse). Different conditions were compared using cuffdiff with default parameters and bias correction ('-G <gtf> -b' options).

Each SWI/SNF or H3K27ac RoE was associated to the closest active TSS as defined above. The connection between ChIP-seq signal change and RNA-seq change was studied for TSS-distal

RoEs. Only RoEs for which the closest active TSS is between 5kb and 100kb are retained. IP signal was quantified as 'total IP signal in RoE per billion mapped reads + pseudocount of 0.1'. RNA signal was quantified as 'normalized gene level count value from cuffdiff + 5'. The ratio of IP signal for Arid1a-deficient divided by WT was used to categorize RoEs to four groups: more than 3-fold signal loss, between 1.5 to 3-fold signal loss, less than 1.5-fold change, and more than 1.5-fold signal increase. The ratio of RNA-seq signal for Arid1a-deficient divided by WT signal was plotted for each category. The heatmaps for RNA-seq results show mean-shifted log expression values: log2(normalized gene level count value from cuffdiff + 5) - log2(average value for the gene across all samples).

Sample preparation for ChIP-Seq

HCT116 isogenic cell lines were dual-crosslinked in 2mM disuccinimidyl glutarate (DSG; Life Technologies #20593) for 30 min then 1% formaldehyde for 10 min, followed by 5 min glycine quenching. Nuclear extracts were generated following 3 washes in PBS. Chromatin was fragmented using Covaris sonication (adaptive focused acoustics; AFA technology). The following antibodies were used for immunoprecipitation of 30ug solubilized chromatin: SMARCC1/BAF155 (Santa Cruz sc9746), SMARCA4/BRG1 (Abcam ab110641), H3K27ac (Abcam ab4729); H3K4me (Abcam ab8895), H3K4me3 (Abcam ab8580); Antibody:chromatin complexes were pulled down with Protein G dynabeads (Life Technologies 10004D), washed, and eluted. Chromatin crosslinks were reversed and samples were treated with Proteinase K and RNAse A. ChIP-DNA was extracted with the Min-Elute PCR purification kit (Qiagen) and

quantified with Quant-iT™ PicoGreen dsDNA Assay Kit (Life Technologies). 10ng of purified ChIP-DNA was used to prepare sequencing libraries for the Illumina Hi-Seq Genome Analyzer.

Dissociated colon epithelial cells pooled from 3 Villin-Cre^{ER-T2} *Arid1a*$^{fl/fl}$ mice and from 3 control *Arid1a*$^{fl/fl}$ mice were fixed in 2% formaldehyde, lysed, and pulsed at 15% amplitude on a tip-sonicator. Sonicated chromatin was immunoprecipitated using antibodies for H3K27ac (Abcam ab4729) and H3K4me3 (Abcam ab8580). Antibody:chromatin complexes were pulled down with Protein G dynabeads (Life Technologies 10004D), washed, and eluted. Chromatin crosslinks were reversed and samples were treated with Proteinase K and RNAse A. ChIP-DNA was extracted with the Min-Elute PCR purification kit (Qiagen) and quantified with Quant-iT™ PicoGreen dsDNA Assay Kit (Life Technologies). 10ng of purified ChIP-DNA was used to prepare sequencing libraries for the Illumina Hi-Seq Genome Analyzer.

Alignment, fragment size estimation, and library complexity

The sequenced reads were aligned to the hg19/mm9 genome assembly using Bowtie (Landmead et al. 2009) 0.12.6, allowing up to 10 matches ('-m 10 --best' options). For HCT116, reads on the 24 assembled chromosomes excluding the ENCODE blacklisted regions were kept for downstream analysis. For the mouse samples, reads on the 19 assembled autosomes excluding a custom 280kb blacklist region were kept for downstream analysis. The custom blacklist regions were selected based on very high signal in input tracks in a parallel study (GSE71509). Peaks of cross-correlation profiles were identified to estimate the typical fragment size for each sample. The typical fragment size for the different samples ranged between 140-180bp. Each read was considered to represent a signal at half typical fragment size from the 5' end. Library complexity

was calculated for each sample as the number of unique bp positions mapped on each strand, divided by the total number of mapped reads. For batches of experiments where the typical library complexity was below 90% (all mouse samples, HCT116 H3K4me1/3 and matching input), only one read mapping to each position was kept.

Identification of regions of enrichment (RoE)

Different ChIP-seq regions of enrichment (RoE) were identified using the SPP package (Kharchenko et al. 2008) in R, with the function get.broad.enrichment.clusters and option window.size=500, with matching input samples for each IP experiment, using appropriate z.thr values for each analysis as specified below. The input samples for samples for WT and ARID1A-deficient mice had relatively low sequencing depth (5.7 and 12.0M reads after selecting one read per position). The two samples appeared to show no more systematic variability than expected from the statistical variability due to low sequencing depth. Therefore the two were merged and the combined input was provided to SPP as a control for both WT and ARID1A-deficient H3K27ac ChIP-seq.

Defining active TSSs and H3K4me3 RoEs

Active TSSs in HCT116 were defined as all TSSs in Ensembl release GRCh37.72 that overlapped H3K4me3 RoEs (z>4) in either condition. Active TSSs in mouse colon were defined as all TSSs in Ensembl release NCBIM37.67 that overlapped an H3K27ac (z>4) peak in either condition (H3K4me3 ChIP-seq failed for ARID1A-deficient mouse colon; the enrichment at

TSSs was relatively low for the WT colon sample). Additionally, H3K4me3 RoEs in mouse colon were defined as the union of RoEs (z>4) called with the WT colon sample and a WT MEF sample from the accompanying study (GSE71509). These two regions, i.e., "active TSSs" and "H3K4me3 RoEs" were used together to conservatively define TSS-proximal and TSS-distal regions in downstream analyses, as specified below.

Identification and classification of SWI/SNF binding sites

SWI/SNF binding sites were identified in two steps: First, overlapping SMARCA4 and SMARCC1 RoEs with z>4 were called for each condition, WT and *ARID1A*$^{-/-}$. Next, the union of the regions for the two conditions was calculated. This approach reduces any bias that may arise in differential RoE calling due to thresholds. The sites that were called in WT, or those where the signal in WT was more than half the signal in *ARID1A*$^{-/-}$ were considered as SWI/SNF binding sites in WT cells; the complementary selection was performed for SWI/SNF binding sites in *ARID1A*$^{-/-}$. SWI/SNF binding sites overlapping both an active TSS and an H3K4me3 RoE were called TSS-proximal. Those more than 1kb away from an H3K4me3 RoE and more than 2kb away from an active TSS were called TSS-distal. Others were ambiguous and excluded from studies specific to TSS-proximal or distal sites. When we evaluated where SWI/SNF binding falls in the genome, we used H3K27ac and H3K4me1 RoEs that are called inclusively with a z>3 threshold. Changes in TSS-distal SWI/SNF binding upon ARID1A loss were evaluated at each binding site, by dividing the library-size normalized IP signal for SMARCA4 and SMARCC1 in *ARID1A*$^{-/-}$ by WT. If the geometric mean of change was greater than 1.5 fold, and both factors showed increased signal, the RoE was called as gained/strengthened. In reverse, if the geometric

mean of change was less than 2/3 fold, and both factors showed decreased signal, the RoE was called as lost/weakened. Other sites were called as unchanged.

Identification and classification of H3K27ac RoEs (promoters and enhancers)

H3K27ac RoEs with z>4 were called to specifically study changes in this mark in HCT116 and colon epithelial cells. Similar to SWI/SNF binding sites, H3K27ac RoEs from different conditions were merged, ones overlapping both an active TSS and an H3K4me3 RoE were defined as promoters; and ones that were more than 1kb away from an H3K4me3 RoE and more than 2kb away from an active TSS were called as enhancers.

Gene ontology analysis for enhancers

To identify enhancers that lost SWI/SNF binding, we used an inclusive definition of SWI/SNF binding as overlapping SMARCA4 and SMARCC1 RoEs with a z>3 threshold. Enhancers with SWI/SNF binding, where the average SMARCA4 and SMARCC1 signal was down more than 1.5 fold were called as SWI/SNF losing enhancers. GO analysis for these SWI/SNF losing enhancers was performed as follows: Gene Ontology databases were downloaded from geneontology.org on 2014/04/29. Each enhancer was associated to the closest active TSS within 100kb. p-values for gene set enrichment for genes associated to SWI/SNF losing enhancers were calculated relative to genes associated with any enhancer using hypergeometric test. q-values were obtained based on Benjamini–Hochberg procedure.

IP efficiency correction for H3K27ac samples

The efficacy of IP pull-down may vary between different ChIP-seq experiments. A number of lines of evidence suggested that the real levels of H3K27ac are unchanged at a large fraction of promoters upon ARID1A deletion: i. we found that a large fraction of promoters show the same amount of fold-change with very small variance; ii. the typical fold-change was independent of SWI/SNF binding at promoters, and was the same as at enhancers with no SWI/SNF binding; iii. in the accompanying MEF study, we observed different fold-changes at promoters, both greater and less than one-fold for replicates of experiments upon *Smarcb1* knockout, while we saw consistent decrease of H3K27ac in western blots and at enhancers. Based on these observations, we applied a small multiplicative factor on H3K27ac samples to set the mode of the log-fold-change distribution at promoters to zero while comparing WT and ARID1A-deficient samples. These factors were HCT116: Parental: 0.95, *ARID1A*$^{-/-}$: 1.31, *ARID1A*$^{+/-}$: 0.82; Mouse colon epithelial cells: WT: 1.08, ARID1A-deficient: 0.92. This normalization procedure does not affect the qualitative observations presented. We refrained from applying a similar normalization for other ChIP-seq sample pairs, since we could not confidently determine a set of regions where they are unaffected upon ARID1A loss.

ChIP-seq visualization

Genomic profiles for visualization were generated using Gaussian smoothing with sigma=100bp after library size normalization. The SWI/SNF binding site heatmaps were centered at the position with highest signal in the smoothed profile obtained by summing the four tracks considered, (WT or *ARID1A*$^{-/-}$, SMARCA4 and SMARCC1). The heatmaps show input

subtracted values, whereas the browser shots show raw smoothed signal. The average profiles for each class were obtained as 0.1-0.9 trimmed linear mean at each position.

Transcription Factor Motif Enrichment

Transcription factor motif maps for hg19 for 4095 motifs (including a redundant set of real transcription factor recognition elements, and shuffled motif control sequences) were downloaded from http://compbio.mit.edu/encode-motifs/ (Kheradpour et al. 2014). The position weight matrix (pwm) for each motif was calculated based on the actual sequences of the provided motif locations. The number of motif occurrences was counted for each motif inside lost/weakened enhancers (H3K27ac fold change< 1/1.5) and unchanged enhancers (1/1.5<H3K27ac fold-change<1.5). A lowess curve was calculated to model the ratio of counts for each motif (sensitive/insensitive) as a function of the GC content of the motif pwm. This curve was used to calculate back the null hypothesis expected number of occurrences for each motif in the lost/weakened. The figure shows observed counts vs. expected counts for the 4095 motifs. pwms for two selected motifs ("AP-1_known3_8mer" and "CTCF_known1_8mer") are displayed on the figure. Motif similarity was assessed based on Pearson correlation values between motif pwms; motifs which are similar to the two selected motifs (r>0.85) are highlighted on the figure.

Correlation of binding of different factors in WT with H3K27ac changes

ChIP-seq signal around the peak point of each H3K27ac RoE was calculated within +/-1.5kb, per million mapped reads; a pseudocount of 1 was added, and the values were logged ($L=\log_2(1+\text{signal})$). The change in H3K27ac was calculated using samples generated in this work as L(ARID1A)-L(WT). For each factor from ENCODE project, the L values were averaged over replicates at each H3K27ac RoE. The Pearson's correlation between each factor and the change in H3K27ac is plotted.

Chapter 5:

ARID1B preserves SWI/SNF function in *ARID1A*-mutant cancer

Abstract

ARID1A and ARID1B are homologous, mutually exclusive subunits of SWI/SNF complexes. ARID1B has been identified as a specific vulnerability in *ARID1A*-mutant cancers by genome-wide synthetic lethality screening. However, the mechanistic basis of synthetic lethality is unknown. Here, we utilize our newly established model systems to investigate the role of ARID1B in *ARID1A*-mutant colon cancer. We find ARID1B is expressed in ARID1A-deficient colon adenocarcinomas in *Arid1a*$^{fl/fl}$ mice and that ARID1B knockdown specifically impairs the proliferation of ARID1A-deficient HCT116 human colon cancer cells. We further demonstrate that the composition and integrity of ARID1B-containing complexes is unaffected by ARID1A inactivation, and that ARID1B-containing SWI/SNF complexes bind enhancers that remain active in ARID1A-deficient cells. These results provide support for targeting residual SWI/SNF complexes in *ARID1A*-mutant colon cancers, and suggest that the function of residual complexes in preserving activity of a subset of enhancers might be specifically targeted for therapy.

Contributions

The following individuals contributed to the work described in this chapter:

Radhika Mathur[1,2], Burak Han Alver[3], Xiaofeng Wang[2], Peter J. Park[3,6], Ramesh A. Shivdasani[4,5], and Charles W. M. Roberts[2,6]

[1]Program in Biological & Biomedical Sciences, Harvard Medical School, Boston MA, 02215, USA

[2]Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, MA 02215, USA

[3]Department of Biomedical Informatics, Harvard Medical School, Boston, MA 02115, USA

[4]Department of Medical Oncology and Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, Boston, MA 02215, USA

[5]Departments of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA

[6]Department of Oncology, St. Jude Children's Research Hospital, Memphis, TN 38105, USA

R.M. and C.W.M.R conceived experiments and study design. Experiments were performed by R.M. X.W provided technical guidance. Computational and statistical analyses were performed by R.M. and B.H.A., with guidance from P.J.P. All authors contributed to data analysis and interpretation. R.M., R.A.S., and C.W.M.R. wrote the manuscript with input from all authors.

Acknowledgements

We thank S.H. Orkin for mentorship, and all members of the Roberts and Orkin labs for discussion. This work was supported by US National Institutes of Health grants R01CA172152 (C.W.M.R.) and R01DK081113 (R.A.S), by a Claudia Adams Barr grant (C.W.M.R.), and by an Innovation Award from Alex's Lemonade Stand (C.W.M.R). R.M. and A.K.S.R. were supported by the US National Institutes of Health predoctoral fellowships (1F31CA199994 and 1F31CA180784). X.W. was supported by the Pathway to Independence Award from the US National Institutes of Health (K99CA197640). The Cure AT/RT Now foundation, the Avalanna Fund, the Garrett B. Smith Foundation, Miles for Mary (C.W.M.R.), and the Lind Family (R.A.S.) provided additional support.

Citation of published work

All data presented in this chapter are published in the article:

Mathur, R., Alver, B.H., San Roman, A.K., Wilson, B.G., Wang, X., Agoston, A.T., Park, P.J., Shivdasani, R.A., and Roberts, C.W.M. (2017) ARID1A loss impairs enhancer-mediated gene regulation and drives colon cancer in mice. *Nature Genetics*, *49*(2), 296–302. http://doi.org/10.1038/ng.3744.

ARID1A and ARID1B are homologous, mutually exclusive subunits of SWI/SNF complexes, containing the "ARID" AT-rich interaction domain, which is capable of binding DNA in a sequence non-specific manner (Nie 2000, Wang 2004). Synthetic lethality-based screening approaches have identified ARID1B as a specific vulnerability in *ARID1A*-mutant cancers (Helming et al. 2014a). Several distinct possibilities have been suggested for the mechanistic basis of this finding. These include functional redundancy of ARID1A and ARID1B-containing SWI/SNF complexes and neomorphic gain-of-function of ARID1B-containing complexes in the absence of ARID1A (Helming et al. 2014b).

Here, we sought to elucidate the mechanistic basis of synthetic lethality between ARID1A and ARID1B utilizing our newly established models of *ARID1A*-mutant colon cancer. We found that ARID1B protein was present in the colon epithelium of wildtype mice and also in ARID1A-deficient colon cancers in both MX1-Cre *Arid1a*$^{fl/fl}$ and Villin-Cre$^{ER\text{-}T2}$ *Arid1a*$^{fl/fl}$ mice **(Figure 5-1).** ARID1B was also expressed in HCT116 WT, *ARID1A*$^{+/-}$, and *ARID1A*$^{-/-}$ cells; shRNA-mediated knockdown of ARID1B selectively impaired the proliferation of *ARID1A*$^{-/-}$ cells **(Figure 5-2)**. These results validated ARID1B as a specific vulnerability and potential therapeutic target in *ARID1A*-mutant colon cancer.

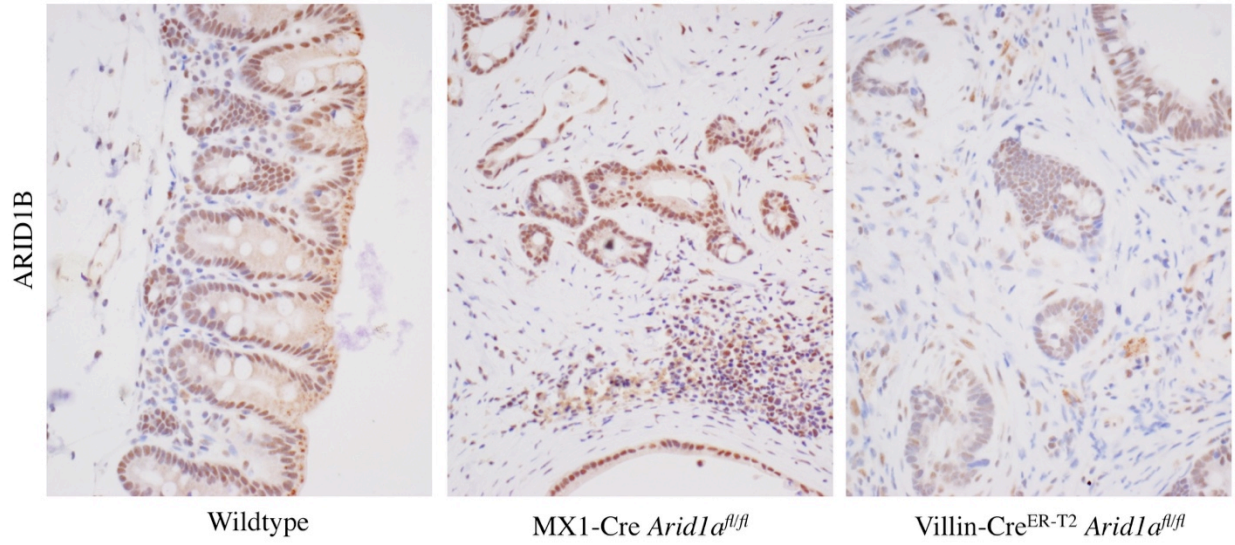


Figure 5-1: ARID1B is expressed in colon tumors from *Arid1a*$^{fl/fl}$ mice. IHC for ARID1B shown for wildtype mouse colon and sections of invasive colon adenocarcinoma from MX1-Cre *Arid1a*$^{fl/fl}$ and Villin-Cre$^{ER\text{-}T2}$ *Arid1a*$^{fl/fl}$ mice.

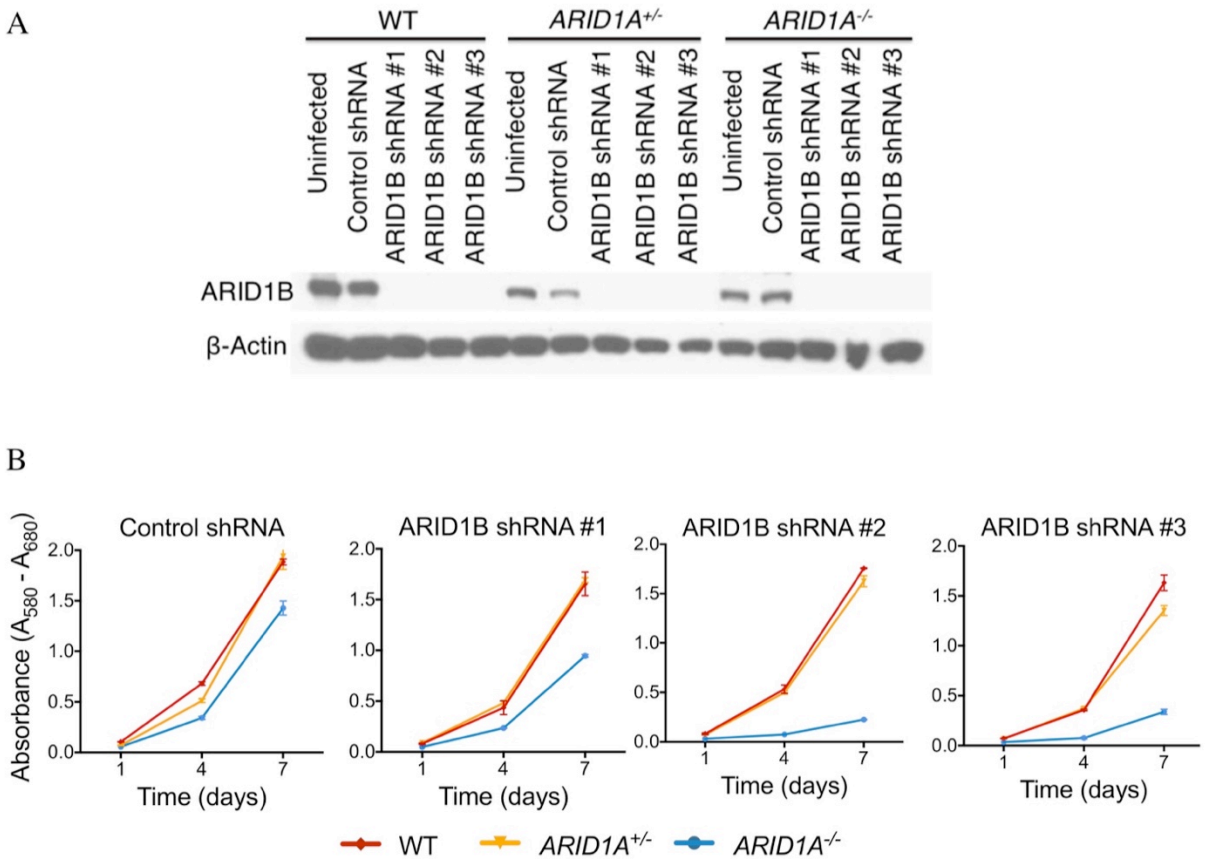


Figure 5-2: ARID1B is a specific vulnerability in ARID1A-deficient human colon cancer cells. (A) Protein levels of ARID1B and β-actin following ARID1B knockdown with 3 independent shRNAs; (B) Proliferation following shRNA-mediated ARID1B knockdown measured by MTT assay.

We next utilized immunoprecipitation to examine the composition of ARID1A and ARID1B-containing complexes in the HCT116 cell model **(Figure 5-3).** As expected for mutually exclusive subunits, ARID1A and ARID1B did not co-immunoprecipitate. Some distinctions were noted in subunit composition of ARID1A and ARID1B-containing complexes;

SMARCA2 was not detected upon immunoprecipitation with ARID1B, suggesting selective incorporation into ARID1A-containing complexes. ARID1A inactivation showed no effect on the protein levels of SWI/SNF subunits or on the incorporation of subunits into ARID1B-containing complexes. The composition and integrity of ARID1B-containing SWI/SNF complexes is thus maintained upon ARID1A inactivation.

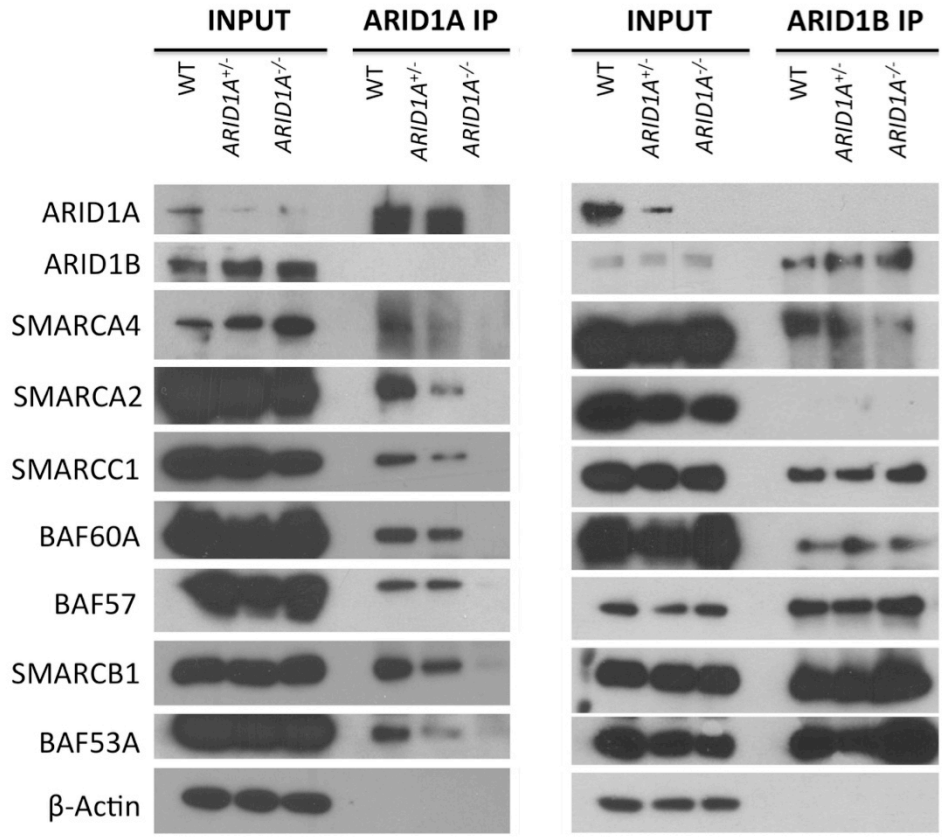


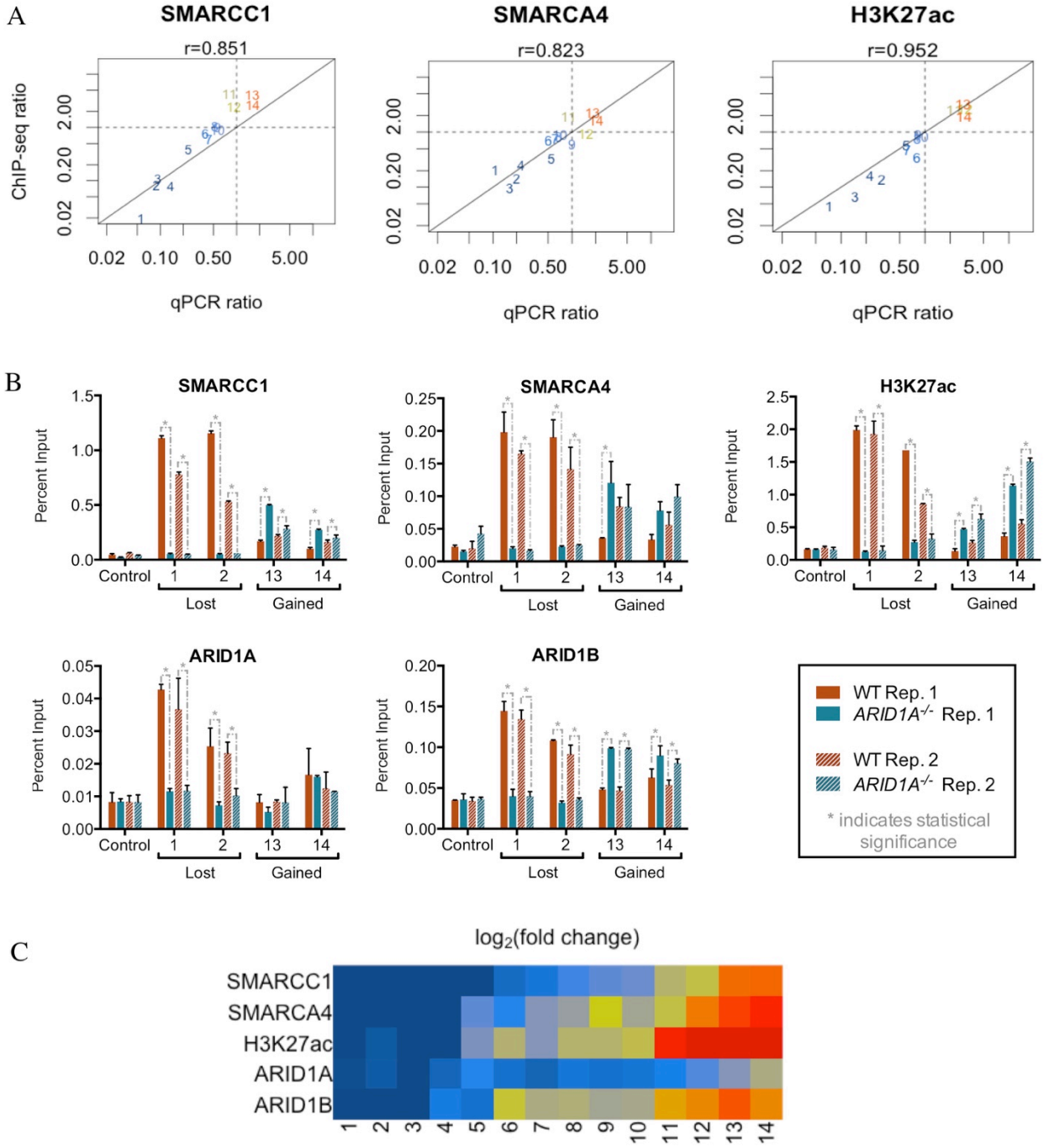
Figure 5-3: ARID1B-containing SWI/SNF complexes remain intact upon ARID1A deficiency. Western blots for SWI/SNF subunits and β-actin (control) following immunoprecipitation with antibodies targeting ARID1A and ARID1B.

To examine targeting of ARID1A and ARID1B containing SWI/SNF complexes to chromatin, we developed a ChIP-qPCR assay to specifically interrogate a set of 14 enhancers selected based upon ChIP-Seq data, which showed a broad range of changes in SWI/SNF binding following ARID1A inactivation. We validated this assay by performing ChIP-qPCR for SMARCA4, SMARCC1, and H3K27ac, finding that similar changes in binding were identified by ChIP-qPCR and by ChIP-Seq **(Figure 5-4A)**. We then examined binding of ARID1A and ARID1B at these sites **(Figure 5-4B-C)**. We found ARID1A was lost from all enhancers that lost SWI/SNF binding. ARID1B was also lost from enhancers that lost SWI/SNF binding, but was present at enhancers that had retained or gained SWI/SNF binding.

While complexity in binding patterns of ARID1A and ARID1B is yet to be fully elucidated, these results reveal that the role of residual ARID1B-containing SWI/SNF complexes is to bind and preserve the activity of a subset of enhancers in *ARID1A*-mutant cancers. These results provide insight into the mechanistic basis of synthetic lethality between ARID1A and ARID1B, supporting a model where enhancers bound by residual ARID1B-containing SWI/SNF complexes are essential for continued proliferation of ARID1A-deficient cells. Therapeutic targeting of these cancers may potentially be achieved by modulation of enhancer activity, in addition to direct inhibition of ARID1B-containing residual SWI/SNF complexes.

Figure 5-4: ARID1B is present at enhancers active in ARID1A-deficient cells. (A) Average log-fold change in *ARID1A*$^{-/-}$ / WT ChIP-qPCR and ChIP-Seq signal of SMARCA4, SMARCC1, and H3K27ac at 14 SWI/SNF-binding sites; (B) Input-normalized ChIP-qPCR signal of SMARCA4, SMARCC1, H3K27ac, ARID1A, and ARID1B at representative SWI/SNF binding sites shown for two independent biological replicate experiments, each performed in duplicate; (C) Average log-fold change in ChIP-qPCR signal in *ARID1A*$^{-/-}$ cells relative to WT of each factor at 14 SWI/SNF-binding sites.

Figure 5-4 (Continued)



Methods

Cell culture

The HCT116 cell line and derivative *ARID1A*$^{+/-}$ and *ARID1A*$^{-/-}$ isogenic cell lines were purchased from Horizon Discovery (HD 104-031 and HD 104-049) and cultured in RPMI 1640 supplemented with 10% FBS as per instructions. These cell lines were negative for mycoplasma and all other infectious agents evaluated under the Mouse/Rat Comprehensive CLEAR Panel (Charles River Research Animal Diagnostic Services).

Knockdown experiments

Lentiviral shRNAs on the PLKO.1 vector targeting ARID1B (clone #1: TRCN0000107361, clone #2: TRCN0000107363, clone #3: TRCN0000107364) were used to infect into HCT116 WT, *ARID1A*$^{+/-}$ and *ARID1A*$^{-/-}$cells. Infected cells were selected in Puromycin for 72 hours and then plated in triplicate onto an MTT assay (Cell Proliferation Kit, Roche # 1465007001) with 5000 cells/well.

Western blotting and co-immunoprecipitation

Whole cell extracts of isogenic HCT116 cell lines were used in Western blots for ARID1B (ABCAM ab57461) and ACTIN (Cell Signaling Technology: 5125). Nuclear extracts for co-immunoprecipitation were prepared using the NE-PER Nuclear and Cytoplasmic Extraction Kit (Thermo Scientific #78835). Nuclear extracts were diluted with RIPA buffer (Life Technologies 89900) to a final concentration of 1 mg/ml (with protease inhibitor cocktails, Roche). Each IP was incubated with SMARCC1/BAF155 antibody (Santa Cruz: sc9746), ARID1A antibody (Millipore PSG3), or ARID1B antibody (Santa Cruz 32762) overnight at 4°C. Protein G

Dynabeads (Life Technologies 10009D) were added and incubated at 4°C for 3 h. Beads were then washed three times with RIPA buffer and resuspended in reducing SDS gel loading buffer. Antibodies to the following proteins were used in the immunoblots: ARID1A (Cell Signaling Technology: 12354); ARID1B (Abcam: ab54761); SMARCA4/BRG1 (Santa Cruz: sc17796); BRM (Cell Signaling Technology: 11966); SMARCC2/BAF170 (Bethyl Laboratories: A301-039A); SMARCD1/BAF60A (Bethyl Laboratories: A301-595A); SMARCE1/ BAF57 (Bethyl Laboratories: A300-810A); SMARCB1/SNF5 (Bethyl Laboratories: A301-087A); ACTL6A/BAF53A (Bethyl Laboratories: A301-391A); ACTIN (Cell Signaling Technology: 5125, 1:3,000).

Sample preparation for ChIP-qPCR

HCT116 isogenic cell lines were dual-crosslinked in 2mM disuccinimidyl glutarate (DSG; Life Technologies #20593) for 30 min then 1% formaldehyde for 10 min, followed by 5 min glycine quenching. Nuclear extracts were generated following 3 washes in PBS. Chromatin was fragmented using Covaris sonication (adaptive focused acoustics; AFA technology). The following antibodies were used for immunoprecipitation of 30ug solubilized chromatin: SMARCC1/BAF155 (Santa Cruz sc9746), SMARCA4/BRG1 (Abcam ab110641), H3K27ac (Abcam ab4729); ARID1A (Santa Cruz 32761 and Millipore PSG3) and ARID1B (Abcam ab57461 and Santa Cruz 32762). Antibody:chromatin complexes were pulled down with Protein G dynabeads (Life Technologies 10004D), washed, and eluted. Chromatin crosslinks were reversed and samples were treated with Proteinase K and RNAse A. ChIP-DNA was extracted with the Min-Elute PCR purification kit (Qiagen) and quantified with Quant-iT™ PicoGreen dsDNA Assay Kit (Life Technologies). ChIP-qPCR was performed on the ViiA7 Real-Time PCR System (Life Technologies) using SYBR Select Master Mix (Life Technologies) using 1ul

of purified ChIP-DNA in duplicate in 384-well format. A minimum of two independent biological replicate experiments were performed for each factor analyzed by ChIP-qPCR.

ChIP-qPCR processing

ChIP-qPCR signals were normalized using the percent-input method: [100*2^(Adjusted input - Ct (IP)]. For each biological replicate experiment, statistical significance in WT v. *ARID1A*$^{-/-}$ ChIP-qPCR signal was determined using the Holm-Sidak method, with alpha=5.000% (GraphPad Prism version 6 for Mac OS X). Computations assume that all rows (individual ChIP-qPCR sites) are sample from populations with the same scatter (SD). Log-fold-change for each biological replicate experiment was calculated using the ratio of percent input for ARID1A-/- over WT (averaged over technical replicates). Averages of *ARID1A*$^{-/-}$ /WT log-fold-change values were calculated for each factor from all biological replicate experiments.

Chapter 6:

Discussion and future directions

SWI/SNF control of enhancer activity

This work implicates control of enhancer activity as the function of SWI/SNF complexes that is defective in *ARID1A*-mutant cancer. Studies from our laboratory have extended this conclusion to other systems, showing that activity of most enhancers is lost upon deletion of SWI/SNF subunits in mouse embryonic fibroblasts (Alver et al. 2017) and in SMARCB1-deficient malignant rhabdoid tumors (Wang et al. 2017). Re-expression of SMARCB1 in SMARCB1-deficient malignant rhabdoid tumor cell lines causes increased binding of SWI/SNF complexes to enhancers, which subsequently gain activity (showing increased H3K27ac and upregulation of nearest genes).

SWI/SNF complexes are evolutionarily conserved regulators of transcription, in control of inducible genes in yeast and homeotic genes in Drosophila. The crucial function of SWI/SNF complexes in mammalian cells appears to also be transcriptional regulation, with enhancers serving as the genomic sites of SWI/SNF function, enabling dynamic and precise gene regulation across cell and tissue types. SWI/SNF complexes interact with transcription factors at enhancers to accept and relay input from signaling pathways and other cellular machinery. It is likely that compositionally distinct complexes interact differently with these factors, such that multiple complexes are required within an individual cell to provide combinatorial complexity in gene expression patterns. Inactivation of individual SWI/SNF subunits may thus perturb enhancer-mediated gene regulation in some cellular contexts, but not in others, leading to the observed context specificity in tumor suppressor roles of individual SWI/SNF subunits.

While this and other recent work from our laboratory have established SWI/SNF complexes in control of enhancer activity, the precise mechanisms by which these complexes exert control remain unclear. While SWI/SNF binding at enhancers affects levels of H3K27ac,

this is likely an indirect effect, as the complex does not contain histone acetyltransferase activity. SWI/SNF complexes might function in recruiting the P300 acetyltransferase to catalyze H3K27acetylation (Alver et al. 2017). However, SWI/SNF complexes might also modulate chromatin structure directly at enhancers, such as by utilizing ATP-dependent chromatin remodeling activity to alter accessibility of nucleosomal DNA. The mechanisms by which enhancers function in gene activation remain poorly understood (Long et al. 2016, Sur and Taipale 2016, Hnisz et al. 2017); integrated analyses of SWI/SNF binding, histone modifications, nucleosome positioning, binding of transcriptional machinery, and promoter-enhancer looping are required to develop a sophisticated understanding of the mechanisms by which SWI/SNF complexes control enhancer activity, and thus gene regulation in mammalian cells. In addition to cancers, developmental disorders and intellectual disabilities have been linked to mutations in SWI/SNF subunits (Tsurusaki et al. 2012, Van Houdt et al. 2012, Shang et al. 2015). SWI/SNF control of enhancer activity is thus likely to be crucial not only in tumor suppression, but also in development, neurocognition, and other biological processes dependent upon enhancer-mediated gene regulation.

SWI/SNF pathway to tumorigenesis

This work demonstrates that ARID1A inactivation drives invasive colon adenocarcinoma in mice without the cooperation of additional mutations in colon cancer-associated genes, and that it antagonizes (rather than cooperates with) tumorigenesis initiated by aberrant activation of Wnt signaling. Defective SWI/SNF targeting and control of enhancer activity causes extensive dysregulation of gene expression in the ARID1A-deficient colon epithelium, with over 1000 genes affected. However, it remains unclear if genes deregulated by this process drive cancer by

classical mechanisms – alteration of discrete oncogenes or tumor suppressor genes – or by a mechanism that is fundamentally distinct. The enrichment of *ARID1A* mutations in a specific molecular subtype of colon cancer with distinct histological features than cancers driven by classical pathways suggests the latter; these features are also recapitulated in *Arid1a*$^{fl/fl}$ mouse models. Notably, developmental enhancers are broadly inactivated across systems upon SWI/SNF subunit mutations, but individual genes affected are variable. Collectively, these results indicate that defective SWI/SNF control of enhancer activity drives tumorigenesis via a novel pathway distinct from established models (**Figure 6-1**).

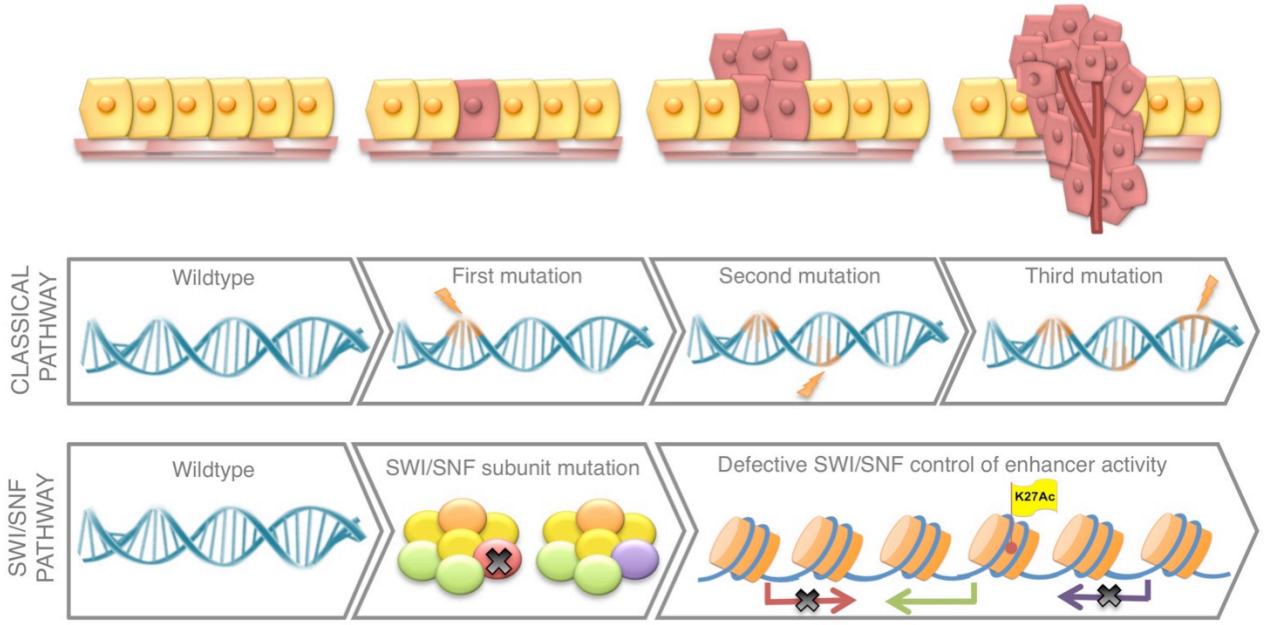


Figure 6-1: SWI/SNF pathway to tumorigenesis. Defective SWI/SNF function drives invasive cancer by impairing enhancer-mediated gene regulation. This pathway to tumorigenesis does not require a sequential accumulation of genetic mutations as in the classical pathway.

While multi-stage genetic models of cancer initiation and progression are best exemplified in colon cancer, they are also thought to underlie pathogenesis of other human

cancers. Malignant rhabdoid tumor and small cell carcinoma of the ovary, hypercalcemic type are driven by biallelic inactivation of *SMARCB1* and *SMARCA4* respectively and have stable genomes, despite their highly aggressive nature (Lee et al. 2012, Jelinic et al. 2014). In contrast, many cancers driven by SWI/SNF subunit mutations show genomic instability and/or cooperating genetic mutations and are therefore difficult to distinguish from cancers driven by classical pathways. *ARID1A*-mutant human colon cancers, for example, are associated with microsatellite-instability, a feature that is not recapitulated by our mouse model and is likely unrelated to the mechanism of tumor suppression by ARID1A. Stratification of cancers driven by SWI/SNF mutation as mechanistically distinct from other cancers will allow for systematic investigation of commonalities in SWI/SNF-mutant cancers, and may have immediate clinical implications for diagnostics, prognostics, and approaches to treatment.

Implications for therapy

Residual SWI/SNF complexes and Polycomb group complexes have each been identified as specific vulnerabilities in SWI/SNF-mutant cancers (Helming et al. 2014a, Helming et al. 2014b, Wilson et al. 2014, Hoffman et al. 2014, Wilson et al. 2010, Knutson et al. 2013, Kim and Roberts 2016, Fillmore et al. 2015, Bitler et al. 2015). These approaches to targeting SWI/SNF-mutant cancers have been validated in cancers with mutations in different SWI/SNF subunits, indicating they target a fundamental mechanism underlying SWI/SNF-mutant cancers. These approaches are attractive relative to chemotherapeutic agents, as they do not cause irreversible damage to DNA or other cellular processes. However, both SWI/SNF and Polycomb group complexes have context-dependent roles in tumor suppression (Masliah-Planchon et al.

2015, Raedt et al. 2014) and the particular sensitivities of SWI/SNF-mutant cancers to these approaches are largely unexplained.

Our investigation into the role of ARID1A inactivation in colon cancer pathogenesis yields a model in which defective SWI/SNF control of enhancer activity drives cancer and confers specific vulnerabilities that might be targeted for therapy **(Figure 6-2)**. Multiple SWI/SNF complexes with diverse subunit assemblies are required to maintain dynamic, spatiotemporally precise control of gene expression programs in a given cell. Defective control of enhancer activity upon SWI/SNF subunit mutation dysregulates these gene expression programs and leads to the creation of an altered transcriptional state. SWI/SNF-mutant cancers are then dependent upon the maintenance of their altered transcriptional state, such that they are especially sensitive to further perturbation.

The concept of oncogene addiction has recently been extended to transcription, as dysregulated gene expression programs acquired during early development of a tumor continue to be absolutely essential to its continued growth (Bradner et al. 2017). In the context of SWI/SNF-mutant cancers, altered transcriptional states are maintained by Polycomb group complexes, which function antagonistically to SWI/SNF complexes in mediating gene silencing, and also by residual complexes, which preserve activity of a subset of enhancers in SWI/SNF-mutant cancers. Transcriptional addiction provides a mechanistic explanation for the specific vulnerabilities of SWI/SNF-mutant cancers to perturbation of Polycomb and residual SWI/SNF function. Transcriptional addiction likely confers additional vulnerabilities yet to be discovered, both in general transcriptional and co-regulatory machinery that can be targeted across SWI/SNF-mutant cancers, and also in tissue-specific regulators upon which particular SWI/SNF-

mutant cancers may be dependent. Advances in chemical biology and drug discovery have created a new emphasis on these non-traditional drug targets (Gonda and Ramsey 2015, Jones et al. 2016), which, in combination with mechanistic understanding of SWI/SNF-mutant cancers, create a hopeful outlook for therapy.

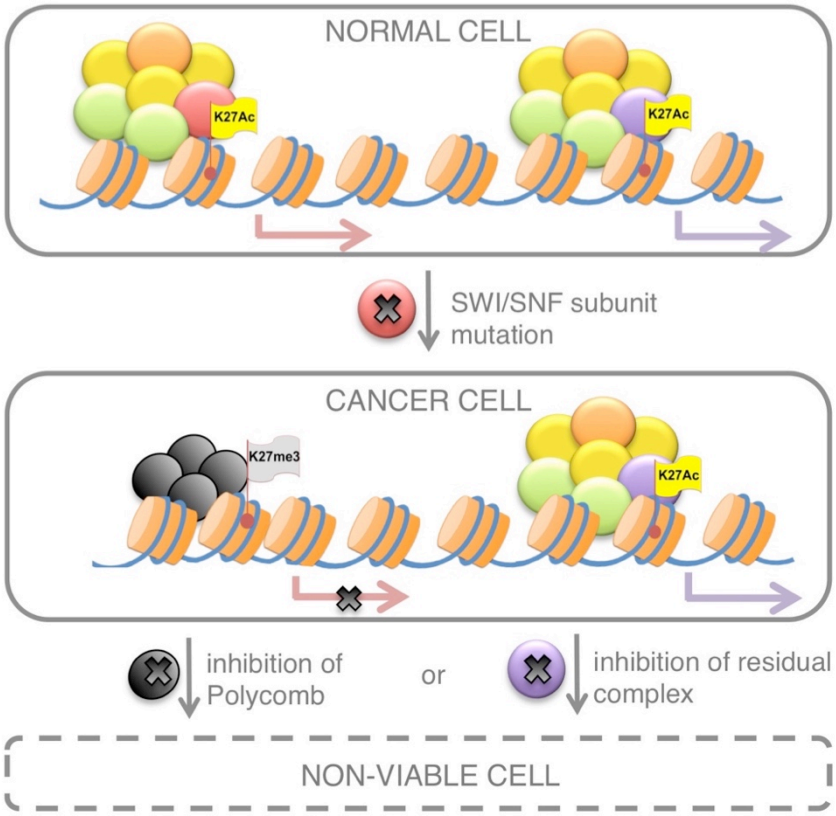


Figure 6-2: Defective SWI/SNF control of enhancer activity drives cancer and confers vulnerability. The transcriptional state of SWI/SNF-mutant cancers is dysregulated as a consequence of defective SWI/SNF control of enhancer activity. These cells are then dependent upon the maintenance of this transcriptional state, such that they are sensitive to additional perturbations such as inhibition of Polycomb group proteins or residual SWI/SNF complexes.

References

Alexander, J. M., Hota, S. K., He, D., Thomas, S., Ho, L., Pennacchio, L. A., & Bruneau, B. G. (2015). Brg1 modulates enhancer activation in mesoderm lineage commitment. *Development*, *142*(8), 1418–1430. http://doi.org/10.1242/dev.109496

Alver, B. H., Kim, K. H., Lu, P., Wang, X., Manchester, H. E., Wang, W., et al. (2017). The SWI/SNF chromatin remodelling complex is required for maintenance of lineage specific enhancers. *Nature Communications*, *8*, 14648. http://doi.org/10.1038/ncomms14648

American Cancer Society. *Cancer Facts & Figures 2017*. Atlanta, Ga: American Cancer Society; 2017.

Bitler, B. G., Aird, K. M., Garipov, A., Li, H., Amatangelo, M., Kossenkov, A. V., et al. (2015). Synthetic lethality by targeting EZH2 methyltransferase activity in ARID1A-mutated cancers. *Nature Medicine*, *21*(3), 231–238. http://doi.org/10.1038/nm.3799

Bossen, C., Murre, C. S., Chang, A. N., Mansson, R., Rodewald, H.-R., & Murre, C. (2015). The chromatin remodeler Brg1 activates enhancer repertoires to establish B cell identity and modulate cell growth. *Nature Immunology*, *16*(7), 775–784. http://doi.org/10.1038/ni.3170

Bradner, J. E., Hnisz, D., & Young, R. A. (2017). Transcriptional Addiction in Cancer. *Cell*, *168*(4), 629–643. http://doi.org/10.1016/j.cell.2016.12.013

Bulger, M., & Groudine, M. (2011). Functional and mechanistic diversity of distal transcription enhancers. *Cell*, *144*(3), 327–339. http://doi.org/10.1016/j.cell.2011.01.024

Cairns, B. R., Kim, Y. J., Sayre, M. H., Laurent, B. C., & Kornberg, R. D. (1994). A multisubunit complex containing the SWI1/ADR6, SWI2/SNF2, SWI3, SNF5, and SNF6 gene products isolated from yeast. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(5), 1950–1954.

Cajuso, T., Hänninen, U. A., Kondelin, J., Gylfe, A. E., Tanskanen, T., Katainen, R., et al. (2014). Exome sequencing reveals frequent inactivating mutations in ARID1A, ARID1B, ARID2 and ARID4A in microsatellite unstable colorectal cancer. *International Journal of Cancer*, *135*(3), 611–623. http://doi.org/10.1002/ijc.28705

Cancer Genome Atlas Network. (2012). Comprehensive molecular characterization of human colon and rectal cancer. *Nature*, *487*(7407), 330–337. http://doi.org/10.1038/nature11252

Chandler, R. L., Damrauer, J. S., Raab, J. R., Schisler, J. C., Wilkerson, M. D., Didion, J. P., et al. (2015). Coexistent ARID1A-PIK3CA mutations promote ovarian clear-cell tumorigenesis through pro-tumorigenic inflammatory cytokine signalling. *Nature Communications*, *6*, 6118. http://doi.org/10.1038/ncomms7118

Cibulskis, K. et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnology (2013).doi:10.1038/nbt.2514

Cheng, S. W., Davies, K. P., Yung, E., Beltran, R. J., Yu, J., & Kalpana, G. V. (1999). c-MYC interacts with INI1/hSNF5 and requires the SWI/SNF complex for transactivation function. *Nature Genetics*, *22*(1), 102–105. http://doi.org/10.1038/8811

Chou, A., Toon, C. W., Clarkson, A., Sioson, L., Houang, M., Watson, N., et al. (2014). Loss of ARID1A expression in colorectal carcinoma is strongly associated with mismatch repair deficiency. *Human Pathology*, *45*(8), 1697–1703. http://doi.org/10.1016/j.humpath.2014.04.009

Côté, J., Quinn, J., Workman, J. L., & Peterson, C. L. (1994). Stimulation of GAL4 derivative binding to nucleosomal DNA by the yeast SWI/SNF complex. *Science*, *265*(5168), 53–60.

De Raedt, T., Beert, E., Pasmant, E., Luscan, A., Brems, H., Ortonne, N., et al. (2014). PRC2 loss amplifies Ras-driven transcription and confers sensitivity to BRD4-based therapies. *Nature*, 1–12. http://doi.org/10.1038/nature13561

Donehower, L. A., Harvey, M., Slagle, B. L., McArthur, M. J., Montgomery, C. A., Butel, J. S., & Bradley, A. (1992). Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. *Nature*, *356*(6366), 215–221. http://doi.org/10.1038/356215a0

Dykhuizen, E. C., Hargreaves, D. C., Miller, E. L., Cui, K., Korshunov, A., Kool, M., et al. (2013). BAF complexes facilitate decatenation of DNA by topoisomerase IIα. *Nature*, *497*(7451), 624–627. http://doi.org/10.1038/nature12146

Dynan, W. S. (1989). Modularity in promoters and enhancers. *Cell*, *58*(1), 1–4.

Elfring, L. K., Daniel, C., Papoulas, O., Deuring, R., Sarte, M., Moseley, S., et al. (1998). Genetic analysis of brahma: the Drosophila homolog of the yeast chromatin remodeling factor SWI2/SNF2. *Genetics*, *148*(1), 251–265.

ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, *489*(7414), 57–74. http://doi.org/10.1038/nature11247

Euskirchen, G. M., Auerbach, R. K., Davidov, E., Gianoulis, T. A., Zhong, G., Rozowsky, J., et al. (2011). Diverse roles and interactions of the SWI/SNF chromatin remodeling complex revealed using global approaches. *PLoS Genetics*, *7*(3), e1002008. http://doi.org/10.1371/journal.pgen.1002008

Fearon, E. R. (2011). Molecular Genetics of Colorectal Cancer. *Annual Review of Pathology: Mechanisms of Disease*, *6*(1), 479–507. http://doi.org/10.1146/annurev-pathol-011110-130235

Fearon, E. R., & Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. *Cell*, *61*(5), 759–767.

Fillmore, C. M., Xu, C., Desai, P. T., Berry, J. M., Rowbotham, S. P., Lin, Y.-J., et al. (2015). EZH2 inhibition sensitizes BRG1 and EGFR mutant lung tumours to TopoII inhibitors. *Nature*, 1–4. http://doi.org/10.1038/nature14122

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. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(18), 6656–6661. http://doi.org/10.1073/pnas.0801802105

Garraway, L. A., & Lander, E. S. (2013). Lessons from the Cancer Genome. *Cell*, *153*(1), 17–37. http://doi.org/10.1016/j.cell.2013.03.002

Gebuhr TC, Kovalev GI, Bultman S, Godfrey V, Su L, Magnuson T. The role of Brg1, a catalytic subunit of mammalian chromatin-remodeling complexes, in T cell development.

Gonda, T. J., & Ramsay, R. G. (2015). Directly targeting transcriptional dysregulation in cancer. *Nature Reviews Cancer*, *15*(11), 686–694. http://doi.org/10.1038/nrc4018

Gong, F., Fahy, D., & Smerdon, M. J. (2006). Rad4-Rad23 interaction with SWI/SNF links ATP-dependent chromatin remodeling with nucleotide excision repair. *Nature Structural & Molecular Biology*, *13*(10), 902–907. http://doi.org/10.1038/nsmb1152

Greenson, J. K., Huang, S.-C., Herron, C., Moreno, V., Bonner, J. D., Tomsho, L. P., et al. (2009). Pathologic Predictors of Microsatellite Instability in Colorectal Cancer. *The American Journal of Surgical Pathology*, *33*(1), 126–133. http://doi.org/10.1097/PAS.0b013e31817ec2b1

Gresh L, Bourachot B, Reimann A, Guigas B, Fiette L, Garbay S et al. The SWI/SNF chromatin-remodeling complex subunit SNF5 is essential for hepatocyte differentiation. EMBO J 2005; 24: 3313–3324.

Guinney, J., Dienstmann, R., Wang, X., de Reyniès, A., Schlicker, A., Soneson, C., et al. (2015). The consensus molecular subtypes of colorectal cancer. *Nature Medicine*, *21*(11), 1350–1356. http://doi.org/10.1038/nm.3967

Hamilton SR, Bosman FT, Boffetta P, et al. Carcinoma of the colon and rectum. In: WHO Classification of Tumours of the Digestive System.

Hanahan, D., & Weinberg, R. A. (2000). The hallmarks of cancer. *Cell*, *100*(1), 57–70.

Hara, R., & Sancar, A. (2002). The SWI/SNF chromatin-remodeling factor stimulates repair by human excision nuclease in the mononucleosome core particle. *Molecular and Cellular Biology*, *22*(19), 6779–6787. http://doi.org/10.1128/MCB.22.19.6779-6787.2002

He, F., Li, J., Xu, J., Zhang, S., Xu, Y., Zhao, W., et al. (2015). Decreased expression of ARID1A associates with poor prognosis and promotes metastases of hepatocellular carcinoma. *Journal of Experimental & Clinical Cancer Research*, *34*(1), 80–8. http://doi.org/10.1186/s13046-015-0164-3

Heintzman, N. D., Hon, G. C., Hawkins, R. D., Kheradpour, P., Stark, A., Harp, L. F., et al. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*, *459*(7243), 108–112. http://doi.org/10.1038/nature07829

Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in

the human genome. *Nature Genetics*, *39*(3), 311–318. http://doi.org/10.1038/ng1966

Helming, K. C., Wang, X., & Roberts, C. W. M. (2014a). Vulnerabilities of Mutant SWI/SNF Complexes in Cancer. *Cancer Cell*, *26*(3), 309–317. http://doi.org/10.1016/j.ccr.2014.07.018

Helming, K. C., Wang, X., Wilson, B. G., Vazquez, F., Haswell, J. R., Manchester, H. E., et al. (2014b). ARID1B is a specific vulnerability in ARID1A-mutant cancers. *Nature Medicine*, *20*(3), 251–254. http://doi.org/10.1038/nm.3480

Hirschhorn, J. N., Brown, S. A., Clark, C. D., & Winston, F. (1992). Evidence that SNF2/SWI2 and SNF5 activate transcription in yeast by altering chromatin structure. *Genes & Development*, *6*(12A), 2288–2298.

Hnisz, D., Abraham, B. J., Lee, T. I., Lau, A., Saint-André, V., Sigova, A. A., et al. (2013). Super-enhancers in the control of cell identity and disease. *Cell*, *155*(4), 934–947. http://doi.org/10.1016/j.cell.2013.09.053

Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K., & Sharp, P. A. (2017). A Phase Separation Model for Transcriptional Control. *Cell*, *169*(1), 13–23. http://doi.org/10.1016/j.cell.2017.02.007

Ho, L., Ronan, J. L., Wu, J., Staahl, B. T., Chen, L., Kuo, A., et al. (2009). An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(13), 5181–5186. http://doi.org/10.1073/pnas.0812889106

Hoffman, G.R., Rahal, R., Buxton, F., Xiang, K., McAllister, G., Frias, E., Bag- dasarian, L., Huber, J., Lindeman, A., Chen, D., et al. (2014). Functional epigenetics approach identifies BRM/SMARCA2 as a critical synthetic lethal target in BRG1-deficient cancers. Proc. Natl. Acad. Sci. USA *111*, 3128–3133.

Hu, G., Schones, D. E., Cui, K., Ybarra, R., Northrup, D., Tang, Q., et al. (2011). Regulation of nucleosome landscape and transcription factor targeting at tissue-specific enhancers by BRG1. *Genome Research*, *21*(10), 1650–1658. http://doi.org/10.1101/gr.121145.111

Imbalzano, A. N., Kwon, H., Green, M. R., & Kingston, R. E. (1994). Facilitated binding of TATA-binding protein to nucleosomal DNA. *Nature*, *370*(6489), 481–485.

http://doi.org/10.1038/370481a0

"International Cancer Genome Consortium." *International Cancer Genome Consortium*. 23 Feb. 2017. < http://icgc.org>

Jagani, Z., Mora-Blanco, E. L., Sansam, C. G., McKenna, E. S., Wilson, B., Chen, D., et al. (2010). Loss of the tumor suppressor Snf5 leads to aberrant activation of the Hedgehog-Gli pathway. *Nature Medicine*, *16*(12), 1429–1433. http://doi.org/10.1038/nm.2251

Jass, J. R., Biden, K. G., Cummings, M. C., Simms, L. A., Walsh, M., Schoch, E., et al. (1999). Characterisation of a subtype of colorectal cancer combining features of the suppressor and mild mutator pathways. *Journal of Clinical Pathology*, *52*(6), 455–460.

Jelinic, P., Mueller, J. J., Olvera, N., Dao, F., Scott, S. N., Shah, R., et al. (2014). Recurrent SMARCA4 mutations in small cell carcinoma of the ovary. *Nature Genetics*, *46*(5), 424–426. http://doi.org/10.1038/ng.2922

Jones, P. A., Issa, J.-P. J., & Baylin, S. (2016). Targeting the cancer epigenome for therapy. *Nature Reviews Genetics*, *17*(10), 630–641. http://doi.org/10.1038/nrg.2016.93

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. *Nature Publishing Group*, *45*(6), 592–601. http://doi.org/10.1038/ng.2628

Kadoch, C., Williams, R. T., Calarco, J. P., Miller, E. L., Weber, C. M., Braun, S. M. G., et al. (2017). Dynamics of BAF-Polycomb complex opposition on heterochromatin in normal and oncogenic states. *Nature Publishing Group*, *49*(2), 213–222. http://doi.org/10.1038/ng.3734

Kennison, J. A. (1995). The Polycomb and trithorax group proteins of Drosophila: trans-regulators of homeotic gene function. *Annual Review of Genetics*, *29*(1), 289–303. http://doi.org/10.1146/annurev.ge.29.120195.001445

Kennison, J. A., & Tamkun, J. W. (1988). Dosage-dependent modifiers of polycomb and antennapedia mutations in Drosophila. *Proceedings of the National Academy of Sciences of the United States of America*, *85*(21), 8136–8140.

Kharchenko, P. V, Tolstorukov, M. Y. & Park, P. J. Design and analysis of ChIP-seq experiments for DNA-binding proteins. Nat. Biotechnol. 26, 1351–9 (2008).

Khavari, P. A., Peterson, C. L., Tamkun, J. W., Mendel, D. B., & Crabtree, G. R. (1993). BRG1 contains a conserved domain of the SWI2/SNF2 family necessary for normal mitotic growth and transcription. *Nature*, *366*(6451), 170–174. http://doi.org/10.1038/366170a0

Kheradpour, P. & Kellis, M. Systematic discovery and characterization of regulatory motifs in ENCODE TF binding experiments. Nucleic Acids Res 42, 2976-87 (2014).

Khoury, G., & Gruss, P. (1983). Enhancer elements. *Cell*, *33*(2), 313–314.

Kia, S. K., Gorski, M. M., Giannakopoulos, S., & Verrijzer, C. P. (2008). SWI/SNF mediates polycomb eviction and epigenetic reprogramming of the INK4b-ARF-INK4a locus. *Molecular and Cellular Biology*, *28*(10), 3457–3464. http://doi.org/10.1128/MCB.02019-07

Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 14, R36 (2013).

Kim, K. H., & Roberts, C. W. M. (2016). Targeting EZH2 in cancer. *Nature Medicine*, *22*(2), 128–134. http://doi.org/10.1038/nm.4036

Kim, T. H., & Ren, B. (2006). Genome-wide analysis of protein-DNA interactions. *Annual Review of Genomics and Human Genetics*, *7*(1), 81–102. http://doi.org/10.1146/annurev.genom.7.080505.115634

Kingston, R. E., & Narlikar, G. J. (1999). ATP-dependent remodeling and acetylation as regulators of chromatin fluidity. *Genes & Development*, *13*(18), 2339–2352.

Kinzler, K. W., & Vogelstein, B. (1996). Lessons from hereditary colorectal cancer. *Cell*, *87*(2), 159–170.

Knutson, S. K., Warholic, N. M., Wigle, T. J., Klaus, C. R., Allain, C. J., Raimondi, A., et al. (2013). Durable tumor regression in genetically altered malignant rhabdoid tumors by inhibition of methyltransferase EZH2. *Proceedings of the National Academy of Sciences of*

the United States of America, *110*(19), 7922–7927. http://doi.org/10.1073/pnas.1303800110

Koboldt, D. C., Zhang, Q., Larson, D. E., Shen, D., McLellan, M. D., Lin, L., et al. (2012). VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Research, 22(3), 568–576. http://doi.org/10.1101/gr.129684.111

Kowenz-Leutz, E., & Leutz, A. (1999). A C/EBP beta isoform recruits the SWI/SNF complex to activate myeloid genes. *Molecular Cell*, *4*(5), 735–743.

Kühn, R., Schwenk, F., Aguet, M., & Rajewsky, K. (1995). Inducible gene targeting in mice. *Science*, *269*(5229), 1427–1429.

Kwon, H., Imbalzano, A. N., Khavari, P. A., Kingston, R. E., & Green, M. R. (1994). Nucleosome disruption and enhancement of activator binding by a human SW1/SNF complex. *Nature*, *370*(6489), 477–481. http://doi.org/10.1038/370477a0

la Serna, de, I. L., Carlson, K. A., & Imbalzano, A. N. (2001). Mammalian SWI/SNF complexes promote MyoD-mediated muscle differentiation. *Nature Genetics*, *27*(2), 187–190. http://doi.org/10.1038/84826

Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, R25 (2009).

Laurent, B. C., Treich, I., & Carlson, M. (1993). The yeast SNF2/SWI2 protein has DNA-stimulated ATPase activity required for transcriptional activation. *Genes & Development*, *7*(4), 583–591.

Lawrence, M. S., Stojanov, P., Polak, P., Kryukov, G. V., Cibulskis, K., Sivachenko, A., et al. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*, *499*(7457), 214–218. http://doi.org/10.1038/nature12213

Lee, D., Kim, J. W., Seo, T., Hwang, S. G., Choi, E.-J., & Choe, J. (2002). SWI/SNF complex interacts with tumor suppressor p53 and is necessary for the activation of p53-mediated transcription. *The Journal of Biological Chemistry*, *277*(25), 22330–22337. http://doi.org/10.1074/jbc.M111987200

Lee, R. S., Stewart, C., Carter, S. L., Ambrogio, L., Cibulskis, K., Sougnez, C., et al. (2012). A remarkably simple genome underlies highly malignant pediatric rhabdoid cancers. *Journal of Clinical Investigation*, *122*(8), 2983–2988. http://doi.org/10.1172/JCI64400

Lessard, J., Wu, J. I., Ranish, J. A., Wan, M., Winslow, M. M., Staahl, B. T., et al. (2007). An essential switch in subunit composition of a chromatin remodeling complex during neural development. *Neuron*, *55*(2), 201–215. http://doi.org/10.1016/j.neuron.2007.06.019

Li H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v2 [q-bio.GN]

Li H. and Durbin R. (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics, 26, 589-595. [PMID: 20080505].

Li, J., Lupat, R., Amarasinghe, K. C., Thompson, E. R., Doyle, M. A., Ryland, G. L., et al. (2012). CONTRA: copy number analysis for targeted resequencing. Bioinformatics (Oxford, England), 28(10), 1307–1313. http://doi.org/10.1093/bioinformatics/bts146

Lickert, H., Takeuchi, J. K., Both, Von, I., Walls, J. R., McAuliffe, F., Adamson, S. L., et al. (2004). Baf60c is essential for function of BAF chromatin remodelling complexes in heart development. *Nature*, *432*(7013), 107–112. http://doi.org/10.1038/nature03071

Long, H. K., Prescott, S. L., & Wysocka, J. (2016). Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell*, *167*(5), 1170–1187. http://doi.org/10.1016/j.cell.2016.09.018

Marjou, el, F., Janssen, K.-P., Chang, B. H.-J., Li, M., Hindie, V., Chan, L., et al. (2004). Tissue-specific and inducible Cre-mediated recombination in the gut epithelium. *Genesis (New York, N.Y. : 2000)*, *39*(3), 186–193. http://doi.org/10.1002/gene.20042

Masliah-Planchon, J., Bièche, I., Guinebretière, J.-M., Bourdeaut, F., & Delattre, O. (2015). SWI/SNF chromatin remodeling and human malignancies. *Annual Review of Pathology*, *10*(1), 145–171. http://doi.org/10.1146/annurev-pathol-012414-040445

Matsumoto L, Banine F, Struve J, Xing R, Adams C, Liu Y et al. Brg1 is required for murine neural stem cell maintenance and gliogenesis. Dev Biol 2006; 289: 372–383.

McCart, A. E., Vickaryous, N. K., & Silver, A. (2008). Apc mice: models, modifiers and mutants. *Pathology, Research and Practice*, *204*(7), 479–490. http://doi.org/10.1016/j.prp.2008.03.004

Miller, E. L., Hargreaves, D. C., Kadoch, C., Chang, C.-Y., Calarco, J. P., Hodges, C., et al. (2017). TOP2 synergizes with BAF chromatin remodeling for both resolution and formation of facultative heterochromatin. *Nature Structural & Molecular Biology*, 1–11. http://doi.org/10.1038/nsmb.3384

Mora-Blanco, E. L., Mishina, Y., Tillman, E. J., Cho, Y.-J., Thom, C. S., Pomeroy, S. L., et al. (2014). Activation of β-catenin/TCF targets following loss of the tumor suppressor SNF5. *Oncogene*, *33*(7), 933–938. http://doi.org/10.1038/onc.2013.37

Nagl, N. G., Patsialou, A., Haines, D. S., Dallas, P. B., Beck, G. R., & Moran, E. (2005). The p270 (ARID1A/SMARCF1) subunit of mammalian SWI/SNF-related complexes is essential for normal cell cycle arrest. *Cancer Research*, *65*(20), 9236–9244. http://doi.org/10.1158/0008-5472.CAN-05-1225

Neigeborn, L., & Carlson, M. (1984). Genes affecting the regulation of SUC2 gene expression by glucose repression in Saccharomyces cerevisiae. *Genetics*, *108*(4), 845–858.

Ni, Z., Abou El Hassan, M., Xu, Z., Yu, T., & Bremner, R. (2008). The chromatin-remodeling enzyme BRG1 coordinates CIITA induction through many interdependent distal enhancers. *Nature Immunology*, *9*(7), 785–793. http://doi.org/10.1038/ni.1619

Nie, Z., Xue, Y., Yang, D., Zhou, S., Deroo, B. J., Archer, T. K., & Wang, W. (2000). A specificity and targeting subunit of a human SWI/SNF family-related chromatin-remodeling complex. *Molecular and Cellular Biology*, *20*(23), 8879–8888.

Ostlund Farrants, A. K., Blomquist, P., Kwon, H., & Wrange, O. (1997). Glucocorticoid receptor-glucocorticoid response element binding stimulates nucleosome disruption by the SWI/SNF complex. *Molecular and Cellular Biology*, *17*(2), 895–905.

Park, J.-H., Park, E.-J., Lee, H.-S., Kim, S. J., Hur, S.-K., Imbalzano, A. N., & Kwon, J. (2006). Mammalian SWI/SNF complexes facilitate DNA double-strand break repair by promoting gamma-H2AX induction. *The EMBO Journal*, *25*(17), 3986–3997. http://doi.org/10.1038/sj.emboj.7601291

Pedersen, T. A., Kowenz-Leutz, E., Leutz, A., & Nerlov, C. (2001). Cooperation between C/EBPalpha TBP/TFIIB and SWI/SNF recruiting domains is required for adipocyte differentiation. *Genes & Development*, *15*(23), 3208–3216. http://doi.org/10.1101/gad.209901

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.

Phelan, M. L., Sif, S., Narlikar, G. J., & Kingston, R. E. (1999). Reconstitution of a core chromatin remodeling complex from SWI/SNF subunits. *Molecular Cell*, *3*(2), 247–253.

Roberts, C. W. M., & Orkin, S. H. (2004). The SWI/SNF complex — chromatin and cancer. *Nature Reviews Cancer*, *4*(2), 133–142. http://doi.org/10.1038/nrc1273

Roberts, C. W. M., 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.

Sansam, C. G., Greulich, H., Evans, J. A., Thom, C. S., Moreau, L. A., Biegel, J. A., & Roberts, C. W. M. (2008). Loss of the epigenetic tumor suppressor SNF5 leads to cancer without genomic instability. *Molecular and Cellular Biology*, *28*(20), 6223–6233. http://doi.org/10.1128/MCB.00658-08

Shain, A. H., & Pollack, J. R. (2013). The spectrum of SWI/SNF mutations, ubiquitous in human cancers. *PLoS ONE*, *8*(1), e55119. http://doi.org/10.1371/journal.pone.0055119

Shang, L., Cho, M. T., Retterer, K., Folk, L., Humberson, J., Rohena, L., et al. (2015). Mutations in ARID2 are associated with intellectual disabilities. *Neurogenetics*, *16*(4), 307–314. http://doi.org/10.1007/s10048-015-0454-0

Shlyueva, D., Stampfel, G., & Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. *Nature Reviews Genetics*, *15*(4), 272–286. http://doi.org/10.1038/nrg3682

Stanton, B. Z., Hodges, C., Calarco, J. P., Braun, S. M. G., Ku, W. L., Kadoch, C., et al. (2017). Smarca4 ATPase mutations disrupt direct eviction of PRC1 from chromatin. *Nature*

Publishing Group, *49*(2), 282–288. http://doi.org/10.1038/ng.3735

Stern, M., Jensen, R., & Herskowitz, I. (1984). Five SWI genes are required for expression of the HO gene in yeast. *Journal of Molecular Biology*, *178*(4), 853–868.

Sur, I., & Taipale, J. (2016). The role of enhancers in cancer. *Nature Reviews Cancer*, *16*(8), 483–493. http://doi.org/10.1038/nrc.2016.62

Taketo, M. M., & Edelmann, W. (2009). Mouse models of colon cancer. *Gastroenterology*, *136*(3), 780–798. http://doi.org/10.1053/j.gastro.2008.12.049

Tamkun, J. W., Deuring, R., Scott, M. P., Kissinger, M., Pattatucci, A. M., Kaufman, T. C., & Kennison, J. A. (1992). brahma: a regulator of Drosophila homeotic genes structurally related to the yeast transcriptional activator SNF2/SWI2. *Cell*, *68*(3), 561–572.

"The Cancer Genome Atlas." *National Institutes of Health*. U.S. Department of Health and Human Services. 23 Feb. 2017. <https://cancergenome.nih.gov/>.

Tolstorukov, M. Y., Sansam, C. G., Lu, P., Helming, K. C., Alver, B. H., Tillman, E. J., et al. (2013). Swi/Snf chromatin remodeling/tumor suppressor complex establishes nucleosome occupancy at target promoters. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(25), 10165–10170. http://doi.org/10.1073/pnas.1302209110

Trapnell, C. et al. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat. Biotechnol. 31, 46–53 (2013).

Tsurusaki, Y., Okamoto, N., Ohashi, H., Kosho, T., Imai, Y., Hibi-Ko, Y., et al. (2012). Mutations affecting components of the SWI/SNF complex cause Coffin-Siris syndrome. *Nature Publishing Group*, *44*(4), 376–378. http://doi.org/10.1038/ng.2219

Van Houdt, J. K. J., Nowakowska, B. A., Sousa, S. B., van Schaik, B. D. C., Seuntjens, E., Avonce, N., et al. (2012). Heterozygous missense mutations in SMARCA2 cause Nicolaides-Baraitser syndrome. *Nature Publishing Group*, *44*(4), 445–9– S1. http://doi.org/10.1038/ng.1105

Versteege, I., Sévenet, N., Lange, J., Rousseau-Merck, M. F., Ambros, P., Handgretinger, R., et al. (1998). Truncating mutations of hSNF5/INI1 in aggressive paediatric cancer. *Nature*, *394*(6689), 203–206. http://doi.org/10.1038/28212

Visel, A., Blow, M. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., et al. (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*, *457*(7231), 854–858. http://doi.org/10.1038/nature07730

Vogelstein, B., Papadopoulos, N., Velculescu, V. E., Zhou, S., Diaz, L. A., & Kinzler, K. W. (2013). Cancer genome landscapes. *Science*, *339*(6127), 1546–1558. http://doi.org/10.1126/science.1235122

Wang, W., Côté, J., Xue, Y., Zhou, S., Khavari, P. A., Biggar, S. R., et al. (1996a). Purification and biochemical heterogeneity of the mammalian SWI-SNF complex. *The EMBO Journal*, *15*(19), 5370–5382.

Wang, W., Xue, Y., Zhou, S., Kuo, A., Cairns, B. R., & Crabtree, G. R. (1996b). Diversity and specialization of mammalian SWI/SNF complexes. *Genes & Development*, *10*(17), 2117–2130.

Wang, X., Haswell, J. R., & Roberts, C. W. M. (2014). Molecular pathways: SWI/SNF (BAF) complexes are frequently mutated in cancer--mechanisms and potential therapeutic insights. *Clinical Cancer Research*, *20*(1), 21–27. http://doi.org/10.1158/1078-0432.CCR-13-0280

Wang, X., Lee, R. S., Alver, B. H., Haswell, J. R., Wang, S., Mieczkowski, J., et al. (2016). SMARCB1-mediated SWI/SNF complex function is essential for enhancer regulation. *Nature Publishing Group*. http://doi.org/10.1038/ng.3746

Wang, X., Nagl, N. G., Wilsker, D., Van Scoy, M., Pacchione, S., Yaciuk, P., et al. (2004). Two related ARID family proteins are alternative subunits of human SWI/SNF complexes. *The Biochemical Journal*, *383*(Pt 2), 319–325. http://doi.org/10.1042/BJ20040524

Wang, X., Sansam, C. G., Thom, C. S., Metzger, D., Evans, J. A., Nguyen, P. T. L., & Roberts, C. W. M. (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 Research*, *69*(20), 8094–8101. http://doi.org/10.1158/0008-5472.CAN-09-0733

Wei, X.-L., Wang, D.-S., Xi, S.-Y., Wu, W.-J., Chen, D.-L., Zeng, Z.-L., et al. (2014). Clinicopathologic and prognostic relevance of ARID1A protein loss in colorectal cancer. *World Journal of Gastroenterology*, *20*(48), 18404–18412. http://doi.org/10.3748/wjg.v20.i48.18404

Wilsker, D., Probst, L., Wain, H. M., Maltais, L., Tucker, P. W., & Moran, E. (2005). Nomenclature of the ARID family of DNA-binding proteins. *Genomics*, *86*(2), 242–251. http://doi.org/10.1016/j.ygeno.2005.03.013

Wilson, B. G., & Roberts, C. W. M. (2011). SWI/SNF nucleosome remodellers and cancer. *Nature Reviews Cancer*, *11*(7), 481–492. http://doi.org/10.1038/nrc3068

Wilson, B. G., Helming, K. C., Wang, X., Kim, Y., Vazquez, F., Jagani, Z., et al. (2014). Residual complexes containing SMARCA2 (BRM) underlie the oncogenic drive of SMARCA4 (BRG1) mutation. *Molecular and Cellular Biology*, *34*(6), 1136–1144. http://doi.org/10.1128/MCB.01372-13

Wilson, B. G., Wang, X., Shen, X., McKenna, E. S., Lemieux, M. E., Cho, Y.-J., et al. (2010). Epigenetic antagonism between polycomb and SWI/SNF complexes during oncogenic transformation. *Cancer Cell*, *18*(4), 316–328. http://doi.org/10.1016/j.ccr.2010.09.006

Wu, J. I., Lessard, J., & Crabtree, G. R. (2009). Understanding the words of chromatin regulation. *Cell*, *136*(2), 200–206. http://doi.org/10.1016/j.cell.2009.01.009

Wu, J. N., & Roberts, C. W. M. (2013). ARID1A mutations in cancer: another epigenetic tumor suppressor? *Cancer Discovery*, *3*(1), 35–43. http://doi.org/10.1158/2159-8290.CD-12-0361

Yan, H.-B., Wang, X.-F., Zhang, Q., Tang, Z.-Q., Jiang, Y.-H., Fan, H.-Z., et al. (2014). Reduced expression of the chromatin remodeling gene ARID1A enhances gastric cancer cell migration and invasion via downregulation of E-cadherin transcription. *Carcinogenesis*, *35*(4), 867–876. http://doi.org/10.1093/carcin/bgt398

Ye, J., Zhou, Y., Weiser, M. R., Gonen, M., Zhang, L., Samdani, T., et al. (2014). Immunohistochemical detection of ARID1A in colorectal carcinoma: loss of staining is associated with sporadic microsatellite unstable tumors with medullary histology and high TNM stage. *Human Pathology*, *45*(12), 2430–2436. http://doi.org/10.1016/j.humpath.2014.08.007

Yu, Y., Chen, Y., Kim, B., Wang, H., Zhao, C., He, X., et al. (2013). Olig2 targets chromatin remodelers to enhancers to initiate oligodendrocyte differentiation. *Cell*, *152*(1-2), 248–261. http://doi.org/10.1016/j.cell.2012.12.006