CHARACTERIZING THE ROLE OF THE CHROMATIN REMODELING COMPONENT ARID1A AS A SUPPRESSOR OF SPORADIC MAMMARY TUMORS IN MICE

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Nithya Kartha

August 2017

CHARACTERIZING THE ROLE OF THE CHROMATIN REMODELING COMPONENT ARID1A AS A SUPPRESSOR OF SPORADIC MAMMARY TUMORS IN MICE

Nithya Kartha, Ph. D.

Cornell University 2017

Abstract

The American Cancer Society estimates that in the year 2017, approximately 40,000 women will die from breast cancer. The vast majority of these breast cancer cases (80-85%) are sporadic in nature, developing spontaneously within the lifetime of a woman. While there is a significant amount of knowledge regarding the genetic drivers of hereditary breast cancers, there is very little known about the genes responsible for driving sporadic breast cancers, largely in part due to the dearth of appropriate mouse models of this disease. The C3H-MCM4^{Chaos3}(*Chaos3*) mouse model bears a single endogenous mutation in a gene encoding a component of the MCM2-7 replication helicase, which our lab has previously shown results in a state of chronic replication stress and downstream genomic instability, leading to a strain-specific phenotype of female mice developing spontaneous mammary adenocarcinomas. In my graduate research work, I have utilized this powerful and unique mouse model to determine the genetic drivers of these sporadic mammary tumors (MTs), based on relevance to human breast cancer data available publicly. My analyses of recurrent genomic alterations present in these *Chaos3* MTs revealed that the majority of them (>80%) contained heterozygous deletions of a gene encoding the chromatin remodeling component *Arid1a*. Importantly, *ARID1A* is also frequently deleted (monoallelically) in a significant

subset of human breast cancers (between 30-50%, depending on the specific study cited), based on data from The Cancer Genome Atlas (TCGA). I have characterized the pathways being altered upon overexpression of *Arid1a* in *Chaos3* MT cells, and have identified potential direct transcriptional targets of *Arid1a* regulation using this *in vitro* system. I have further shown that the heterozygous loss of *Arid1a* is a critical maintenance factor for MT growth in this model, and that endogenous induction of *Arid1a* expression to wild-type levels is sufficient to significantly slow down MT cell proliferation *in vitro*. This is suggestive of a haploinsufficient role for *Arid1a* tumor suppression, in a manner similar to TP53, and offers an intriguing therapeutic opportunity of inducing the remaining *ARID1A* allele to potentially reduce MT growth in the subset of human breast cancers that retain an intact copy of this powerful tumor suppressor gene.

BIOGRAPHICAL SKETCH

Nithya was born in San Jose, CA and raised in Bangalore, India. In high school, she volunteered for one summer at a school for children with genetic disorders, which is what piqued her interest in the field of Genetics. She went on to obtain her Bachelors of Science degree in Biotechnology, minoring in Chemistry and Botany, from Bangalore University. Following this, she moved to the United Kingdom to earn a Masters of Science degree in Molecular Genetics, from the University of Leicester. She then returned to the United States to gain valuable work experience in the field of Cancer Genetics, first through two back-to-back internships at Genentech, Inc., and then as a research associate at the University of California, San Francisco. At this point, having decided that she wanted to pursue a Ph.D. in Genetics, she moved from California to Ithaca, NY, to join the graduate program of Genetics, Genomics and Development (GGD) at Cornell University in September 2012. She joined the laboratory of Dr. John Schimenti, to study the genetic drivers of sporadic breast cancer using his lab's mouse model for this disease. In 2013, she was awarded a scholarship from the Centre for Vertebrate Genomics, Cornell University. In the same year, she was also awarded the best cancer research poster award at the International Mammalian Genome Conference held in Bar Harbor, ME. In 2016, she received the Lorraine Flaherty trainee award for best oral presentation at the Allied Genetics Conference held in Orlando, FL. In 2017, she was awarded the LPS best paper award for the field of GGD. After graduating with a Ph.D., she will be starting her post-doctoral research at the Fred Hutchinson Cancer Research Center, in the lab of Dr. Sita Kugel, with a focus on pancreatic cancer and a goal of developing more efficient targeted therapies for this deadly disease.

This dissertation is dedicated to my parents for their unconditional love and support, to cancer patients around the world, and to research scientists of the past, present and future, who dedicate their lives to the pursuit of evidence-based truth.

ACKNOWLEDGMENTS

I would like to first thank Dr. Schimenti for his mentorship and support throughout my graduate dissertation work. He has helped me grow into the scientist I am today, teaching me to always be skeptical of my own results, while at the same time reminding me never to underestimate myself. I will value his advice and guidance for the rest of my scientific career.

Secondly, I would like to acknowledge my committee members, Dr. Robert Weiss and Dr. Scott Coonrod, for their encouragement and input towards my research experiments. Their feedback has greatly helped with the progress I've made on my dissertation project, and their supportive and approachable demeanors have made all my committee meetings a pleasurable and stimulating experience.

Finally, I would like to acknowledge the generous grant we received for this project from the Breast Cancer Coalition of Rochester, which helped fund my final year of research here at Cornell University. Their successful efforts at raising money locally for breast cancer research is an inspiration, reminding me of why I continue to work hard towards finding better treatments for this deadly disease.

TABLE OF CONTENTS

CHAPTER 1 Introduction 1
1.1 Mouse models for breast cancer 1
1.2 The *Chaos3* mouse model for sporadic breast cancer 2
1.3 Relevance of the *Chaos3* mouse model to human breast cancer 4
1.4 Genomic instability as a hallmark of cancer 5
1.5 Epigenetic mechanisms of tumor suppression 7
1.6 References 9

CHAPTER 2 The chromatin remodeling component *Arid1a* is a suppressor of spontaneous mammary tumors in mice 12
2.1 Abstract 13
2.2 Introduction 14
2.3 Results and discussion 16
2.4 Materials and methods 32
2.5 References 37
2.6 Supporting information 40
.
CHAPTER 3 Transcriptional targets of *Arid1a* tumor suppression 46
3.1 Abstract 47
3.2 Introduction 48
3.3 Results 50
3.4 Discussion 60
3.5 Materials and methods 61
3.6 References 63

CHAPTER 4 Induction of *Arid1a* expression as a suppressive mechanism of mammary tumor growth 70
4.1 Abstract 71
4.2 Introduction 72
4.3 Results 73
4.4 Discussion 83
4.5 Materials and methods 85
4.6 References 89

CHAPTER 5 Summary and discussion 92
5.1 Chromatin remodelers in cancer 92
5.2 *ARID1A* Mutations in Human Cancer 93
5.3 Mechanisms of *ARID1A* Tumor Suppression 94
5.4 A potential haploinsufficient role for *ARID1A* tumor suppression 96
5.5 Therapeutic intervention in *ARID1A* mutant cancers 97
5.6 References 99

APPENDIX 107

LIST OF FIGURES

Figure 2-1 *Arid1a* is recurrently deleted in *Chaos3* MTs 18
Figure 2-2 Tumors hemizygous for *Arid1a* have less *Arid1a* mRNA 20
Figure 2-3 Overexpression of *Arid1a* in *Chaos3* MT cell line reduces proliferation rate and prevents tumor growth 23
Figure 2-4 Cell-cycle analysis of *Chaos3* MT cell lines 25
Figure 2-5 Senescence characteristics of AB-C1 cancer cells 28
Figure 2-6 Overexpression of *Arid1a* in the TRP53-deficient MCN1 has no effect on proliferation rate or tumor growth 31
Figure S2-1 SNP genotyping of *Chaos3* MTs 40
Figure S2-2 *ARID1A* copy number in human breast cancers 41
Figure S2-3 IPA analyses of RNA-Seq DE genes 42
Figure 3-1 ATAC-seq peaks identified in *Chaos3* MT cells overexpressing *Arid1a* 52
Figure 3-2 Overlapping gene set representing direct transcriptional targets of *Arid1a* 54
Figure 4-1 *Arid1a* induction using synergistic activation mediators (SAM) 75
Figure 4-2 Candidate guides tested to determine induction efficiency 77
Figure 4-3 Inducing *Arid1a* significantly reduces tumor cell proliferation rates 80
Figure 4-4 siRNA knock-down of *Arid1a* in induced clonal cell line reverses reduced growth phenotype 82
Figure A-A1 Top DE genes with roles in tumorigenesis 111
Figure A-B1 Genotyping results from two rounds of speed congenics 114
Figure A-B2: Schematic matings for *Arid1a* CKO cancer screens 115

LIST OF TABLES

Table S2-1 *Arid1a* CNVs across *Chaos3* MTs and controls determined by ddPCR.....43
Table S2-2 Comparison of *Arid1a* CNVs and expression across individual *Chaos3* MTs.....44
Table S2-3 Primer Sequences.....45
Table 3-1 Potential direct transcriptional targets of *Arid1a* gene regulation... 58
Table 3-2 Cancer-specific associations of genes present within overlap dataset..... 59
Table A1 Differentially expressed genes in *Tln1* mutant mammary gland cells 110

CHAPTER 1

Introduction

As of 2017, breast cancer accounts for 30% of all new cancer diagnoses in women, resulting in the second highest number of cancer-related deaths in women annually in the United States alone[1]. The vast majority of these tumors (~70%) develop spontaneously within the lifetime of a woman, and are characterized as sporadic breast cancers[2]. While extensive research over the past few decades has revealed the most important genetic drivers of hereditary/familial breast cancers, the genetic basis for sporadic breast cancers in women is still unclear. This dissertation will focus on characterizing a potential genetic driver of sporadic breast cancer, *Arid1a*, using a unique mouse model for this disease.

1.1. Mouse models for Breast Cancer

Recent large-scale human cancer genome screens such as The Cancer Genome Atlas (TCGA) have catalogued in great magnitude and detail the genetic alterations found in several different cancer types. While this information has greatly increased our understanding of the comparative landscapes of human tumor genotypes, the significant challenge of differentiating between driver events that play a causal role in carcinogenesis and passenger events that are not required for tumor growth/maintenance still remains. The diversity of human patient populations and the heterogeneity of tumor etiology within these patients renders this kind of analyses near

impossible, which is why the use of mouse models is necessary to be able to further validate/invalidate these candidate cancer genes and elucidate their mechanistic functions.

Genetically engineered mouse models (GEMMs) have contributed extensively to our knowledge and understanding of the function of several important breast cancer genes, including tumor suppressors such as *Brca1/2*, *Tp53* and *Pten*, and oncogenes such as *Erbb2*, *Hras*, *Myc*, and *Ccnd1*[3]. GEMMs of breast cancer have improved greatly through advances in genetic engineering over the years, which have enabled greater spatiotemporal control of mammary tumor onset, most commonly by using mammary-specific promoters (*Wap* or MMTV) to drive Cre-mediated gene deletion/activation. While this has facilitated the targeting of tumor initiating events to a smaller fraction of cells within the mammary gland, it still does not fully recapitulate the natural development of sporadic breast cancers, which are believed to evolve from initiating events within a single cell[4]. Additionally, the vast majority of GEMMs of breast cancer are based on the induction of specific molecular changes that are known to result in sustained cell proliferation, with very little focus on lesser understood hallmarks that are now accepted to be central to cancer development, such as DNA replication stress and genomic instability[5,6].

Thus, there is a scarcity of information available for the genetic/epigenetic basis of sporadic breast carcinogenesis, mainly due to the lack of appropriate mouse models of the disease.

1.2. The *Chaos3* Mouse Model for Sporadic Breast Cancer

The Chromosome Aberrations Occurring Spontaneously 3 (*Chaos3*) mouse model for sporadic breast cancer is unique in that these mice develop spontaneous mammary tumors at a very high frequency without the specific targeting of known breast cancer genes or treatment with external carcinogens.

The model is derived from a single nonsynonymous point mutation in the minichromosome maintenance 4 (*Mcm4*) gene. *Mcm4* is a highly conserved subunit of the MCM2-7 replication helicase complex, which is essential for initiating DNA replication in all eukaryotic cells[7]. The nature of the viable hypomorphic *Chaos3* allele results in destabilization of the MCM2-7 helicase in a manner that compromises normal DNA replication licensing by reducing the number of dormant replication origins[8]. This in turn causes downstream replication stress via the accumulation of stalled replication forks, and ultimately leads to an increase in genomic instability and cancer incidence in mouse strains bearing this allele[8-10].

When congenic in the C3HeB/FeJ (C3H) strain, nearly all female mice with *Chaos3* homozygosity succumb to mammary adenocarcinomas with an average latency of 12 months[11]. The mammary tumor phenotype is strain-specific, with progeny from hybrid crosses of C57BL/6J (B6) mice bearing the *Chaos3* allele resulting in a different spectrum of tumor phenotypes[11]. This observation led to the identification of a nonsynonomous point mutation in the *Talin1* (*Tln1*) gene present in our *Chaos3* model that predisposes to mammary tumor formation, following QTL mapping of susceptibility and resistance loci in C3H x B6 offspring (MD Wallace et al., unpublished; see Appendix). Other studies using this model have shown that

mouse viability is also affected by the genetic background in which the *Chaos3* allele exists, with increased lethality observed in B6 mice, due to strain-specific dynamics between the number and density of licensed and dormant replication origins[10]. Taken together, these genetic and biochemical data in large part help explain the unique mammary tumor phenotype of the *Chaos3* mouse model.

1.3. Relevance of the *Chaos3* Model to Human Breast Cancer

Previous work in the lab has compared gene expression profiles of the *Chaos3* mouse mammary tumors (MTs) with that of other well established GEMMs of breast cancer, and found that the expression signature of *Chaos3* MTs most closely resembles that of mature human mammary luminal cells, and when compared to all the major subtypes of human breast cancer based on clinical diagnoses, the *Chaos3* MT signature was most highly expressed in the luminal A subtype[12]. Luminal A tumors are the most prevalent subtype of human breast cancers, accounting for 64% and 48% of breast cancers diagnosed in populations of white and African American women, respectively[13].

The controlled genetic background and consistent etiology of the *Chaos3* MTs enables the characterization of recurrent genetic events that may play driving roles in mammary carcinogenesis. Array-based Comparative Genomic Hybridization (aCGH) experiments were conducted by previous lab members and myself, with the goal of identifying DNA copy number variations (CNVs) specific to *Chaos3* MTs. These experiments have led to the discovery of CNVs spanning large regions of the genome that are recurrently amplified or deleted within the *Chaos3* MTs. Several of these

regions also correspond to CNVs frequently present in human breast cancers, as characterized by TCGA studies[12,14].

In summary, these data demonstrate the usefulness and relevance of the *Chaos3* mouse model as a means to identify potential drivers of sporadic breast cancer.

1.4 Genomic Instability as a Hallmark of Cancer

Genomic instability (GIN) is predominantly characterized by an increased rate of both structural and numerical chromosomal aberrations, which is found to occur at a much higher frequency in cancer cells relative to normal cells over time[15]. In hereditary cancers, GIN is driven by germline mutations in DNA repair genes such as *BRCA1*, *BRCA2*, *PALB2* and *BRIP1* (breast and ovarian cancers)[16], *NBS1* and *RAD50* (non-Hodgkin lymphoma and brain cancers)[17], *WRN, BLM* and *REQL4* (broad spectrum of cancers)[18], and Fanconi Anaemia genes (acute myeloid leukemia)[19]. These genetic observations in individuals predisposed to cancer strongly supports the mutator hypothesis, which states that GIN may be present in precancerous lesions and can drive tumor progression by increasing the rate of spontaneous muations[20].

In sporadic cancers, the molecular basis and significance of GIN is less clear. Recent advances in the molecular and genetic analysis of the genomes of cancer cells have provided the most compelling evidence for enduring GIN in tumor progression. aCGH experiments in particular have revealed prevalent genomic aberrations (amplifications and deletions) across different tumor types, and importantly have identified recurrent CNVs at specific genomic locations, suggestive of the presence of

genes whose alteration favors neoplastic formation/progression[21]. Specifically, with respect to recurrent genomic deletions in cancer, proteins encoded by genes present within these regions have been found to play important roles in maintaining genomic integrity through the DNA damage response (DDR)[22,23]. Additionally, recent next-generation sequencing of cancer genomes has identified widespread alterations/mutations in known DNA repair genes, such as *BRCA1/BRCA2* (homologous repair genes), *KU70/80* (non-homologous repair genes), and *MLH1/MLH2* (mismatch repair genes) in leukemias, prostate and breast cancers[24-26]. The organ-specific cancer risk associated with these types of genetic aberrations is suggestive of cell-type specific vulnerabilities to extrinsic/intrinsic factors, and more work needs to be done to elucidate these mechanisms.

Thus, there is growing evidence for a model of carcinogenesis in which loss of function of genes involved with the DDR can act as an instigator of GIN, which then leads to the acquisition of additional mutational events that may be selected for during tumor progression. One important set back in proving this model is the limited knowledge we have of all potential "caretaker genes" that play a critical role in maintaining genomic integrity. As technology advances and we garner more information from large-scale cancer genome screens, it will be essential to further characterize the function of the genes identified as being altered/mutated at high frequencies, as this could add to our incomplete understanding of pathways/mechanisms that are involved with protecting the genome against malignant transformation.

For example, only recently have we begun to understand the critical roles that ATPase-dependent mammalian SWItch/ Sucrose Non-Fermentable (mSWI/SNF) chromatin remodeling complexes and their individual components play as "caretakers" of the genome. Multiple studies published in the last five years or so have shown differential involvement of mSWI/SNF subunits in DNA double-strand break repair, non-homologous end joining, replication-associated decatenation, sister chromatid cohesion and DNA damage-induced transcriptional repression[27-30]. Since we know that reduced or loss of function of these important pathways can lead to chromosome mis-segregation/aneupoliody, structural chromosomal instability, and abnormal growth signaling, all of which contribute to carcinogenesis, the loss of mSWI/SNF function may in fact lead to compromised genome integrity as a result of the simultaneous aberration of these crucial pathways[27].

1.5 Epigenetic Mechanisms of Tumor Suppression

In recent years, our knowledge of the processes by which genes are regulated has greatly broadened to include the epigenetic effects of chromatin structure and its impact on gene function. Currently, epigenetics is defined as 'the study of heritable changes in gene expression that occur independent of changes in the primary DNA sequence'[31]. Most of these changes are heritable and stable within cells, and are required during normal mammalian development for establishing and maintaining tissue-specific gene expression. Failure to maintain these inherited epigenetic marks has been found to result in the dysregulation of important signaling pathways, resulting in diseases such as cancer[32].

Epigenetic aberrations in cancer may occur in the form of disordered DNA methylation patterns, histone modifications and/or nucleosome positioning, all of which ultimately result in dysregulated gene expression[33]. TCGA data has revealed that in addition to numerous genetic alterations, tumors of various types also harbor widespread epigenetic alterations at a very high frequency[34,35]. Like genetic changes, epigenetic changes affect cancer development at all stages, and in most cases, they appear to work together to promote cancer progression[33,36,37]. Thus, taken together, the evidence suggests an intertwining of genetic and epigenetic events as the most prevalent model of carcinogenesis.

Relevant to my research is the idea that genetic alterations in epigenetic regulators may lead to an altered epigenome and ultimately cellular transformation. The most compelling evidence for this is the recent discovery that nucleosome modelers are mutated/altered at very high frequencies in human cancers of various types[39-42]. These types of mutations/alterations may have catastrophic effects on gene expression, based on the role of nucleosome modifiers in genome-wide transcriptional regulation[39,40,42]. One of the most recurrently deleted chromosomal regions identified by aCGH analysis of *Chaos3* MTs bears the gene *Arid1a*, which encodes a member of the mSWI/SNF chromatin remodeling complex, the most highly altered/mutated nucleosome modeler in cancer[43-46]. Experimentally, our results have shown that genetic loss of *Arid1a* is a maintenance factor in MT cells derived from this model, and that *Arid1a* may behave as a haploinsufficient tumor suppressor, with dosage effects playing a key role in cell proliferation and growth. Although mounting functional evidence has led to the acceptance of ARID1A as a bona fide tumor

suppressor gene[47,48], the epigenetic extent of its role in altering chromatin structure by contributing to mSWI/SNF activity, modulating gene transcription and ultimately suppressing cancer formation is yet to be fully elucidated. My research work described in the following chapters attempts to shed more light on the tumor suppressive mechanisms of *Arid1a* activity.

1.6 References

1. Siegel, R.L., Miller, K.D. & Jemal, A. Cancer Statistics, 2017. *CA Cancer J Clin* **67**, 7-30 (2017).
2. Filippini, S.E. & Vega, A. Breast cancer genes: beyond BRCA1 and BRCA2. *Front Biosci (Landmark Ed)* **18**, 1358-72 (2013).
3. Vargo-Gogola, T. & Rosen, J.M. Modelling breast cancer: one size does not fit all. *Nat Rev Cancer* **7**, 659-72 (2007).
4. Jonkers, J. & Berns, A. Conditional mouse models of sporadic cancer. *Nat Rev Cancer* **2**, 251-65 (2002).
5. Macheret, M. & Halazonetis, T.D. DNA replication stress as a hallmark of cancer. *Annu Rev Pathol* **10**, 425-48 (2015).
6. Gaillard, H., Garcia-Muse, T. & Aguilera, A. Replication stress and cancer. *Nat Rev Cancer* **15**, 276-89 (2015).
7. Tye, B.K. MCM proteins in DNA replication. *Annu Rev Biochem* **68**, 649-86 (1999).
8. Chuang, C.H., Wallace, M.D., Abratte, C., Southard, T. & Schimenti, J.C. Incremental genetic perturbations to MCM2-7 expression and subcellular distribution reveal exquisite sensitivity of mice to DNA replication stress. *PLoS Genet* **6**, e1001110 (2010).
9. Chuang, C.H. *et al.* Post-transcriptional homeostasis and regulation of MCM2-7 in mammalian cells. *Nucleic Acids Res* **40**, 4914-24 (2012).
10. Kawabata, T. *et al.* A reduction of licensed origins reveals strain-specific replication dynamics in mice. *Mamm Genome* **22**, 506-17 (2011).
11. Shima, N. *et al.* A viable allele of Mcm4 causes chromosome instability and mammary adenocarcinomas in mice. *Nat Genet* **39**, 93-8 (2007).
12. Wallace, M.D. *et al.* Comparative oncogenomics implicates the neurofibromin 1 gene (NF1) as a breast cancer driver. *Genetics* **192**, 385-96 (2012).
13. O'Brien, K.M. *et al.* Intrinsic breast tumor subtypes, race, and long-term survival in the Carolina Breast Cancer Study. *Clin Cancer Res* **16**, 6100-10 (2010).
14. Kartha, N., Shen, L., Maskin, C., Wallace, M. & Schimenti, J.C. The Chromatin Remodeling Component Arid1a Is a Suppressor of Spontaneous Mammary Tumors in Mice. *Genetics* **203**, 1601-11 (2016).

15. Negrini, S., Gorgoulis, V.G. & Halazonetis, T.D. Genomic instability--an evolving hallmark of cancer. *Nat Rev Mol Cell Biol* **11**, 220-8 (2010).
16. Ripperger, T., Gadzicki, D., Meindl, A. & Schlegelberger, B. Breast cancer susceptibility: current knowledge and implications for genetic counselling. *Eur J Hum Genet* **17**, 722-31 (2009).
17. Schuetz, J.M. *et al.* Genetic variation in the NBS1, MRE11, RAD50 and BLM genes and susceptibility to non-Hodgkin lymphoma. *BMC Med Genet* **10**, 117 (2009).
18. Bachrati, C.Z. & Hickson, I.D. RecQ helicases: suppressors of tumorigenesis and premature aging. *Biochem J* **374**, 577-606 (2003).
19. Kennedy, R.D. & D'Andrea, A.D. DNA repair pathways in clinical practice: lessons from pediatric cancer susceptibility syndromes. *J Clin Oncol* **24**, 3799-808 (2006).
20. Loeb, L.A. Mutator phenotype may be required for multistage carcinogenesis. *Cancer Res* **51**, 3075-9 (1991).
21. Hanahan, D. & Weinberg, R.A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-74 (2011).
22. Gao, G. & Smith, D.I. Very large common fragile site genes and their potential role in cancer development. *Cell Mol Life Sci* **71**, 4601-15 (2014).
23. Hazan, I., Hofmann, T.G. & Aqeilan, R.I. Tumor Suppressor Genes within Common Fragile Sites Are Active Players in the DNA Damage Response. *PLoS Genet* **12**, e1006436 (2016).
24. Pritchard, C.C. *et al.* Inherited DNA-Repair Gene Mutations in Men with Metastatic Prostate Cancer. *N Engl J Med* **375**, 443-53 (2016).
25. Nik-Zainal, S. *et al.* Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47-54 (2016).
26. Nilles, N. & Fahrenkrog, B. Taking a Bad Turn: Compromised DNA Damage Response in Leukemia. *Cells* **6**(2017).
27. Brownlee, P.M., Meisenberg, C. & Downs, J.A. The SWI/SNF chromatin remodelling complex: Its role in maintaining genome stability and preventing tumourigenesis. *DNA Repair (Amst)* **32**, 127-33 (2015).
28. Huang, H.T., Chen, S.M., Pan, L.B., Yao, J. & Ma, H.T. Loss of function of SWI/SNF chromatin remodeling genes leads to genome instability of human lung cancer. *Oncol Rep* **33**, 283-91 (2015).
29. Chai, B., Huang, J., Cairns, B.R. & Laurent, B.C. Distinct roles for the RSC and Swi/Snf ATP-dependent chromatin remodelers in DNA double-strand break repair. *Genes Dev* **19**, 1656-61 (2005).
30. Allard, S., Masson, J.Y. & Cote, J. Chromatin remodeling and the maintenance of genome integrity. *Biochim Biophys Acta* **1677**, 158-64 (2004).
31. Sharma, S., Kelly, T.K. & Jones, P.A. Epigenetics in cancer. *Carcinogenesis* **31**, 27-36 (2010).
32. Egger, G., Liang, G., Aparicio, A. & Jones, P.A. Epigenetics in human disease and prospects for epigenetic therapy. *Nature* **429**, 457-63 (2004).
33. You, J.S. & Jones, P.A. Cancer genetics and epigenetics: two sides of the same coin? *Cancer Cell* **22**, 9-20 (2012).

34. Gnad, F., Doll, S., Manning, G., Arnott, D. & Zhang, Z. Bioinformatics analysis of thousands of TCGA tumors to determine the involvement of epigenetic regulators in human cancer. *BMC Genomics* **16 Suppl 8**, S5 (2015).
35. Shah, M.A., Denton, E.L., Arrowsmith, C.H., Lupien, M. & Schapira, M. A global assessment of cancer genomic alterations in epigenetic mechanisms. *Epigenetics Chromatin* **7**, 29 (2014).
36. Jones, P.A. & Laird, P.W. Cancer epigenetics comes of age. *Nat Genet* **21**, 163-7 (1999).
37. Sandoval, J. & Esteller, M. Cancer epigenomics: beyond genomics. *Curr Opin Genet Dev* **22**, 50-5 (2012).
38. Feinberg, A.P., Ohlsson, R. & Henikoff, S. The epigenetic progenitor origin of human cancer. *Nat Rev Genet* **7**, 21-33 (2006).
39. Nair, S.S. & Kumar, R. Chromatin remodeling in cancer: a gateway to regulate gene transcription. *Mol Oncol* **6**, 611-9 (2012).
40. Skulte, K.A., Phan, L., Clark, S.J. & Taberlay, P.C. Chromatin remodeler mutations in human cancers: epigenetic implications. *Epigenomics* **6**, 397-414 (2014).
41. Yaniv, M. Chromatin remodeling: from transcription to cancer. *Cancer Genet* **207**, 352-7 (2014).
42. Zhang, C., Lu, J. & Zhang, P. The Roles of Chromatin Remodeling Proteins in Cancer. *Curr Protein Pept Sci* **17**, 446-54 (2016).
43. Biegel, J.A., Busse, T.M. & Weissman, B.E. SWI/SNF chromatin remodeling complexes and cancer. *Am J Med Genet C Semin Med Genet* **166c**, 350-66 (2014).
44. Kadoch, C. & Crabtree, G.R. Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics. *Sci Adv* **1**, e1500447 (2015).
45. Masliah-Planchon, J., Bieche, I., Guinebretiere, J.M., Bourdeaut, F. & Delattre, O. SWI/SNF chromatin remodeling and human malignancies. *Annu Rev Pathol* **10**, 145-71 (2015).
46. Wilson, B.G. & Roberts, C.W. SWI/SNF nucleosome remodellers and cancer. *Nat Rev Cancer* **11**, 481-92 (2011).
47. Wu, J.N. & Roberts, C.W. ARID1A mutations in cancer: another epigenetic tumor suppressor? *Cancer Discov* **3**, 35-43 (2013).
48. Wu, R.C., Wang, T.L. & Shih Ie, M. The emerging roles of ARID1A in tumor suppression. *Cancer Biol Ther* **15**, 655-64 (2014).

CHAPTER 2

The Chromatin Remodeling Component *Arid1a* is a Suppressor of Spontaneous Mammary Tumors in Mice

Nithya Kartha[1], Lishuang Shen[2], Carolyn Maskin[1], Marsha Wallace[3] and John C. Schimenti[1]

Affiliations:

[1] Cornell University, College of Veterinary Medicine, Dept. of Biomedical Sciences, Ithaca, NY 14853, USA

[2] Current address: Children's Hospital of Los Angeles, Los Angeles, CA, 90057

[3] Current address: Ludwig Institute for Cancer Research, University of Oxford, OXF, GB, OX3 7DQ

Author Contributions:

NK and JCS conceived and designed the experiments. NK performed all experiments, with assistance from CM on q-RT-PCR. LS conducted aCGH data analyses. NK and JCS wrote the manuscript.

This chapter was published in Genetics, and was reprinted with permission here.

Kartha, N., Shen, L., Maskin, C., Wallace, M. & Schimenti, J.C. The Chromatin Remodeling Component Arid1a Is a Suppressor of Spontaneous Mammary Tumors in Mice. *Genetics* **203**, 1601-11 (2016).

2.1 Abstract

Human cancer genome studies have identified the SWI/SNF chromatin remodeling complex member *ARID1A* as one of the most frequently altered genes in several tumor types. Its role as an ovarian tumor suppressor has been supported in compound knockout mice. Here, we provide genetic and functional evidence that *Arid1a* is a bona fide mammary tumor suppressor, using the *Chaos3* (Chromosome aberrations occurring spontaneously 3) mouse model of sporadic breast cancer. About 70% of mammary tumors that formed in these mice contained a spontaneous deletion removing all or part of one *Arid1a* allele. Restoration of *Arid1a* expression in a *Chaos3* mammary tumor line with low *Arid1a* levels greatly impaired its ability to form tumors following injection into cleared mammary glands, indicating that ARID1A insufficiency is crucial for maintenance of these *Trp53*-proficient tumors. Transcriptome analysis of tumor cells before and after re-introduction of *Arid1a* expression revealed alterations in growth signaling and cell-cycle checkpoint pathways, in particular the activation of the TRP53 pathway. Consistent with the latter, *Arid1a* re-expression in tumor cells led to increased *p21* (*Cdkn1a*) expression and dramatic accumulation of cells in G2 phase of the cell cycle. These results not only provide *in vivo* evidence for a tumor suppressive and/or maintenance role in breast cancer, but also indicate a potential opportunity for therapeutic intervention in ARID1A-deficient human breast cancer subtypes that retain one intact copy of the gene and also maintain wild-type TRP53 activity.

2.2 Introduction

The identification of genetic drivers of specific types and subtypes of cancer continues to be an important goal of cancer biology research. Major efforts including The Cancer Genome Atlas (TCGA) project have catalogued mutations in diverse human tumors. This wealth of data has been instrumental in identifying genes that may be playing a direct or indirect role in carcinogenesis by virtue of them being commonly altered in a particular cancer type. However, proving causality of these candidate "driver" genes, and elucidating their roles in tumorigenesis, requires relevant experimental validation.

*ARID1A (*also called *BAF250a*), encoding an important component of the mammalian SWI/SNF complex, has emerged as one of the most commonly mutated or downregulated genes in diverse tumors, including gastrointestinal[1,2], endometrial[3,4], ovarian clear cell[5,6], pancreatic[7], lung[8] and breast[9,10]. ARID1A impacts epigenetic gene regulation by altering chromatin structure around promoters of specific loci in conjunction with its associated SWI/SNF complex components[11,12]. Therefore, its downregulation or mutation in somatic cells can have profound consequences including inappropriate proliferation[13]. Despite the accumulating correlative data implicating *ARID1A* as a tumor suppressor, functional proof has been lacking in part due to the fact that knockout of *Arid1a* in mice causes embryonic lethality even in the heterozygous state[14]. However, two recent reports have shown that conditional biallelic knockout of *Arid1a* in ovarian surface epithelial cells, in conjunction with either conditional expression of a mutant phosphoinositide 3-kinase

catalytic subunit PIK3CA[15], or conditional disruption of *Pten*[16], caused carcinomas resembling clear cell in the former, and endometriod/undifferentiated in the latter. In both studies, deletion of *Arid1a* alone, or deletion of only one *Arid1a* allele in the compound mutant situations, was insufficient to cause cancer. While these studies provided compelling evidence for the tumor suppressive role of *Arid1a* in ovarian cancer, they (and most other genetically-engineered cancer models) do not model the process of sporadic cancer development. Furthermore, the dependency of bi-allelic *Arid1a* inactivation upon mutation of *Pten* or *Pik3ca* in driving tumor formation in these models seems to be specific to the pathogenesis of endometrium-related ovarian neoplasms[17] and does not appear to apply to several of the other human cancers in which *ARID1A* is commonly mutated[18]. Thus, it is important to validate cancer genes/ pathways in the context of their tumor-type-specific environments, as the behavior of these genes and pathways may vary by tissue type. Sporadic breast cancer (i.e., not associated with inherited neoplasia-driving mutations) accounts for the vast majority of breast cancer cases worldwide (90-95%)[19]. Although *ARID1A* has not yet been widely recognized as key suppressor of breast carcinogenesis, it is heterozygously deleted in a substantial fraction of tumors[9,10], and low *ARID1A* expression in tumors of breast cancer patients correlates significantly with poorer prognosis and overall survival[20]. Here, we report functional evidence that *Arid1a* loss is critical for mammary tumorigenesis in a mouse model of spontaneous breast cancer, and present data on how this leads to deregulated cancer cell growth.

2.3 Results and Discussion

The *Chaos3* mouse, bearing a missense allele (*Mcm4*^{Chaos3}) of the DNA replication gene *Mcm4*, exhibits high levels of genomic instability caused by the mutation's destabilization of the MCM2-7 replicative helicase complex [21-23]. Most females homozygous for the *Chaos3* mutation congenic in the C3HeB/FeJ strain background ("*Chaos3*") develop spontaneous mammary tumors (MTs) with an average latency of twelve months[21]. Array Comparative Genomic Hybridization (aCGH) analyses of 9 *Chaos3* MTs revealed interstitial deletions common to a small number of chromosomal regions[24]. Almost all tumors were missing both copies of *Nf1*, a tumor suppressor that functions as negative regulator of *Ras*, but positive for *Trp53*[25].

Those aCGH data, plus an additional 12 reported here, indicated that most (18/21) MTs also contained deletions involving part or all of an ~100kb region on Chr4 containing *Arid1a* (Figure 2-1). To further validate the aCGH results, we performed digital droplet PCR (ddPCR) on DNA from the same 12 MTs + 3 non-MTs using probes situated at both ends of *Arid1a*. All 15 calls for Probe 2, and 13/15 for Probe 1 (Figure 2-1A) were consonant. The two discrepancies were in MTs 7 and 12, which according to aCGH results, have an identical deletion breakpoint within *Arid1a*. It is possible that in these cases, the breakpoint is actually proximal to that called by the software. As an alternative confirmation of *Arid1a* hemizygosity in these tumors, we took advantage of genetic polymorphisms in two F1 (C3HeB/FeJ x C57BL/6J) MTs deleted for *Arid1a* (#1 and #8, Figure 2-1) and an F2 MT having no deletion (#2

in Figure 2-1), based on aCGH calls. Genotyping of SNPs at the 3' end of *Arid1a* revealed agreement with the aCGH and ddPCR data (Figure 2-1A; S2-1).

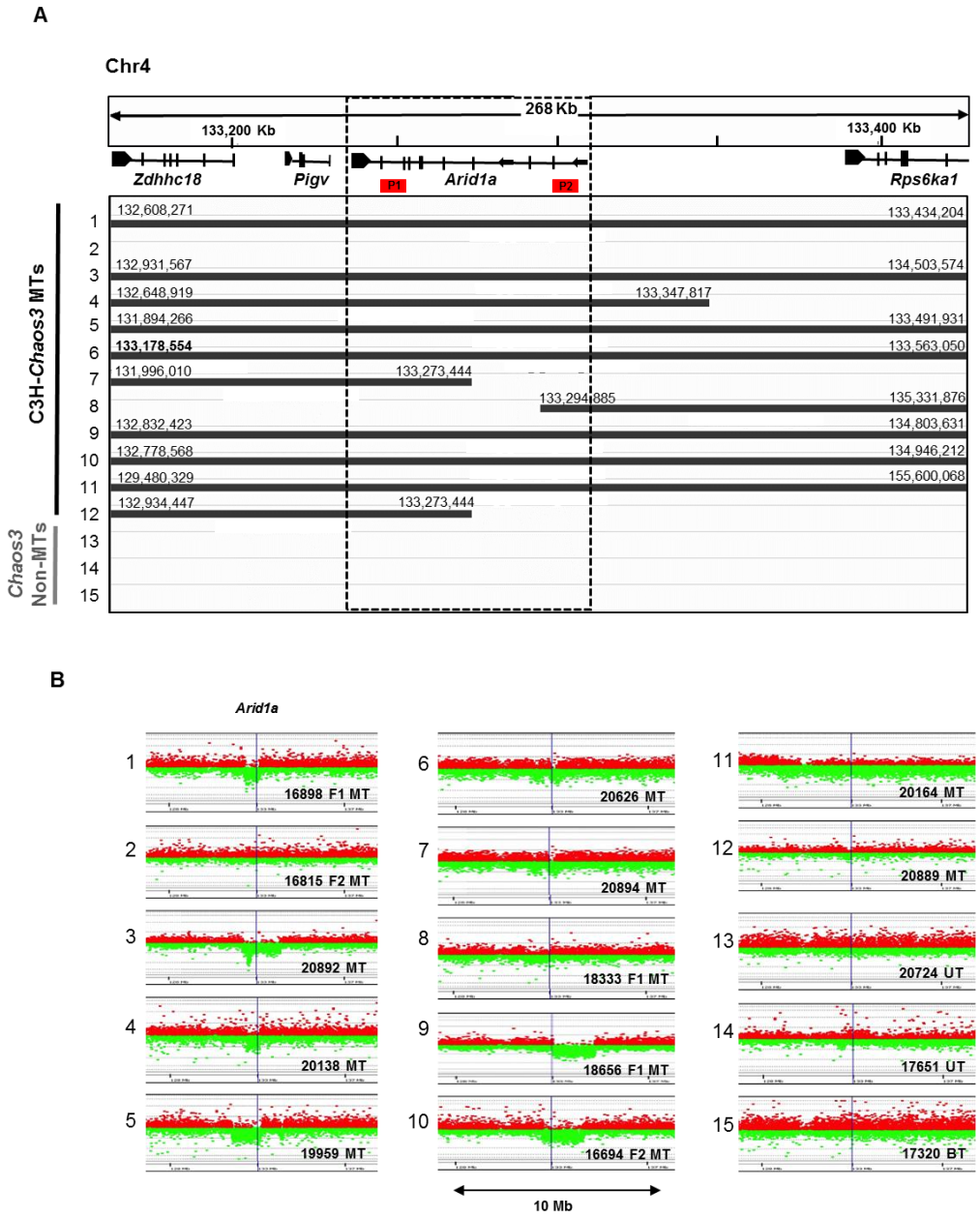


Figure 2-1: *Arid1a* is recurrently deleted in *Chaos3* mammary tumors. (A) Overview of aCGH data near the *Arid1a* locus from 15 tumor samples, adapted from an IGV depiction. Solid lines denote stretches of contiguous probes with reduced hybridization signal, thus representing deleted regions. Nucleotide coordinates of deletion endpoints are indicated, and correspond to the last probe with reduced hybridization signal on the array. MT = mammary tumor. The control non-MTs consist of 2 uterine tumors and 1 bone tumor. (B) Plot of probe intensities near the *Arid1a* from aCGH hybridization. Each dot is a probe on the array, with the green and red representing control vs tumors, respectively. Locations of primer pairs used for *Arid1a* CNV analyses by ddPCR are depicted in the figure as P1 and P2 (see Methods).

We next scored 33 additional *Chaos3* MTs and 5 cell lines derived from *Chaos3* MTs for deletions in the *Arid1a* coding region by ddPCR. In total, ~70% of the *Chaos3* MTs analyzed had monoallelic deletions for all or part of *Arid1a* (Table S2-1). Hemizygosity for *ARID1A* also appears to be common in human breast carcinomas at frequencies as high as ~40% depending on the dataset[26-28] (Fig. S2-2).

If hemizygosity of *Arid1a* contributes to tumorigenesis, then either it is haploinsufficient (i.e., 50% expression contributes to the transformed state), or the non-deleted allele is also altered in a genetic or epigenetic manner that reduces *Arid1a* expression to a level below that which is necessary to prevent transformation and/or tumor growth. To test this, we quantified *Arid1a* mRNA in hemizygous and non-deleted *Chaos3* MTs. On average, transcript levels in 24 *Arid1a*-deleted tumors, but not non-deleted tumors, averaged about half that present in WT mammary tissue (Fig. 2-2; Table S2-2). The results suggest that ARID1A reduction, but not elimination, may contribute to tumorigenesis or tumor maintenance. Interestingly, the two *Chaos3* non-mammary tumors tested had ~5 fold more *Arid1a* than the deleted MTs (Fig. 2-2).

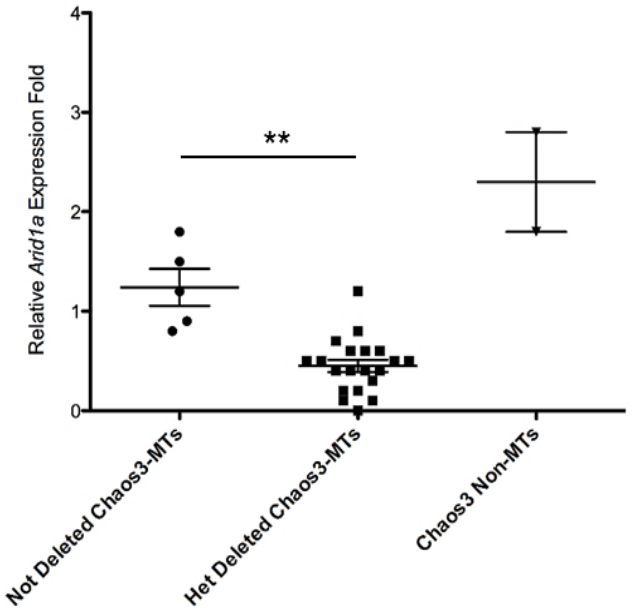


Figure 2-2. Tumors hemizygous for *Arid1a* have less *Arid1a* mRNA.
Plotted are qRT-PCR expression data from the *Chaos3* non-MTs (two uterine tumors) and *Chaos3* MTs that were either heterozygously deleted (n = 21) or not deleted (n = 5) for ARID1A, based on ddPCR data. Expression levels for each sample were calculated relative to an average of two WT RNA tissue samples. Significant differences were calculated using a two-tailed Student's t-test (** $P < 0.001$).

The genetic and molecular data described above implicate, but do not prove, that decreased ARID1A expression is involved in either neoplastic transformation or maintenance of the transformed state. To address this question functionally, we conducted experiments to restore *Arid1a* expression in ARID1A-deficient *Chaos3* MT cells, followed by analyses of the *in vitro* and *in vivo* consequences. First, we generated a *Chaos3* mammary tumor cell line (23116 MT) that has one copy of *Arid1a* deleted (Table S2-1) and very low levels of *Arid1a* expression (Fig. 2-3A, B), then stably introduced an *Arid1a* cDNA expression construct into these cells using lentivirus-mediated transduction. These transformed lines were termed "Addback" (AB) cells. We then assessed cell proliferation activity via EdU incorporation in the parental vs. three transduced cell lines, and found that ectopic *Arid1a* expression caused a dramatic decrease (~3 fold) in EdU incorporation in each of the lines tested (Fig. 2-3C).

To determine if ARID1A is required for tumorigenicity, we tested whether one of the transduced clones (AB-C1) exhibiting elevated levels of mRNA (Fig. 2-3A) and protein (Fig. 2-3B) would reduce/abolish the ability of 23116 MT cells to form tumors when transplanted into host animals. The parental and AB-C1 cancer cells were injected into cleared mammary fat pads of WT C3H female mice (23116 MT on one side, and AB-C1 on the other of each mouse, see Materials and Methods), and monitored for tumor formation. Overexpression of *Arid1a* significantly decreased mammary tumor formation frequency and size (Fig. 2-3D). These results indicate that loss/reduction of *Arid1a* expression is crucial for the growth and/or formation of *Chaos3* mammary tumors. As shown below, ectopic *Arid1a* overexpression did not

inhibit tumor formation in an unrelated non-*Chaos3* MT cell line MCN1, indicating that excessive ARID1A itself is not cell toxic.

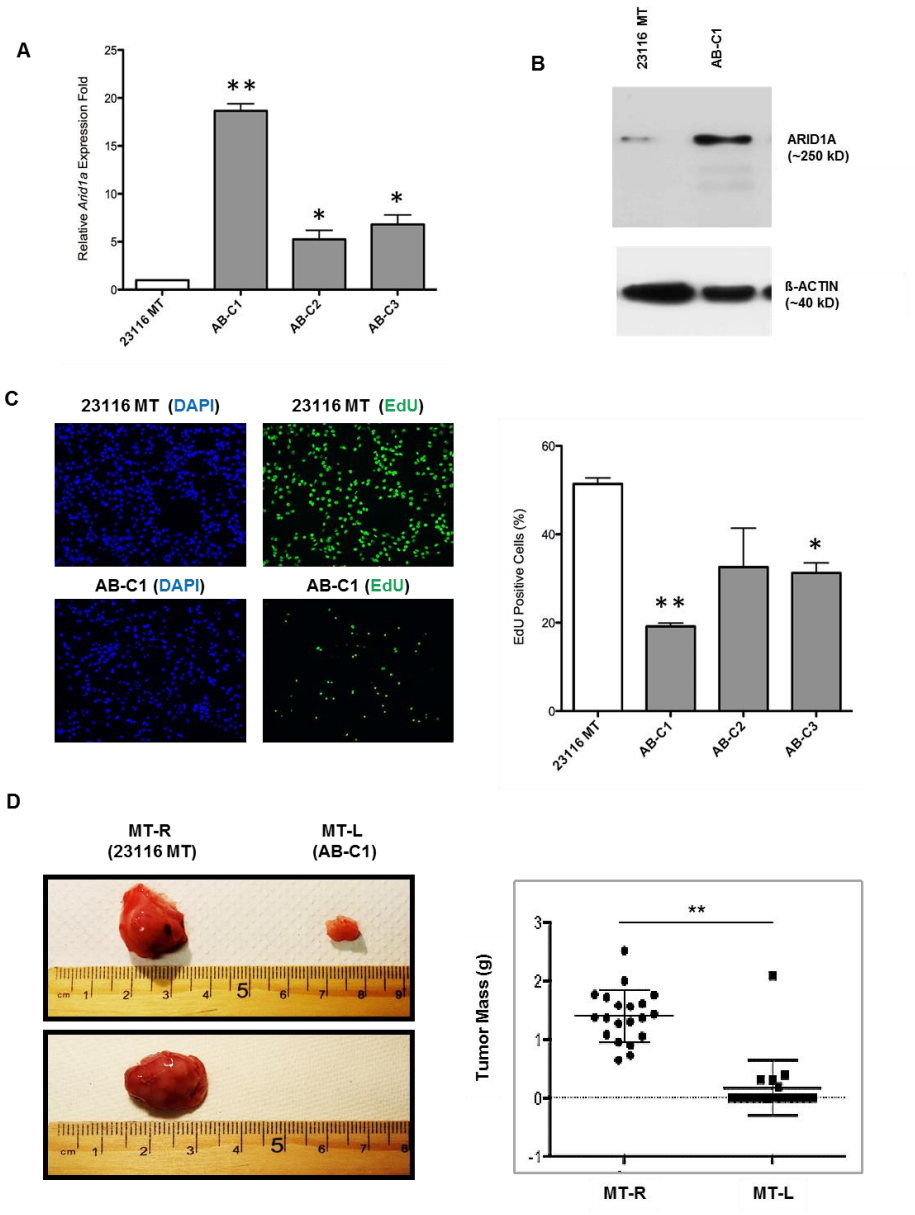


Figure 2-3. Overexpression of *Arid1a* in *Chaos3* MT cell line reduces proliferation rate and prevents tumor growth. (A) Arid1a expression levels quantified by qRT-PCR in untransduced Chaos3 MT cell line (23116 MT) and three individual clonal lines (AB-C1, AB-C2, and AB-C3) transduced with the Arid1a expression vector. Results are shown as the mean 6 SEM of three technical replicates. Significant differences were calculated using a two-tailed Student's t-test (* $P < 0.05$; ** $P < 0.001$). Values are relative to untransduced parental cell line 23116 MT. (B) Arid1a expression levels quantified by immunoblotting in indicated cell lines. (C) EdU incorporation assays of 23116 MT vs. AB clones (C1–C3). Error bars signify the SEM of three experimental replicates, with >1000 cells counted per sample for each replicate. Significant differences were calculated using a two-tailed Student's t-test.

To gain insight into the mechanisms by which *Arid1a* loss promotes tumorigenesis in the *Chaos3* model, we considered data showing that *Arid1a* is required for efficient functioning of the DNA damage response (DDR), specifically the G2/M cell-cycle checkpoint that helps suppress genomic instability (GIN) and tumorigenesis[29,30]. Since *Chaos3* cells have chronic replication stress and GIN[21,23,31], it is possible that a loss or reduction of ARID1A in a cell allows escape from DDR-mediated growth arrest or apoptosis, thus promoting carcinogenesis. Therefore, we examined the cell cycle of AB-C1 cultures. This revealed an accumulation of cells in the G2 phase (Fig. 2-4A & B), suggesting that *Arid1a* overexpression might be inducing a checkpoint response and consequent growth arrest.

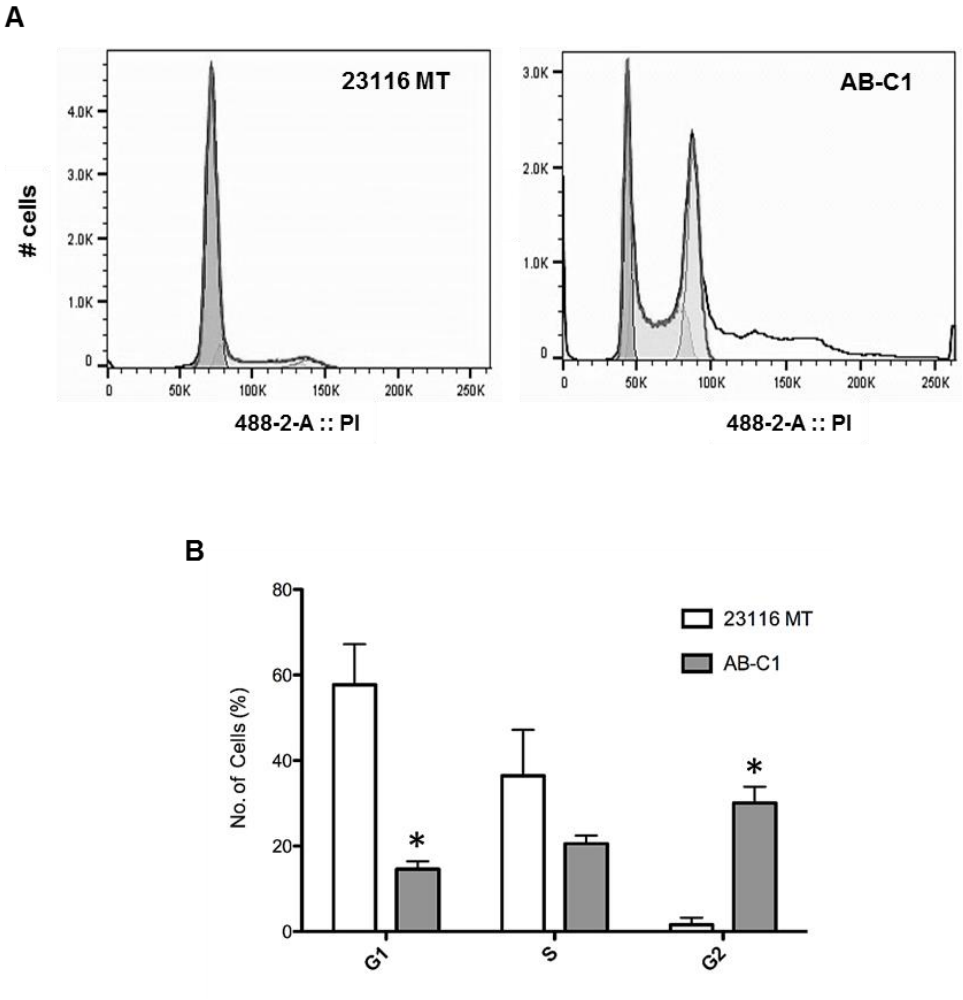


Figure 2-4. Cell-cycle analysis of *Chaos3* mammary tumor cell lines. (A) Cell-cycle profiles of 23116 and AB-C1 mammary tumor lines. (B) Percentage of cells in different phases of the cell cycle (G1, S, and G2) based on FloJo statistical analyses. Significant differences were calculated using a two-tailed Student's t-test (* $P < 0.05$; ** $P < 0.001$). Values are relative to untransduced parental cell line 23116 MT.

Since ectopic *Arid1a* expression in ARID1A-deficient MT cells caused cell cycle arrest, we assessed whether either senescence or apoptosis was triggered as a consequence. TUNEL assays did not reveal a significant increase in apoptosis, based on relative percentages of positively stained cells (23116 MT = 0.6%, ± 0.2% S.E.M.; AB-C1 = 1.5%, ± 0.4%), raising the possibility that these cells were instead senescing. Another indication was the dramatic change in morphological features of the cells in which *Arid1a* was overexpressed. They appeared larger and flatter (Fig. 5A), characteristic of cells undergoing senescence [32]. Finally, we conducted senescence-associated beta-galactosidase (SABG) assays, showing that the AB-C1 cell population had nearly 10 fold more SABG-positive cells than the parental cultures (Fig. 2-5B & C).

Multiple studies have demonstrated that DNA damage-induced G2 arrest activates cellular senescence in a TRP53- and p21-dependent manner[33-35]. mRNA levels of *p21*, which is transcriptionally regulated by TRP53, was ~5-fold higher in AB-C1 MT cells compared to the 23116 parental line (Fig. 2-5D). Taken together, these data indicate that restoring or overexpressing *Arid1a* in *Chaos3* MT cells enables G2/M arrest and subsequent cellular senescence. This is consistent with data indicating that ARID1A functions as both a "gatekeeper" in its control of cell proliferation, and a "caretaker" in its maintenance of genomic integrity [36].

To better understand the mechanism by which restoration/overexpression of *Arid1a* impairs growth and tumorigenesis, we conducted RNA-seq comparing the transcriptomes of 23116 MT vs AB-C1 and AB-C2 cells. A total of 554 genes were significantly differentially expressed (DE) between the parental stock and the *Arid1a*-

transduced lines (FPKM > 5; Log_2 Fold > 1; Table S3). Ingenuity Pathway Analysis (IPA) of these DE genes revealed that the TRP53 pathway was activated in AB-C1/C2 cells, while the TGF Beta pathway was repressed (Fig. 2-S3).

RNA-seq data also showed that the most highly upregulated genes within the TRP53 pathway in AB-C1 cells were *Igfbp5*, *Igfbp2* and *Serpinb2* (*Pai-2*). We validated these data by qRT-PCR (Fig. 2-5E). All three genes have been implicated in tumor growth suppression[37-39]. *IGFBP5* was found to activate cellular senescence through a TRP53-dependent mechanism in human endothelial cells[40]. These data further support the idea of a TRP53-dependent senescence checkpoint response being activated when *Arid1a* is overexpressed in these cancer cells.

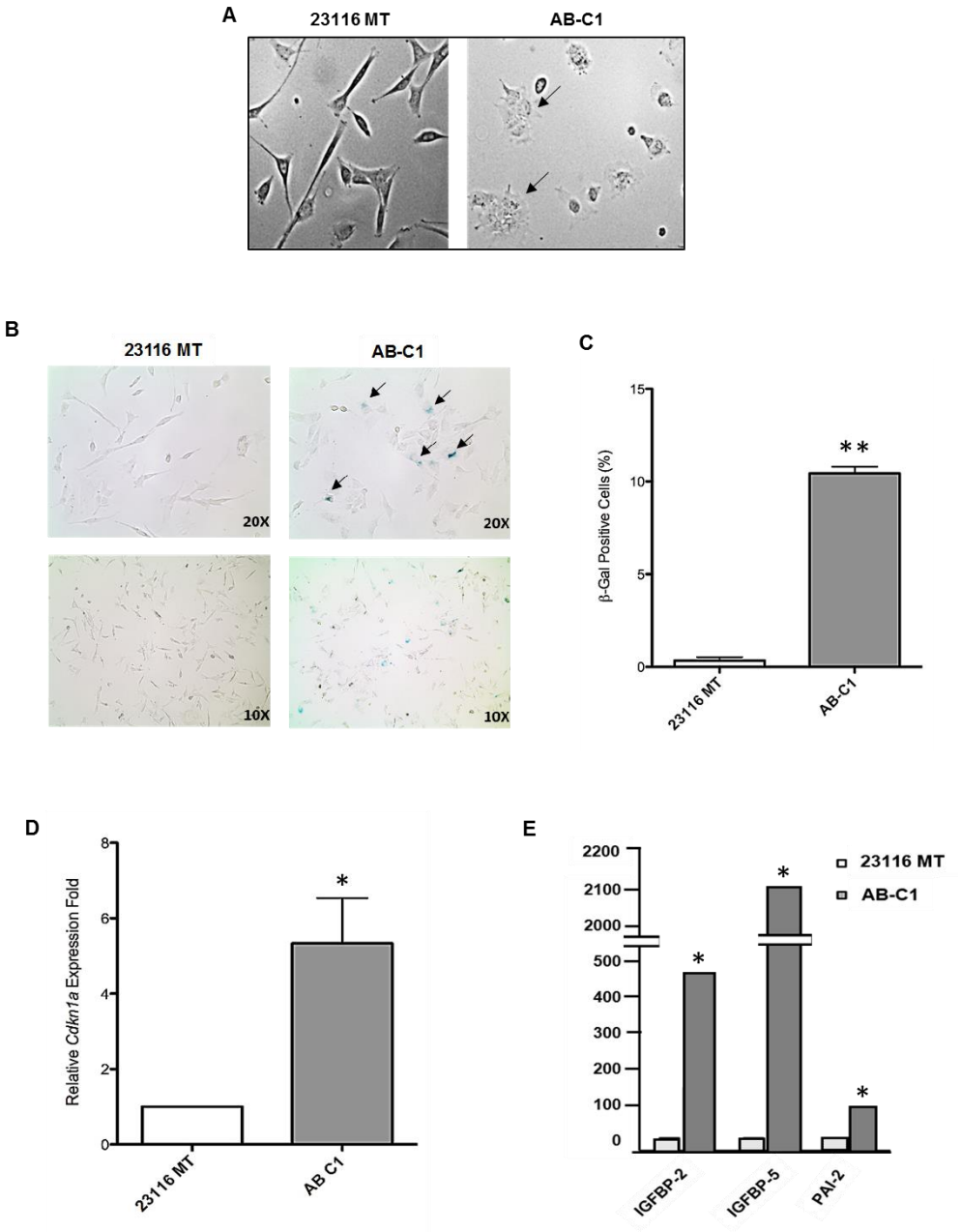


Figure 2-5. Senescence characteristics of AB-C1 cancer cells. (A) Morphological comparison of 23116 MT vs. AB-C1 cells. (B) Representative images of indicated cells stained for senescence-associated ß-galactosidase activity. (C) Average percentages of positively stained cells (blue color) calculated from four technical replicates. (D) qRT-PCR validation of senescence-associated genes. (E) qRT-PCR assay comparing relative transcript levels of p21 in AB-C1 vs. 23116 MT cells. Significant differences were calculated using a two-tailed Student's t-test (* $P < 0.05$; ** $P < 0.001$). Values are relative to untransduced parental cell line 23116 MT.

TRP53 was reported to interact physically with ARID1A and the rest of the SWI/SNF complex, enabling transcriptional regulation of downstream gene targets 41. Several human cancer studies have found that loss of ARID1A expression correlates with high amounts of potentially functionally inactive TRP53[2,42-44], suggesting they have co-dependent tumor suppressive functions, and that ARID1A deficiency would have a similar effect as losing TRP53 activity. A similar mutually exclusive pattern of Trp53 and *Arid1a* expression seems to exist in *Chaos3* MTs. They exhibit high levels of TRP53[25], but it does not seem to be effective or functional in the sense of its established role in suppressing uncontrolled cell growth or malignant transformation. Based on this hypothesis as well as the RNA-seq data, it is possible that re-expressing Arid1a in the *Chaos3* tumor cells restores the ability of TRP53 to regulate downstream target genes.

To test this hypothesis, we overexpressed *Arid1a* in a *Trp53* null mouse mammary tumor cell line MCN1[45], which we found to not only have low *Arid1a* expression (Figure 2-6A, B), but also apparent hemizygosity of *Arid1a* (Table S2-1). EdU incorporation assays comparing the untransduced and AB cells showed that unlike the TRP53-proficient *Chaos3* cell line 23116, proliferation was unaltered upon overexpressing *Arid1a* in MCN1 (Figure 2-6C), and subsequent mammary fat pad growth assays revealed that tumorigenecity *in vivo* was also unaffected (Figure 2-6D). This is consistent with the hypothesis that active TRP53 is necessary for ARID1A to function in its tumor suppressive role.

Chaos3 tumors have a manageable number of recurring spontaneous alterations, making it feasible to unravel the molecular events leading to tumor

formation – a crucial question in cancer genetics. The experiments here provide genetic and functional evidence that ARID1A is a suppressor of mammary tumorigenesis, and particularly, that it is required for maintenance of tumorigenic potential as revealed by transplantation assays. This role in tumor maintenance also appears to be the case in human ovarian cancer, where it was shown that re-introduction of the gene into tumor cells bearing *ARID1A* mutations inhibited xenograft growth[41]. Our finding that *Arid1a* is almost exclusively mono-allelically (not bi-allelically) deleted in the *Chaos3* model of spontaneous breast cancer, apparently similar to the situation in human breast cancers, is important in terms of potential therapeutic intervention. We showed that over-expressing *Arid1a* ectopically in MT cells greatly suppresses tumor growth in a TRP53-dependent manner. Therefore, in breast cancer cases that retain an intact *ARID1A* allele in *trans* to a mutant/deleted allele, and also contain wild-type *TRP53*, it may be possible to employ methods for specific reactivation of the remaining allele, thus suppressing tumor growth and triggering cell-cycle arrest. Recent development of sequence-specific, chimeric transcriptional regulators offers one such potential avenue to accomplish this[46-50].

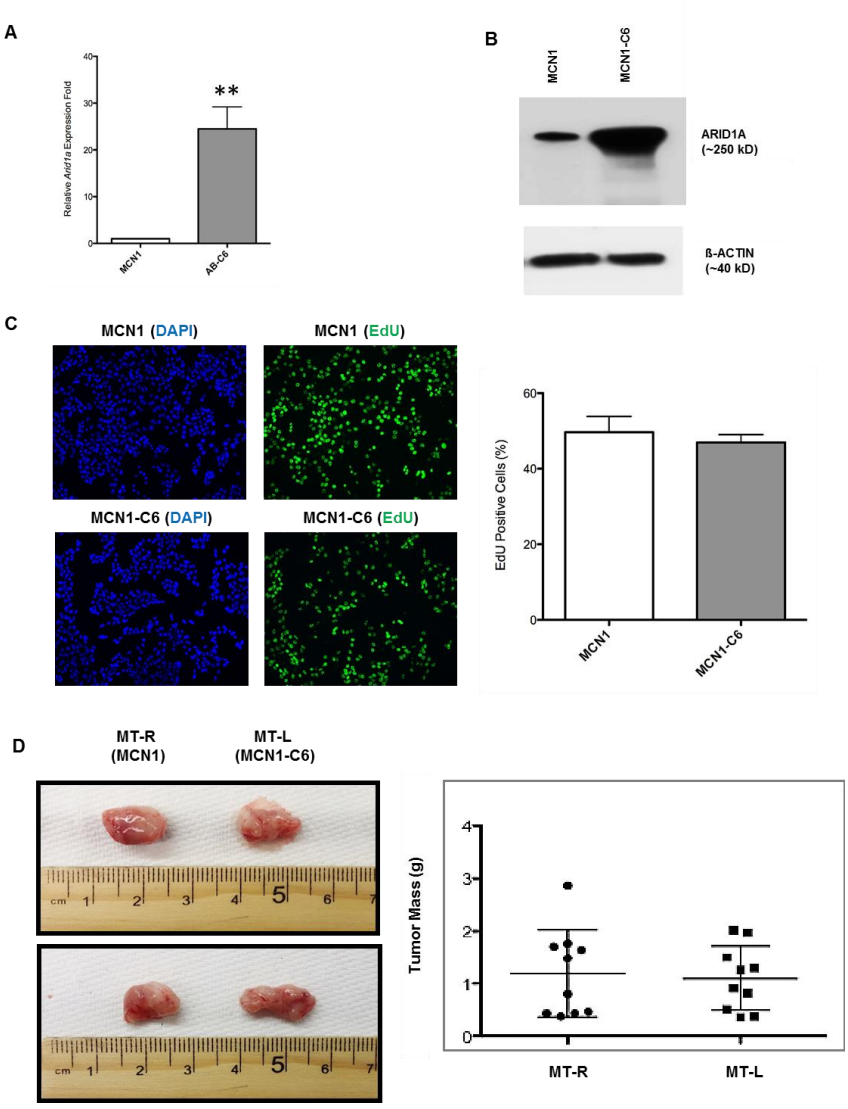


Figure 2-6. Overexpression of *Arid1a* in the TRP53-deficient MCN1 has no effect on proliferation rate or tumor growth. *Arid1a* expression levels were quantified by (A) qRT-PCR and (B) immunoblotting in MCN1 MT cell line and its transduced counterpart cell line (MCN1-C6). Significant differences were calculated using a two-tailed Student's t-test (** $P < 0.001$). Values are relative to untransduced parental cell line MCN1. (C) EdU incorporation assays of MCN1 vs. MCN1-C6. Error bars signify SEM of three experimental replicates, with >1000 cells counted per sample for each replicate. (D) Representative images of tumors arising from transplantation of MCN1 (MT-R) and MCN1-C6 (MT-L) cells into cleared fad pads of recipient syngeneic FVB/N recipient females. MT-R and MT-L refer to right and left sides of mouse, respectively (see Materials and Methods). The right panel plots individual weights of tumors (n = 10 mice, 20 tumors).

2.4 Materials and Methods

Cancer cell lines

The 23116 MT cell line was generated from a primary mammary tumor that arose in a C3H-*Chaos3* mouse. The tumor was dissected and mechanical pulverized, then seeded on gelatin-coated culture dishes in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% Fetal Bovine Serum (FBS). “Add-back” (AB) cell lines were generated by lentiviral transduction of *Arid1a* expression vector pLenti-puro-ARID1A (Addgene plasmid # 39478), followed by puromycin selection (2µg/ml) and growth of clonal lines (AB-C1/C2/C3) expressing ectopic *Arid1a*. All qRT-PCR primers are shown in Table S4.

Array comparative genomic hybridization (aCGH)

Genomic DNA was isolated from primary tumors by solubilizing in lysis buffer (50 mM Tris-HCl pH 8.0; 100 mM EDTA pH 8.0; 100 mM NaCl; 1% SDS; 0.5 mg/ml Proteinase K) for 3 hours at 55°, phenol/chloroform extraction, precipitation of the DNA in 0.8 volumes isopropanol, followed by spooling and washes in 70% then 100% ethanol. One µg of genomic DNA was used for labeling and hybridization to the SurePrint G3 Mouse Genome CGH Microarray, 1 x 1M (Agilent Product # G4838A). Two independent reference WT DNAs (from strain C3HeB/FeJ mammary tissue) were used as hybridization controls. This array consists of 60-mer probes with an overall median spacing of 1.8kb (1.5kb in Refseq genes). Content for probe design was sourced from UCSC mm9 (NCBI Build 37). DNA labeling, hybridization, and post-hybridization processing were performed according to the manufacturer’s

protocol. Images were scanned using Agilent's SureScan Microarray Scanner. Agilent's Cytogenomics software was used for spot identification and signal quantification, following normalization of test/reference ratios and background correction. Criteria for calling amplifications/deletions were as follows: minimum number of contiguous probes >= 3, minimum avg. absolute $\log_2$ ratio >= 0.25. CNAs were visualized using the Integrative Genomics Viewer software package[51].

Droplet Digital PCR

Droplet digital PCR (ddPCR) was carried out using the QX200 droplet digital PCR system (Bio-Rad Laboratories, Inc., Hercules, CA, USA). ~60ng genomic DNA extracted from 51 different tumor samples (*Chaos3* MTs and *Chaos3* non-MT controls) was used per reaction. Individual tumor samples were analyzed for CNVs occurring in the target gene *Arid1a*, by probing two different genomic locations spanning the length of the entire gene. Primer and probe combinations used for the assay are shown in Table S4. *Gapdh* was used as a reference gene in CNV analyses. Droplet generation and droplet reading for ddPCR were carried out using Bio-Rad equipment and reagents, according to the manufacturer's instructions. Results were analyzed using QuantaSoft Software (Bio-Rad) and represented as concentration of DNA (copies/µl). Each DNA sample was run in duplicate. Results for all samples analyzed are shown Table S2-1.

qRT-PCR

Total RNA was isolated from cells and tissues using the E.Z.N.A Total RNA Kit I (Omega Biotek). 500ng of RNA was used for cDNA synthesis using the qScript

cDNA Supermix kit (Quantabio). qRT-PCR analyses was done using Fast SYBR Green Master Mix (Life Technologies) and custom designed primers (Table S4), using GAPDH as an endogenous reference. Assays were run on the CFX96 Touch™ Real-Time PCR Detection System (BIO-RAD). Each sample was run in triplicate wells, from which mean Ct values were obtained. Relative quantification of gene expression was calculated using the ΔΔCt method. At least two technical replicates were run for each experiment to obtain standard error values.

EdU proliferation assay

Cells were grown O/N on coverslips and pulse labeled with 10μM EdU for 30 minutes. Cells were then fixed with formaldehyde (final concentration of 1%) for 10 minutes, followed by permeabilization (0.3% Triton X-100 in PBS) for 15 minutes. The 'Click' reaction cocktail [10mM (+)-sodium-L-ascorbate; 0.1mM 6-Caboxyfluorescein-TEG azide; 2mM CuSO4] was added to cells and incubated for 30 minutes at room temperature. Nuclei were counterstained with DAPI and coverslips were mounted on slides for EdU-positive cell counting by fluorescence microscopy. Experiments were conducted in triplicate, with >1000 cells counted / replicate.

Mammary fat pad injection surgeries

MT cell lines were injected into the #4 and #9 inguinal fat pads of 3-week old nulliparous WT C3HeB/FeJ female mice, following clearance of the endogenous epithelium. Volume of cells injected / fat pad was 10μl, at a concentration of $1x10^6$ cells/ml. Tumors were allowed to develop until one grew to ~2cm in diameter, at

which point the animals were sacrificed and tissues were collected for analyses. The time to tumor formation following surgery varied from 3-12 weeks.

Cell cycle analyses

One million cells were centrifuged and resuspended in 200µl of a cold hypotonic solution of propidium iodide (50µg/ml PI, 1mg/ml sodium citrate, and 1µl/ml Triton X-100). Cells were incubated at 4°C overnight for complete lysis and staining of nuclei. Cell cycle profiles were analyzed using a flow cytometer (BD LSR II), with 488-nm excitation and emission collected with a 576/26 band-pass filter. Using a PI signal-specific width vs. area plot, only single nuclei were included in the analyses of all profiles.

Senescence-associated beta-galactosidase assay

AB-C1 and 23116 MT cells were seeded onto a 6-well dish at a concentration of 0.2×10^6 cells/ml. The next day, cells were stained using a senescence detection kit (Abcam, ab65351) as per the manufacturer's instructions. In brief, cells were fixed for 10 min at RT with the provided Fixation Solution, and then stained for 16 hrs at 37C. The next day, they were visualized using light microscopy for development of blue color. Images were taken at 10X magnification and the number of positive cells were counted using ImageJ software. Experiments were carried out in triplicate to calculate the average percentage values.

RNA sequencing

Total RNA was isolated from replicate samples of 23116 MT and AB clones (C1 and C2) cells using the E.Z.N.A. Total RNA Kit I (Omega Biotek). 500 ng/sample was used to prepare cDNA libraries, with the NEBNext Poly(A) mRNA Magnetic Isolation Module and the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (both from NEB). The libraries were then sequenced on Illumina's Hi-Seq platform, generating single-end 100 bp reads. Reads were aligned to the mouse genome (UCSC mm10) using Tophat v2.0.13[52]. Significant differentially-expressed (DE) genes between 23116 MT and AB clones were determined with the help of the TopHat tool CuffDiff (v2.2.1), which uses a q-value cutoff of 0.05 (q-value = p-value corrected for multiple hypothesis testing)[52]. DE genes were additionally sorted based on more stringent criteria where at least one condition (avg. of replicates) must have FPKM>=5 and the minimum log2 (fold-change) between conditions is 2-fold (up or down).

Statement on data and reagent availability: Cell lines and constructs are available upon request. RNA-seq data and aCGH data are available at the Gene Expression Omnibus (accession numbers GSE81575 and GSE81967, respectively).

Acknowledgements: The authors are grateful to Alex Nikitin for providing MCN1 cells, Jennifer Grenier for the RNA-seq analysis, Christa Heyward and Rodica Bunaciu for assistance with flow cytometry, and Robert Weiss and Adrian McNairn for feedback on the manuscript. This work was supported by grant NIH R21CA175961 to JCS.

2.5 References

1. Cajuso, T. *et al.* Exome sequencing reveals frequent inactivating mutations in ARID1A, ARID1B, ARID2 and ARID4A in microsatellite unstable colorectal cancer. *Int J Cancer* **135**, 611-23 (2014).
2. Wang, K. *et al.* Exome sequencing identifies frequent mutation of ARID1A in molecular subtypes of gastric cancer. *Nat Genet* **43**, 1219-23 (2011).
3. Cancer Genome Atlas Research, N. *et al.* Integrated genomic characterization of endometrial carcinoma. *Nature* **497**, 67-73 (2013).
4. Liang, H. *et al.* Whole-exome sequencing combined with functional genomics reveals novel candidate driver cancer genes in endometrial cancer. *Genome Res* **22**, 2120-9 (2012).
5. Wiegand, K.C. *et al.* ARID1A mutations in endometriosis-associated ovarian carcinomas. *N Engl J Med* **363**, 1532-43 (2010).
6. Jones, S. *et al.* Frequent mutations of chromatin remodeling gene ARID1A in ovarian clear cell carcinoma. *Science* **330**, 228-31 (2010).
7. Waddell, N. *et al.* Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature* **518**, 495-501 (2015).
8. Imielinski, M. *et al.* Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* **150**, 1107-20 (2012).
9. Mamo, A. *et al.* An integrated genomic approach identifies ARID1A as a candidate tumor-suppressor gene in breast cancer. *Oncogene* **31**, 2090-100 (2012).
10. Cornen, S. *et al.* Mutations and deletions of ARID1A in breast tumors. *Oncogene* **31**, 4255-6 (2012).
11. Inoue, H. *et al.* Target genes of the largest human SWI/SNF complex subunit control cell growth. *Biochem J* **434**, 83-92 (2011).
12. Chandler, R.L. *et al.* ARID1a-DNA interactions are required for promoter occupancy by SWI/SNF. *Mol Cell Biol* **33**, 265-80 (2013).
13. Romero, O.A. & Sanchez-Cespedes, M. The SWI/SNF genetic blockade: effects in cell differentiation, cancer and developmental diseases. *Oncogene* **33**, 2681-9 (2014).
14. Gao, X. *et al.* ES cell pluripotency and germ-layer formation require the SWI/SNF chromatin remodeling component BAF250a. *Proc Natl Acad Sci U S A* **105**, 6656-61 (2008).
15. Chandler, R.L. *et al.* Coexistent ARID1A-PIK3CA mutations promote ovarian clear-cell tumorigenesis through pro-tumorigenic inflammatory cytokine signalling. *Nat Commun* **6**, 6118 (2015).
16. Guan, B. *et al.* Roles of deletion of Arid1a, a tumor suppressor, in mouse ovarian tumorigenesis. *J Natl Cancer Inst* **106**(2014).
17. Maeda, D. & Shih Ie, M. Pathogenesis and the role of ARID1A mutation in endometriosis-related ovarian neoplasms. *Adv Anat Pathol* **20**, 45-52 (2013).
18. Kandoth, C. *et al.* Mutational landscape and significance across 12 major cancer types. *Nature* **502**, 333-9 (2013).

19. ACS. Breast cancer facts and figures 2013-2014 (American Cancer Society, cancer.org, 2014).
20. Cho, H.D. *et al.* Loss of Tumor Suppressor ARID1A Protein Expression Correlates with Poor Prognosis in Patients with Primary Breast Cancer. *J Breast Cancer* **18**, 339-46 (2015).
21. Shima, N. *et al.* A viable allele of Mcm4 causes chromosome instability and mammary adenocarcinomas in mice. *Nat Genet* **39**, 93-8 (2007).
22. Chuang, C.H. *et al.* Post-transcriptional homeostasis and regulation of MCM2-7 in mammalian cells. *Nucleic Acids Res* **40**, 4914-24 (2012).
23. Kawabata, T. *et al.* Stalled fork rescue via dormant replication origins in unchallenged S phase promotes proper chromosome segregation and tumor suppression. *Molecular Cell* **41**, 543-53 (2011).
24. Wallace, M.D. *et al.* Comparative oncogenomics implicates the neurofibromin 1 gene (NF1) as a breast cancer driver. *Genetics* **192**, 385-96 (2012).
25. Wallace, M.D., Southard, T.L., Schimenti, K.J. & Schimenti, J.C. Role of DNA damage response pathways in preventing carcinogenesis caused by intrinsic replication stress. *Oncogene* **33**, 3688-95 (2014).
26. Ciriello, G. *et al.* Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. *Cell* **163**, 506-19 (2015).
27. Eirew, P. *et al.* Dynamics of genomic clones in breast cancer patient xenografts at single-cell resolution. *Nature* **518**, 422-6 (2015).
28. TCGA. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61-70 (2012).
29. Shen, J. *et al.* ARID1A Deficiency Impairs the DNA Damage Checkpoint and Sensitizes Cells to PARP Inhibitors. *Cancer Discov* **5**, 752-67 (2015).
30. Lobrich, M. & Jeggo, P.A. The impact of a negligent G2/M checkpoint on genomic instability and cancer induction. *Nat Rev Cancer* **7**, 861-9 (2007).
31. Bai, G., Smolka, M.B. & Schimenti, J.C. Chronic DNA Replication Stress Reduces Replicative Lifespan of Cells by TRP53-Dependent, microRNA-Assisted MCM2-7 Downregulation. *PLoS Genet* **12**, e1005787 (2016).
32. Kuilman, T., Michaloglou, C., Mooi, W.J. & Peeper, D.S. The essence of senescence. *Genes Dev* **24**, 2463-79 (2010).
33. Krenning, L., Feringa, F.M., Shaltiel, I.A., van den Berg, J. & Medema, R.H. Transient activation of p53 in G2 phase is sufficient to induce senescence. *Mol Cell* **55**, 59-72 (2014).
34. Bunz, F. *et al.* Requirement for p53 and p21 to sustain G2 arrest after DNA damage. *Science* **282**, 1497-501 (1998).
35. Mao, Z., Ke, Z., Gorbunova, V. & Seluanov, A. Replicatively senescent cells are arrested in G1 and G2 phases. *Aging (Albany NY)* **4**, 431-5 (2012).
36. Wu, R.C., Wang, T.L. & Shih Ie, M. The emerging roles of ARID1A in tumor suppression. *Cancer Biol Ther* **15**, 655-64 (2014).
37. Pereira, J.J. *et al.* Bimolecular interaction of insulin-like growth factor (IGF) binding protein-2 with alphavbeta3 negatively modulates IGF-I-mediated migration and tumor growth. *Cancer Res* **64**, 977-84 (2004).

38. Butt, A.J., Dickson, K.A., McDougall, F. & Baxter, R.C. Insulin-like growth factor-binding protein-5 inhibits the growth of human breast cancer cells in vitro and in vivo. *J Biol Chem* **278**, 29676-85 (2003).
39. Andreasen, P.A., Kjoller, L., Christensen, L. & Duffy, M.J. The urokinase-type plasminogen activator system in cancer metastasis: a review. *Int J Cancer* **72**, 1-22 (1997).
40. Kim, K.S. *et al.* Induction of cellular senescence by insulin-like growth factor binding protein-5 through a p53-dependent mechanism. *Mol Biol Cell* **18**, 4543-52 (2007).
41. Guan, B., Wang, T.L. & Shih Ie, M. ARID1A, a factor that promotes formation of SWI/SNF-mediated chromatin remodeling, is a tumor suppressor in gynecologic cancers. *Cancer Res* **71**, 6718-27 (2011).
42. Bitler, B.G., Fatkhutdinov, N. & Zhang, R. Potential therapeutic targets in ARID1A-mutated cancers. *Expert Opin Ther Targets* **19**, 1419-22 (2015).
43. Zang, Z.J. *et al.* Exome sequencing of gastric adenocarcinoma identifies recurrent somatic mutations in cell adhesion and chromatin remodeling genes. *Nat Genet* **44**, 570-4 (2012).
44. Bosse, T. *et al.* Loss of ARID1A expression and its relationship with PI3K-Akt pathway alterations, TP53 and microsatellite instability in endometrial cancer. *Mod Pathol* **26**, 1525-35 (2013).
45. Cheng, L. *et al.* Rb inactivation accelerates neoplastic growth and substitutes for recurrent amplification of cIAP1, cIAP2 and Yap1 in sporadic mammary carcinoma associated with p53 deficiency. *Oncogene* **29**, 5700-11 (2010).
46. Gilbert, L.A. *et al.* Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell* **159**, 647-61 (2014).
47. Chavez, A. *et al.* Highly efficient Cas9-mediated transcriptional programming. *Nat Methods* **12**, 326-8 (2015).
48. Zhang, Y. *et al.* CRISPR/gRNA-directed synergistic activation mediator (SAM) induces specific, persistent and robust reactivation of the HIV-1 latent reservoirs. *Sci Rep* **5**, 16277 (2015).
49. Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583-8 (2015).
50. Maeder, M.L. *et al.* Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. *Nat Biotechnol* **31**, 1137-42 (2013).
51. Robinson, J.T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24-6 (2011).
52. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7**, 562-78 (2012).

2.6 Supporting Information

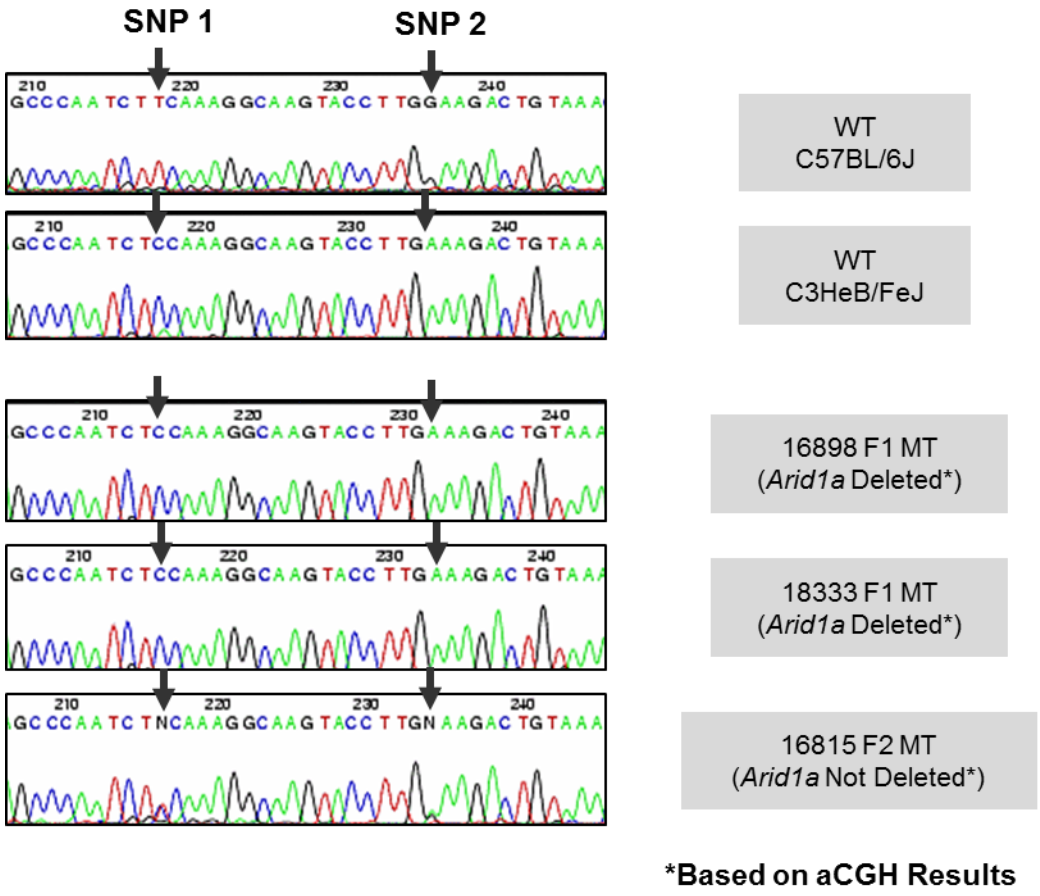


Figure S2-1. **SNP genotyping of *Chaos3*-MTs.** Three individual *Arid1a* SNPs specific to C3HeB/FeJ and C57BL/6J strains were genotyped in F1 and F2 MTs derived from mice generated by crossing these two strains, for which aCGH analyses was also conducted.

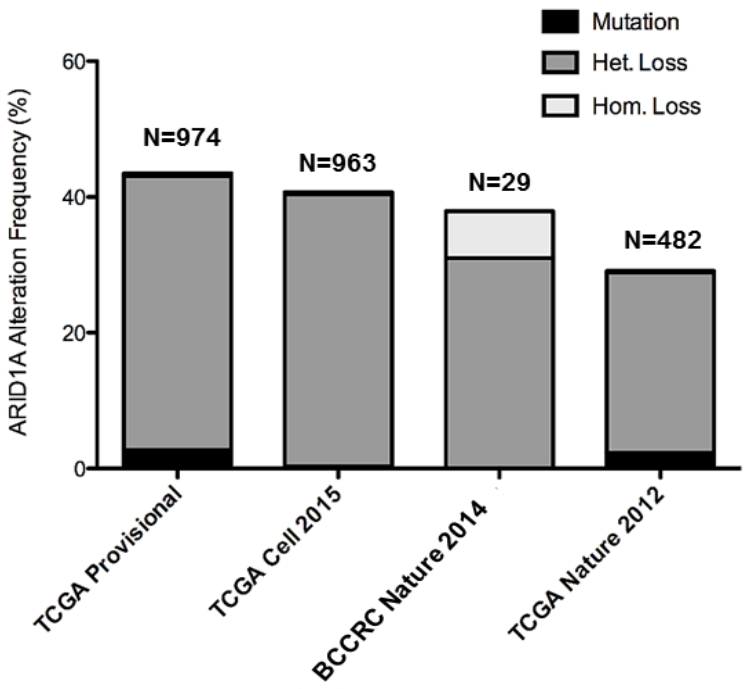


Figure S2-2. *ARID1A* copy number in human breast cancers. Data and analytical tools from the cBio portal (http://www.cbioportal.org) were used for generating the graph. "Mutation" refers to intragenic mutations such as point changes; "Het. Loss" = 1 *ARID1A* copy scored as missing; "Hom. Loss" = no *ARID1A* copies remaining.

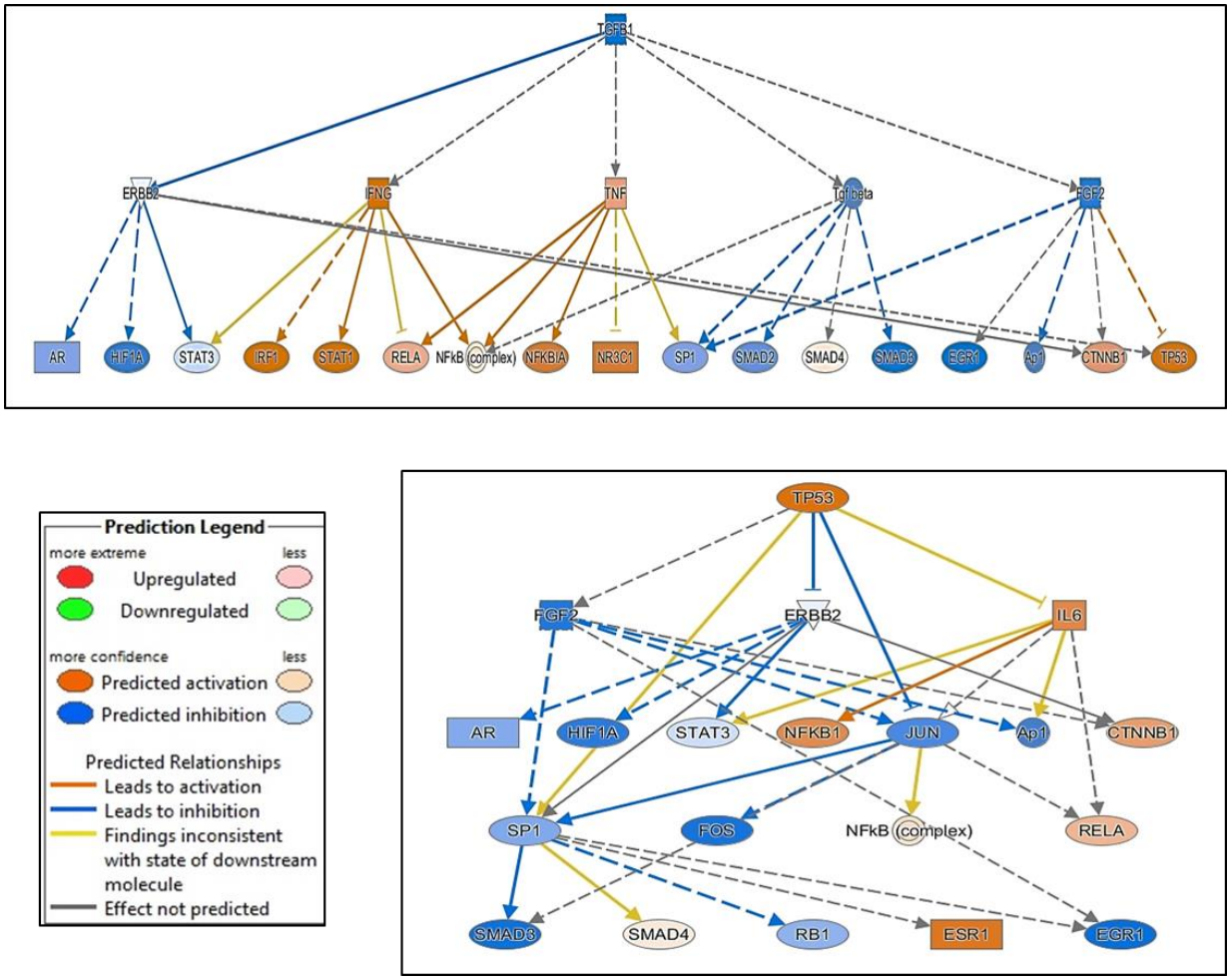


Figure S2-3: IPA analyses of RNA-Seq DE genes. Ingenuity Pathway Analyses of gene networks altered within the TGFB1 and TP53 pathways in AB-C1and AB-C2 cells.

Table S2-1. *Arid1a* CNVs across *Chaos3* MTs and controls determined by ddPCR.

	Arid1a Primer + Probe Set #1							*Arid1a* Primer + Probe Set #2						
	R1 (copies/ul)		R2 (copies/ul)		RATIO		Avg. (R1 & R2)	R1 (copies/ul)		R2 (copies/ul)		RATIO		Avg. (R1 & R2)
Sample	FAM (*Arid1a*)	HEX (*Gapdh*)	FAM (*Arid1a*)	HEX (*Gapdh*)	R1 (FAM/HEX)	R2 (FAM/HEX)		FAM (*Arid1a*)	HEX (*Gapdh*)	FAM (*Arid1a*)	HEX (*Gapdh*)	R1 (FAM/HEX)	R2 (FAM/HEX)	
18333 F1 MT	266	260	264	263	1.0	1.0	1.0	370	698	283	545	0.5	0.5	0.5
16898 F1 MT	345	624	455	835	0.6	0.5	0.5	381	719	398	743	0.5	0.5	0.5
18656 F2 MT	230	448	697	1257	0.5	0.6	0.5	275	501	388	728	0.5	0.5	0.5
16694 F2 MT	219	403	271	526	0.5	0.5	0.5	319	664	1396	2990	0.5	0.5	0.5
16815 F2 MT	150	141	147	154	1.1	1.0	1.0	154	154	134	138	1.0	1.0	1.0
20626 MT	46.3	55.4	29.9	61.5	0.8	0.5	0.7	26.2	48.9	28.2	53.4	0.5	0.5	0.5
20894 MT	500	580	664	802	0.9	0.8	0.8	618	672	1119	1204	0.9	0.9	0.9
20164 MT	96	148	88	152	0.6	0.6	0.6	83	146	114	212	0.6	0.5	0.6
20889 MT	526	541	593	624	1.0	1.0	1.0	684	725	800	824	0.9	1.0	1.0
20892 MT	567	974	371	606	0.6	0.6	0.6	1158	1962	772	1327	0.6	0.6	0.6
20138 MT	660	1190	724	1297	0.6	0.6	0.6	691	1264	1900	3570	0.5	0.5	0.5
19959 MT	586	1193	379	736	0.5	0.5	0.5	1787	3510	1369	2680	0.5	0.5	0.5
18224 MT	189	197	157	166	1.0	0.9	1.0	191	219	678	818	0.9	0.8	0.9
20337 MT	143	199	150	209	0.7	0.7	0.7	2189	3440	2160	3420	0.6	0.6	0.6
21809 MT	54.6	99	199	343	0.6	0.6	0.6	670	1240	1267	2640	0.5	0.5	0.5
21038 MT	920	1625	1056	1912	0.6	0.6	0.6	2960	2920	2740	4580	1.0	0.6	0.8
21333 MT	84	101	112	128	0.8	0.9	0.9	1007	1222	1046	1276	0.8	0.8	0.8
22420 MT	548	668	621	787	0.8	0.8	0.8	523	811	573	887	0.6	0.6	0.6
21041 MT	41.6	40	28.2	32.7	1.0	0.9	1.0	363	518	1072	1522	0.7	0.7	0.7
21419 MT	120	145	75.4	99	0.8	0.8	0.8	1160	1770	986	1536	0.7	0.6	0.6
20459 MT	208	157	217	151	1.3	1.4	1.4	2700	2600	3070	3010	1.0	1.0	1.0
22476 MT	1026	1368	1144	1497	0.8	0.8	0.8	1110	1840	1596	2650	0.6	0.6	0.6
22180 MT	256	222	574	474	1.2	1.2	1.2	2630	2670	1989	2234	1.0	0.9	0.9
19660 MT	26.3	29.3	6330	11300	0.9	0.6	0.7	3370	3410	4100	4060	1.0	1.0	1.0
22238 MT	75.8	114	860	1337	0.7	0.6	0.7	1160	2090	1056	2020	0.6	0.5	0.5
22417 MT	132	109	155	124	1.2	1.3	1.2	204	238	137	165	0.9	0.8	0.8
21416 MT	756	919	917	1097	0.8	0.8	0.8	2043	2700	-	-	0.8	-	0.8
20890 MT	13.4	18.5	629	872	0.7	0.7	0.7	577	984	-	-	0.6	-	0.6
22235 MT	41.4	189	44.9	204	0.2	0.2	0.2	113	725	-	-	0.2	-	0.2
20163 MT	544	421	471	373	1.3	1.3	1.3	1330	1650	-	-	0.8	-	0.8
22414 MT	609	456	647	490	1.3	1.3	1.3	307	577	659	1333	0.5	0.5	0.5
22182 MT	88	88	280	268	1.0	1.0	1.0	690	810	1419	1685	0.9	0.8	0.8
21039 MT	1545	1912	839	1189	0.8	0.7	0.8	818	1514	-	-	0.5	-	0.5
21124 MT	157	199	260	347	0.8	0.7	0.8	702	1251	1354	2403	0.6	0.6	0.6
21597 MT	162	206	305	388	0.8	0.8	0.8	1764	2880	1772	2960	0.6	0.6	0.6
20318 MT	186	243	289	365	0.8	0.8	0.8	231	375	216	399	0.6	0.5	0.6
22416 MT	220	185	132	124	1.2	1.1	1.1	80	82	119	133	1.0	0.9	0.9
21810 MT	291	424	340	505	0.7	0.7	0.7	368	574	2016	3640	0.6	0.6	0.6
21255 MT	516	653	921	1126	0.8	0.8	0.8	1220	1900	534	871	0.6	0.6	0.6
21254 MT	110	125	159	194	0.9	0.8	0.8	729	1154	151	241	0.6	0.6	0.6
22418 MT	214	303	486	678	0.7	0.7	0.7	1193	2162	1251	2376	0.6	0.5	0.5
21417 MT	51.7	49.4	170	167	1.0	1.0	1.0	713	812	1111	1281	0.9	0.9	0.9
21809 MT	10.6	16.6	319	394	0.6	0.8	0.7	1620	1570	468	773	1.0	0.6	0.8
21123 MT	101	127	106	138	0.8	0.8	0.8	2600	4550	1851	3030	0.6	0.6	0.6
21811 MT	240	279	299	349	0.9	0.9	0.9	1718	2510	267	388	0.7	0.7	0.7
22166 MT	182	216	107	130	0.8	0.8	0.8	1087	1655	1517	2329	0.7	0.7	0.7
C3H-*Chaos3* Non-MT Controls														
20724 UT	446	408	437	418	1.1	1.0	1.1	404	396	507	513	1.0	1.0	1.0
17651 UT	119	99	105	91	1.2	1.2	1.2	104	108	100	99	1.0	1.0	1.0
17320 BT	90	89	89	91	1.0	1.0	1.0	91	86	105	107	1.1	1.0	1.0
WT Reference Controls														
C3H Ref DNA	76	68.1	98	99	1.1	1.0	1.1	47	48	38	40	1.0	1.0	1.0
B6 Ref DNA	1192	1199	835	870	1.0	1.0	1.0	456	506	515	541	0.9	1.0	0.9
C3H + B6 Ref DNA	329	311	177	175	1.1	1.0	1.0	109	108	93	101	1.0	0.9	1.0
***Chaos3* and non-*Chaos3* MT Cell Lines**														
20212 MT CL	187	225	213	244	0.8	0.9	0.9	491	566	345	399	0.9	0.9	0.9
21040 MT CL	889	1024	1165	1350	0.9	0.9	0.9	5340	6140	-	-	0.9	-	0.9
21253 MT CL	4060	4040	755	776	1.0	1.0	1.0	2770	2780	2810	3430	1.0	0.8	0.9
22168 MT CL	1491	2066	2460	3590	0.7	0.7	0.7	2640	3910	2870	4860	0.7	0.6	0.6
23116 MT CL	1511	2610	1407	2410	0.6	0.6	0.6	No Call	No Call	2760	5290	-	0.5	0.5
MCN1	1597	2530	-	-	0.6	-	0.6	1251	2340	-	-	0.5	-	0.5

Table S2-2. Comparison of *Arid1a* CNVs and expression across individual *Chaos3* MTs.

Sample	ddPCR Probe 1	ddPCR Probe 2	Deletion Status of *Arid1a*	Relative *Arid1a* Expression
20724 UT	1.1	1.0	ND	1.0
17651 UT	1.0	1.0	ND	1.0
20889 MT	1.0	1.0	ND	0.8
20459 MT	1.4	1.0	ND	0.4
20212 MT CL	0.9	0.9	ND	0.7
21040 MT CL	0.9	0.9	ND	0.5
21253 MT CL	1.0	0.9	ND	0.4
18333 F1 MT	1.0	0.5	het deleted	0.2
16898 F1 MT	0.5	0.5	het deleted	0.04
16694 F2 MT	0.5	0.5	het deleted	0.1
20164 MT	0.6	0.6	het deleted	0.2
19959 MT	0.5	0.5	het deleted	0.2
21890 MT	0.6	0.5	het deleted	0.4
21038 MT	0.6	0.8	het deleted	0.03
22420 MT	0.8	0.6	het deleted	0.003
21041 MT	1.0	0.7	het deleted	0.5
19660 MT	0.7	1.0	het deleted	0.5
22238 MT	0.7	0.5	het deleted	0.1
20890 MT	0.7	0.6	het deleted	0.2
20163 MT	1.3	0.8	het deleted	0.3
21039 MT	0.8	0.5	het deleted	0.2
20318 MT	0.8	0.6	het deleted	0.2
21810 MT	0.7	0.6	het deleted	0.3
21809 MT	0.7	0.8	het deleted	0.3
21123 MT	0.8	0.6	het deleted	0.1
21811 MT	0.9	0.7	het deleted	0.4
22166 MT	0.8	0.7	het deleted	0.1
22168 MT CL	0.7	0.6	het deleted	0.2
23116 MT CL	0.6	0.5	het deleted	0.2

Table S2-3. Primer Sequences.

Gene	Purpose	Forward Primer	Probe	Reverse Primer
*Arid1a*_DNA	ddPCR	GTTCAAGGAAGGCCTGAACT	TGGCAGAGCCTATAAGCCTCCAGTAGTCGC	AGTCTGGGCAGGAAAGAGTA
*Arid1a*_DNA	ddPCR	TGGATGTCCTGGAAGTCTGA	ACCTGAAATTCACCTCCCCCTGCCTCCC	ACCCACTTCTTTGCACCTAC
*Gapdh*_DNA	ddPCR	TCCCTCGAACTAAGGGGAAA	GGGGAGCAGGGTGGAGAGCCCG	TCCATCCTCCAGAAACCAGA
*Arid1a*_DNA	SNP Genotyping	GAAGCCGCATCTTGTCTACT	-	GCAGGTTGGTTTGGTTCTTG
Arid1a_m_RNA	qRT-PCR	TCCCAGCAAACTGCCTATTC	-	CATATCTTCTTGCCCTCCCTTAC
Arid1a_h_RNA	qRT-PCR	CAGTACCTGCCTCGCACATA	-	GCCAGGAGACCAGACTTGAG
Cdkn1a	qRT-PCR	AAGTGTGCCGTTGTCTCTTCG	-	AGTCAAAGTTCCACCGTTCTCG
Serpinb2	qRT-PCR	ACTTAATGGGCTTTATCCTTTCC	-	TGCGTCCTCAATCTCATCG
Igfbp2	qRT-PCR	GGCGCGGGTACCTGTGAAAA	-	TCTCCTGCTGCTCGTTGTAG
Igfbp5	qRT-PCR	CAAGCACACTCGCATT	-	CAGGTACACAGCACGG
Gapdh	qRT-PCR	CTTTGTCAAGCTCATTTCCTGG	-	TCTTGCTCAGTGTCCTTGC

CHAPTER 3

Transcriptional Targets of *Arid1a* Tumor Suppression

Nithya Kartha[1,2] and John C. Schimenti[1,2,3]

Affiliations:

[1] Department of Biomedical Sciences

[2] Department of Molecular Biology and Genetics

[3] Center for Vertebrate Genomics

Cornell University, Ithaca, NY 14853, USA

One Sentence Summary: Potential transcriptional targets of *Arid1a* are identified in *Chaos3* MT cells, through overlapping RNA-sequencing and ATAC-sequencing datasets.

3.1 Abstract

ARID1A encodes a mutually exclusive (with *ARID1B*) component of the mammalian SWI/SNF (BAF) chromatin remodeling complex. Recent cancer genome screens have identified mutations/alterations in multiple mSWI/SNF components across a spectrum of cancer types, with the highest frequency of mutations/alterations occurring in *ARID1A*. Multiple studies since have shown that ARID1A functions as a *bona fide* tumor suppressor, using different models for carcinogenesis. These studies have revealed the various pathways in which ARID1A plays a regulatory role, such as the PI3K/AKT, TP53 and TGFß signaling pathways, suggestively in a tissue-specific manner. Here, we attempted to identify genes that are direct transcriptional targets of *Arid1a* in the context of mammary tumorigenesis, using an *in vitro* system derived from a mouse model for sporadic breast cancer, in which *Arid1a* is recurrently deleted. By overlapping datasets generated from RNA-seq and ATAC-seq experiments conducted using *Chaos3* MT cells overexpressing *Arid1a*, we identified a significant subset of genes that are both differentially expressed and correspond to increased "chromatin accessible" peaks upon overexpressing *Arid1a*. Amongst these, the most differentially up or down regulated (log2fold $\geq$ or $\leq$ 5) included several known cancer-associate genes that have previously not been associated with ARID1A molecular function, and thus could be potential novel targets of ARID1A transcriptional regulation.

3.2 Introduction

Since the beginning of 2010, TCGA has revealed several novel candidate genes that may potentially play a causative role in carcinogenesis, based on their frequency of alteration in different types of human cancers. Amongst these, the genes encoding subunits of the mSWI/SNF complex were identified as the most highly mutated out of all the ATP-dependent chromatin remodelers[1]. The mSWI/SNF complexes are essential in regulating genetic programs and signaling pathways controlling replication, transcription, damage repair, differentiation, migration and development, amongst others[2,3]. The broad range of function and extensive cancer mutation spectrum of these complexes suggests that they play a widespread role in tumor suppression by protecting fundamental cellular activity and survival.

ARID1A (BAF250a, SMARCF1, p270) is a component of the BRG1-associated factor (BAF) complex, which is a subfamily of the SWI/SNF chromatin remodelers. Its paralog, ARID1B, appears to be differentially expressed in embryonic tissues, suggesting these mutual exclusive proteins may confer target and lineage specificity to the BAF complexes during development[4,5]. Additionally, ARID1A and ARID1B demonstrate nonsynchronous kinetics during cell cycle progression and appear to bind different transcriptional targets under normal circumstances[5-7]. A recent screen for targets that conferred synthetic lethality to ARID1A mutant cell lines identified ARID1B as its top hit, and showed that it is required for stable assembly of the BAF complexes in ARID1A-deficient cells[8].

ARID1A contains a highly conserved DNA-binding domain called ARID (AT-rich interactive domain) that spans approximately 100 amino acids of the protein.

Although ARID domains in general bind AT-rich sequences preferentially, it has been shown that mammalian ARID1A binds DNA without sequence specificity[9,10]. Disruption of ARID1A-DNA binding has been found to result in embryonic lethality in mice, due to impaired nucleosome substrate binding and promoter occupancy by mutant ARID1A-containing mSWI/SNF complexes[6].

ARID1A is the most frequently altered SWI/SNF subunit in cancer, with inactivating mutations/deletions occurring in a spectrum of tumor types. Despite all the evidence pointing towards an important tumor suppressive role in various cancers, the exact mechanism by which ARID1A prevents tumor growth in different tissue contexts is still unclear. Functional studies using different model systems have shown that ARID1A may behave as both a "gatekeeper" and a "caretaker" tumor suppressor, as it plays essential roles in cell cycle regulation and DNA damage repair, respectively[11]. Studies using GEMMs for cancer in which ARID1A has been conditionally deleted or mutated within specific tissues have revealed crosstalk/collaboration with different signaling pathways, including PI3K/AKT and TP53[12-14].

Mechanistically, ARID1A is thought to regulate gene function in two ways – (1) by facilitating the recruitment of BAF complexes to DNA regulatory elements through interaction with other transcription factors/cofactors (2) by direct transcriptional regulation of specific targets through its interaction with histone-modifying enzymes[15]. In my research work described in this chapter, I have attempted to differentiate between these direct and indirect targets of ARID1A transcriptional regulation by overlapping datasets derived from transcriptome sequencing and ATAC

(Assay for Transposase-Accessible Chromatin) sequencing, using a *Chaos3* MT cell line in which *Arid1a* has been ectopically overexpressed[16].

3.3 Results

Mapping of open chromatin following Arid1a overexpression

I followed the experimental scheme outlined in Figure 3-1A, with the goal of correlating mapped regions of open chromatin with differential gene expression. Mouse MT cells representing the untransduced (UT) and the add-back (AB) with low and high *Arid1a* levels, respectively, were dispersed into single cell suspensions and subjected to barcoding following integration of the Tn5 transposase bearing sequencing adapters. Based on the principle of transposase activity, only open/easily accessible chromatin regions will be targeted for sequencing, and not tightly packed/closed genomic regions[17]. The ATAC-seq libraries generated by each sample replicate produced the expected distribution of fragment lengths, with the majority representing smaller fragments of open inter-nucleosome regions, and very few representing larger fragments spanning intact/positioned nucleosomes, as has been previously described[18].

Approximately 70% of ATAC-seq peaks identified in the UT (control) replicates were also present in the AB (experimental) replicates, but expectedly there was a significant increase in the number of ATAC-seq peaks in the AB cells as a direct consequence of overexpressing *Arid1a*, with 13,536 AB-specific peaks detected compared to only 4,968 UT-specific peaks detected (Figure 3-1B).

Annotation of ATAC-seq peaks revealed that overexpressing *Arid1a* in the *Chaos3* MT cells resulted in a two-fold increase in the number of peaks associated with the transcription start sites (TSS) and 5' untranslated region (5' UTR) regions of genes (Figure 3-1C), suggestive of a dramatic increase in gene transcriptional activity attributed directly to *Arid1a* function.

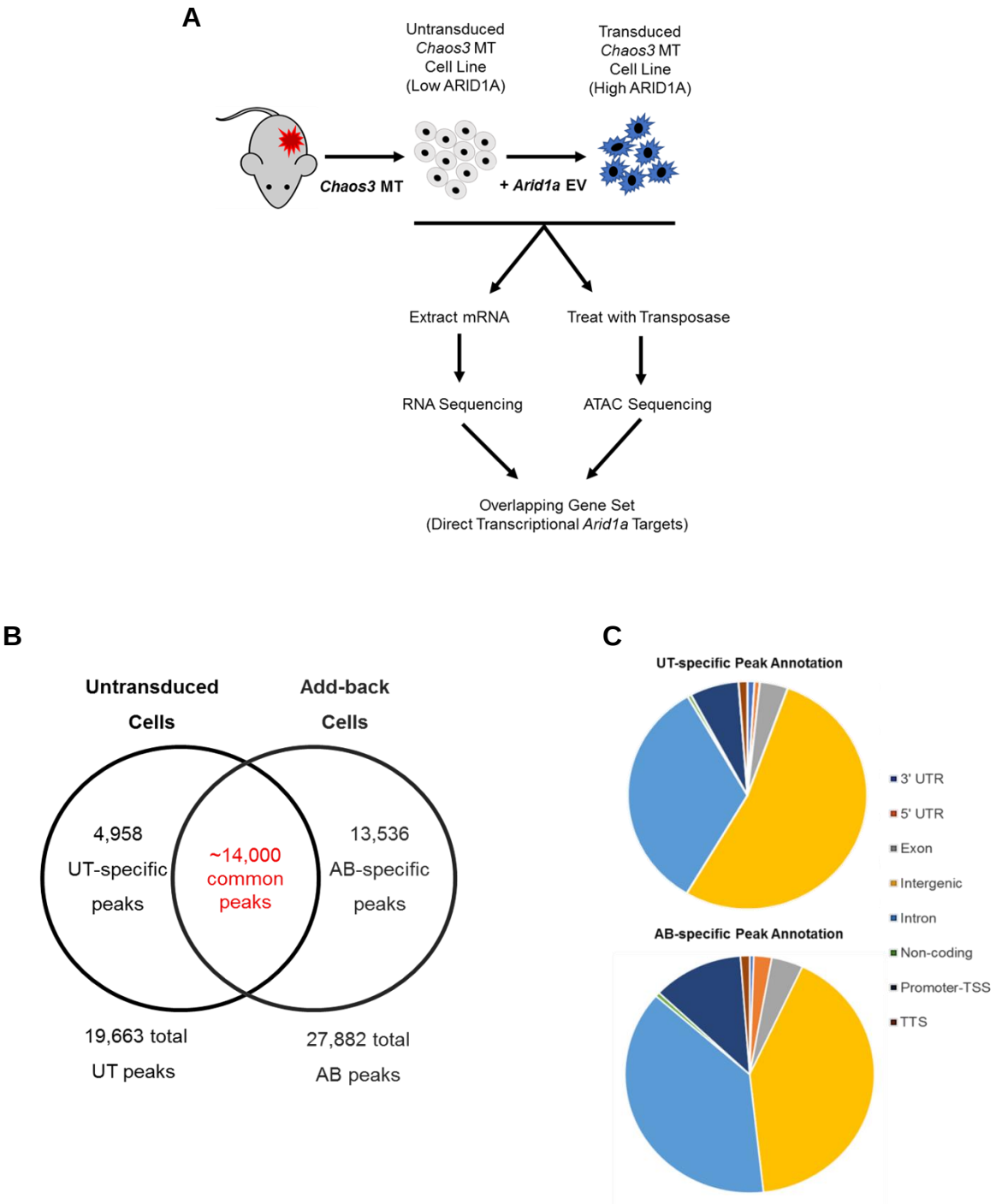


Figure 3-1. ATAC-seq peaks identified in *Chaos3* MT cells overexpressing *Arid1a*. (A) Schematic outline of experiment to determine direct transcriptional targets of ARID1A. (B) Venn diagram indicating the numbers of ATAC-seq peaks identified in untransduced (UT) and *Arid1a* add-back (AB) MT cells. (C) Pie charts representing annotation of UT-specific and AB-specific peaks. 3' UTR = 3' Untranslated region; 5' UTR = 5' Untranslated region; Promoter-TSS =Promoter transcriptional start site; TTS = Translational termination site

Integration of ATAC-seq and RNA-seq results in tumor cells overexpressing Arid1a

In previously published work, I have conducted mRNA-sequencing experiments to identified genes that are differentially expressed (DE) upon ectopic introduction of *Arid1a* in mouse mammary tumor cells that are hemizygously deleted for it[16]. Pathway analyses of the DE genes revealed alterations in critical growth and proliferation pathways, including TGF Beta and TP53. While this experiment was useful for identifying potential targets of *Arid1a* regulation, it is not sufficient on its own to differentiate between those that are direct or indirect.

To determine whether open regions of chromatin correlated with gene expression changes, we integrated the ATAC-seq and RNA-seq datasets to identify overlapping genes. Of the 1,064 DE genes identified by RNA-seq (UT cells vs. AB cells) previously, we found that 593 genes (~56 %) corresponded to annotated open chromatin regions identified by ATAC-seq (AB cell-specific peaks) (Figure 3-2A). The overall representation of annotated peaks did not vary significantly between the AB-specific ATAC-seq peak data and the overlapping DE gene set (Figure 3-2B).

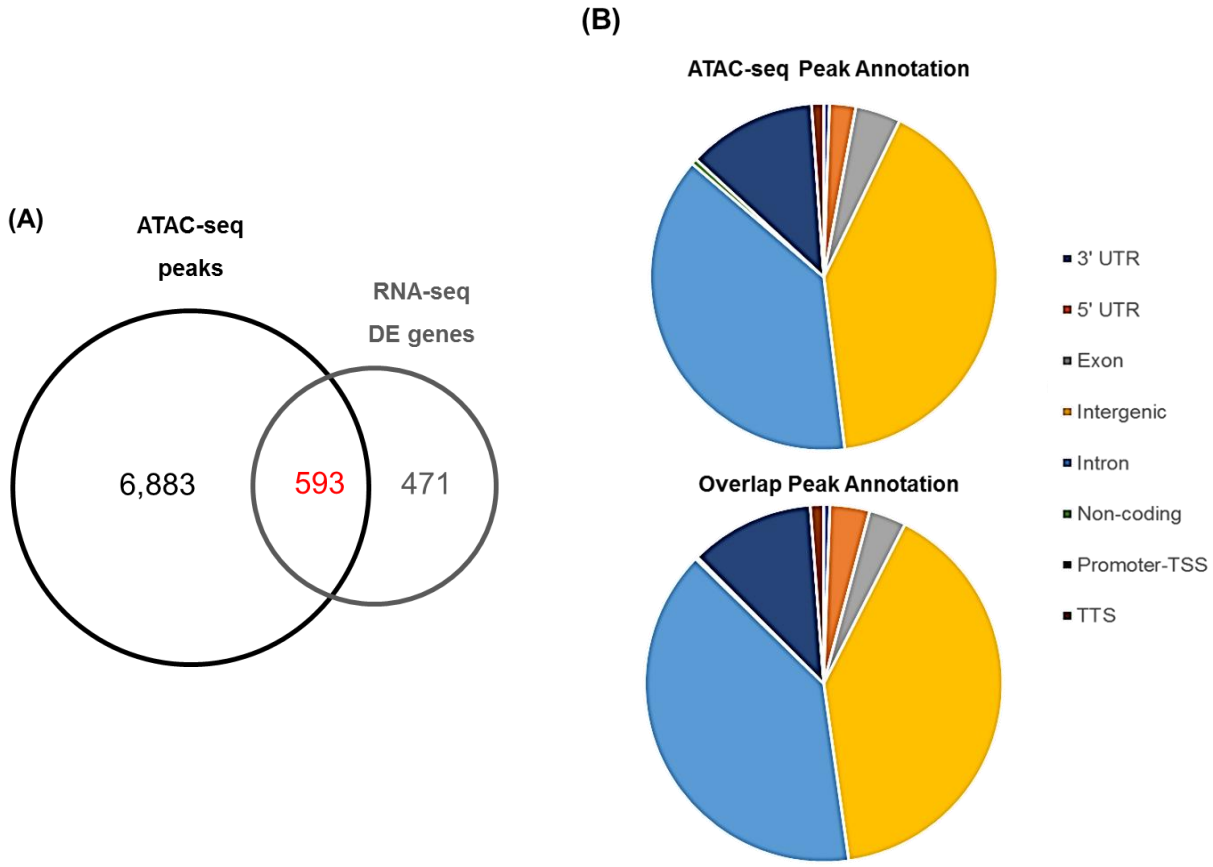


Figure 3-2. Overlapping Gene Set Representing Direct Transcriptional Targets of *Arid1a*. (A) Venn diagram showing the number of overlapping genes in the ATAC-seq and RNA-Seq datasets. (B) Pie charts showing annotation of ATAC-seq peaks and overlapping peaks with RNA-seq DE genes. 3' UTR = 3' Utranslated region; 5' UTR = 5' Untranslated region; Promoter-TSS =Promoter transcriptional start site; TTS = Translational termination site.

Genes Identified as Direct transcriptional targets of Arid1a

To further narrow down the integrated RNA-seq and ATAC-seq data to potential direct transcriptional targets of *Arid1a*, we identified the most highly differentially expressed genes within the overlapping dataset ($\geq$ or $\leq$ 5 Log2 fold expression; RPKM >5). This resulted in a list of 25 gene transcripts (Table 1). Several of these genes (9/25) had peaks located within 1Kb of the TSS, and an equal number (9/25) had more than one type of annotated peak, suggesting multiple access points for ARID1A binding at those genes.

Amongst this set of transcripts are three genes encoding type 1 collagen (*Col23a1*, *Col1a1* and *Col1a2*), which are a group of transmembrane proteins representing a major component of the extracellular matrix (ECM), whose expression is controlled in a tissue-specific manner during normal development and carcinogenesis[19]. COL23A1 was first identified as a biomarker in metastatic tumor cells[20], and is a biomarker for detecting recurrent non-small cell lung and prostrate cancer[21,22]. Abnormal COL1A1 and COL2A1 expression has been identified in breast cancer[23-28], renal cancer[29], hepatocellular cancer[30], melanoma[31] and gastric cancer[32].

Also included are genes encoding multiple growth factor and growth factor receptor proteins, namely fibroblast growth factor receptor 2 (*Fgfr2*), insulin-like growth factor receptor (*Igf1*), multiple epidermal growth factor-like domains protein 6 (*Megf6*) and nerve growth factor receptor (*Ngfr*). FGFR signaling is important for both embryonic development and adult tissue homeostasis, and is found to be dysregulated in a variety of human cancers[33], including breast[23,34]. IGF1 signaling also plays a critical role in development and carcinogenesis, particularly with respect to hormone-

sensitive cancers such a breast, ovarian, endometrial and pancreatic carcinomas[35-38]. Finally, NGFR is a prognostic marker for luminal subtypes of breast cancer[39], as well as many other cancer types[40-42], and has been identified as a transcriptional target of TP53[43].

Other interesting genes from the list included chordin-like 1 (*Chrdl1*), which encodes an antagonist of bone morphogenic protein 4 (*Bmp4*) and has been recently identified as a tumor suppressor in breast cancer[44,45], gastric cancer[46], lung cancer[47] and melanoma[48]. C-X-C chemokine receptor type 4 (*Cxcr4*) is a G-protein coupled receptor that has been well characterized as an oncogenic factor in more than 23 different human cancers, and is considered an important target for therapeutic intervention[49]. Plasminogen activator inhibitor-1 (*Pai1/Serpine1*) encodes a member of the plasminogen-plasmin system that plays a critical role in cell migration, cell proliferation and tissue remodeling, and has been implicated as a valuable prognostic indicator in a spectrum of cancer types[50], particularly in breast cancer[50,51]. Semaphorin 4D (*Sema4D*) encodes a member of the semaphorin family of secreted membrane proteins, that has been shown to function as a potent oncogene and metastatic factor in breast cancer[52,53], as well as playing a prognostic role in several other cancer types[54-57]. Cellular retinoic acid-binding protein 1 (*Crabp1*) plays a key role in tumor response to retinoic acid and is often hypermethylated in cancer, with its low expression serving as a poor prognosis indicator in multiple cancer types[58-62], particularly in breast cancer[63]. Pituitary Tumor Transforming Gene 1 (*Ptt1*) behaves as an oncogene in a variety of endocrine and non-endocrine cancers[64], and was found to promote tumor cell growth by inhibiting TGFß1/SMAD3 signaling in breast, prostate and lung cancers[65-67].

Alpha-smooth muscle actin (*Acta2*) is a stromal biomarker of cancer-associated fibroblasts[68,69], and its increased expression is associated with accelerated invasiveness and metastasis of breast cancer cells[70]. Ras homolog gene family, member C (*RhoC*) encodes a small signaling GTPase, and behaves as a potent oncogenic factor in several different cancer types, including breast cancer[71-78]. Dickkopf WNT Signaling Pathway Inhibitor 2 (*Dkk2*) encodes a protein that is involved in embryonic development through its interactions with the Wnt signaling pathway, and is frequently dysregulated in several different cancer types[79-84]. Finally, Small Nuclear Ribonucleoprotein Polypeptide N (*Snrpn*) encodes a protein that plays a role in pre-mRNA processing and tissue-specific alternative splicing events, and appears to behave as an oncogenic factor in medulloblastomas and pancreatic cancer[85,86]. The cancer-associated roles of these genes have been summarized in Table 3-2.

Gene ID	Log2 Fold Differential Expression	Annotated Peak within 1Kb of TSS	More than One Annotated Peak
Col23a1	6.2	X	
Serpine1	6.2		
Pknox2	5.8		
Fgfr2	5.7	X	
Chrdl1	5.5	X	
Igf1	5.3		X
Megf6	5.3		
Sema4d	5.3		
Asprv1	5.2		
Cxcr4	5.2	X	X
Ngfr	5.2	X	
Fam198b	-5.2		X
Col1a1	-5.3		X
Itgb3	-5.3	X	X
Crabp1	-5.6		
Pttg1	-5.6		
Acta2	-5.8	X	
Rhoc	-6.1		X
Col1a2	-6.2		X
Prl2c2	-6.4		
Stac2	-7.2	X	X
Slc44a1	-7.8		X
Dkk2	-10		
Gng11	-10	X	
Snrpn	-10		

Table 3-1. Potential direct transcriptional targets of *Arid1a* gene regulation. Highest DE genes (≥ or ≤ 5 Log2 fold expression; RPKM >5) present within the overlap dataset. X= presence of annotated peaks from ATAC-seq dataset.

GENE ID	GENE DESCRIPTION	ASSOCIATED CANCER TYPES	REFERENCES
Col23a1	collagen, type XXIII, alpha 1	Gastric cancer, prostate cancer, NSCLC	20 - 22
Serpine1	serine peptidase inhibitor, clade E, member 1	Breast cancer, several other cancer types	50, 51
Pknox2	Pbx/knotted 1 homeobox 2	---	
Fgfr2	fibroblast growth factor receptor 2	Breast cancer, lung cancer, gastric cancer, ovarian cancer, and endometrial cancer	33
Chrdl1	chordin-like 1	Breast cancer, gastric cancer, lung cancer, melanoma	44 - 48
Igf1	insulin-like growth factor 1	Breast cancer, ovarian cancer, endometrial cancer and pancreatic cancer	35 - 38
Megf6	multiple EGF-like-domains 6	---	
Sema4d	sema domain, immunoglobulin domain (Ig), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 4D	Breast cancer, pancreatic cancer, colorectal cancer, cervical cancer and ovarian cancer	52 - 57
Asprv1	aspartic peptidase, retroviral-like 1	---	
Cxcr4	chemokine (C-X-C motif) receptor 4	More than 23 different cancer types, including breast cancer	49
Ngfr	nerve growth factor receptor (TNFR superfamily, member 16)	Breast cancer, colorectal cancer, melanoma, ovarian cancer	39 - 42
Fam198b	family with sequence similarity 198, member B	---	
Col1a1	collagen, type I, alpha 1	Breast cancer, renal cancer, hepatocellular cancer, gastric cancer, melanoma	23 - 32
Itgb3	integrin beta 3	---	
Crabp1	cellular retinoic acid binding protein I	Breast cancer, ovarian cancer, esophageal squamous cell cancer, colorectal cancer, cervical cancer, papillary thyroid cancer	58 - 63
Pttg1	pituitary tumor-transforming gene 1	Breast cancer, prostate cancer, lung cancer	65 - 67
Acta2	actin, alpha 2, smooth muscle, aorta	Breast cancer, multiple cancer-associated fibroblasts	68 - 70
Rhoc	ras homolog family member C	Multiple cancer types, including breast cancer	71 - 78
Col1a2	collagen, type I, alpha 2	Breast cancer, renal cancer, hepatocellular cancer, gastric cancer, melanoma	23 - 32
Prl2c2	prolactin family 2, subfamily c, member 2	---	
Stac2	SH3 and cysteine rich domain 2	---	
Slc44a1	solute carrier family 44, member 1	---	
Dkk2	dickkopf WNT signaling pathway inhibitor 2	Ovarian cancer, Ewing sarcoma, prostate cancer, renal cancer, colorectal cancer, hepatocellular cancer	79 - 84
Gng11	guanine nucleotide binding protein (G protein), gamma 11	---	
Snrpn	small nuclear ribonucleoprotein N	Medulloblastoma, pancreatic cancer	85, 86

Table 3-2. Cancer-specific associations of genes present within overlap dataset.

Cancer-specific associations of highest DE genes (≥ or ≤ 5 Log2 fold expression; RPKM >5) present within the overlap dataset.

3.4 Discussion

In this chapter, I have described an experimental approach to determine direct transcriptional targets of *Arid1a* gene regulation, using a previously published *in vitro* model system of mammary tumor cells in which *Arid1a* is ectopically overexpressed[16]. By comparing ATAC-seq and RNA-seq profiles, we generated an overlapping dataset of annotated genes that (a) correspond to open chromatin regions and (b) are differentially expressed, both as a direct result of overexpressing *Arid1a* in these tumor cells. While RNA-sequencing reveals the differential abundance of global mRNA transcripts upon overexpressing *Arid1a*, ATAC-sequencing measures genome-wide differences in chromatin accessibility within the same context. Thus, the overlapping dataset represents open chromatin regions that also correspond to differentially expressed genes, and serves as a proxy for direct transcriptional gene regulation by *Arid1a*.

Amongst the most differentially expressed (both upregulated and downregulated) genes present in the overlapping dataset, several of them were found to be cancer-associated genes that have well established roles in human carcinogenesis. More important and relevant to my research question, some of these cancer-associated genes are also known to be associated with SWI/SNF activity. *Fgfr2* has been found to be upregulated in human fibroblasts with impaired mSWI/SNF function, and has been proposed as a therapeutic target in SNF5 (SMARCB1)-deleted malignant rhabdoid tumors[87]. Reduced expression of *Igf1* has been found to occur in BRG1 (SMARCA4)-depleted melanoma cells, and restoring its expression in these cells is sufficient to relieve the stalled cell proliferation imposed by BRG1 depletion[88].

Cxcr4 has been found to be dysregulated in the context of SNF5-deficient atypical teratoid/ rhabdoid (aggressive brain) tumors[89]. In a drosophila model system, *Col1a2* transcription has been shown to be regulated by TGFß-induced chromatin remodeling, brought about by the SWI/SNF complex[90]. Finally, chromatin immunoprecipitation of BRG1 in embryonic stem cells showed localized mSWI/SNF binding near the *Crabp1* promoter, and its expression was found to be decreased upon RNAi targeting of BRG1, suggesting that *Crabp1* is under the direct transcriptional control of BRG1-containing mSWI/SNF complexes[91]. To date, none of these genes have been implicated directly with ARID1A gene regulation, either independently or within its context of mSWI/SNF activity. We propose that they may be novel targets of *ARID1A* regulation based on the significant correlation of chromatin accessibility and differential gene expression. Follow-up experiments validating the physical binding of ARID1A at these specific genetic loci, as well as functional validation of target gene activity upon further perturbation of *Arid1a* expression in this *in vitro* model system would be necessary to unequivocally conclude that they are indeed transcriptional targets of *Arid1a* gene regulation.

3.5 Methods

ATAC Sequencing Library Preparation

Homogenous, single-cell suspensions of both control (untransduced) and experimental (*Arid1a* add-back) samples were generated using 75,000 cells/sample. Cells were centrifuged and lysis buffer (10 mM Tris·Cl, pH 7.4; 10mM NaCl; 3mM MgCl2; 0.1% (v/v) Igepal CA-630) was added to them. Following lysis, cells were

treated with TDE1 ((Nextera Tn5 Transposase from Nextera kit; Illumina, cat. no. FC-121-1030). The transposition reaction was carried out at 37C for 30 mins. Following transposase treatment, DNA was purified using the Qiagen MinElute PCR Purification Kit. Purified DNA was eluted in 10ul Buffer EB. Purified DNA was amplified via PCR in 10 cycles, using barcoded primers described in Supplementary Table 1. Amplified DNA was re-purified using the Qiagen MinElute PCR Purification Kit, and eluted in 20ul Buffer EB. Quality of purified DNA was checked by Bioanalyzer High-Sensitivity DNA Analysis kit (Agilent), and quantification was done using Qubit Fluorophore. Replicate libraries were prepared for both control and experimental samples. All libraries were sequenced on the Illumina HiSeq 2500 with 50bp single-end reads. Approximately 50 million reads were generated/library.

ATAC Sequencing Analysis

Base calling and initial data processing were performed using the standard Illumina protocol. Reads quality control check was investigated by FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Adapter was removed by software trimmomatic[92]. The reads were aligned to mouse reference genome sequence (mm10) using the bowtie2 (2.3.0) aligner[93] with '--local' option specified. Duplicates were removed using Picard and mapping quality of >10 were retained. The peak-calling was done by the ENCODE ATAC-seq pipeline (https://www.encodeproject.org/atac-seq/), with mapping quality parameter >10. Significant and reproducible ATAC-seq peaks produced by Irreproducible Discovery

Rate (IDR) were used for downstream analysis. The peaks overlapped genome features were performed by annotatepeaks.pl function from HOMER[94].

RNA Sequencing

RNA-seq data used for determining the overlapping gene dataset was generated previously in my published research[16].

3.6 References

1. Kadoch, C. & Crabtree, G.R. Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics. *Sci Adv* **1**, e1500447 (2015).
2. Wilson, B.G. & Roberts, C.W. SWI/SNF nucleosome remodellers and cancer. *Nat Rev Cancer* **11**, 481-92 (2011).
3. Hargreaves, D.C. & Crabtree, G.R. ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. *Cell Res* **21**, 396-420 (2011).
4. Ho, L. & Crabtree, G.R. Chromatin remodelling during development. *Nature* **463**, 474-84 (2010).
5. Flores-Alcantar, A., Gonzalez-Sandoval, A., Escalante-Alcalde, D. & Lomeli, H. Dynamics of expression of ARID1A and ARID1B subunits in mouse embryos and in cells during the cell cycle. *Cell Tissue Res* **345**, 137-48 (2011).
6. Chandler, R.L. *et al.* ARID1a-DNA interactions are required for promoter occupancy by SWI/SNF. *Mol Cell Biol* **33**, 265-80 (2013).
7. Nagl, N.G., Jr., Wang, X., Patsialou, A., Van Scoy, M. & Moran, E. Distinct mammalian SWI/SNF chromatin remodeling complexes with opposing roles in cell-cycle control. *Embo j* **26**, 752-63 (2007).
8. Helming, K.C. *et al.* ARID1B is a specific vulnerability in ARID1A-mutant cancers. *Nat Med* **20**, 251-4 (2014).
9. Dallas, P.B. *et al.* The human SWI-SNF complex protein p270 is an ARID family member with non-sequence-specific DNA binding activity. *Mol Cell Biol* **20**, 3137-46 (2000).
10. Wilsker, D. *et al.* The DNA-binding properties of the ARID-containing subunits of yeast and mammalian SWI/SNF complexes. *Nucleic Acids Res* **32**, 1345-53 (2004).
11. Wu, R.C., Wang, T.L. & Shih Ie, M. The emerging roles of ARID1A in tumor suppression. *Cancer Biol Ther* **15**, 655-64 (2014).

12. Bosse, T. *et al.* Loss of ARID1A expression and its relationship with PI3K-Akt pathway alterations, TP53 and microsatellite instability in endometrial cancer. *Mod Pathol* **26**, 1525-35 (2013).
13. Chandler, R.L. *et al.* Coexistent ARID1A-PIK3CA mutations promote ovarian clear-cell tumorigenesis through pro-tumorigenic inflammatory cytokine signalling. *Nat Commun* **6**, 6118 (2015).
14. Guan, B., Wang, T.L. & Shih Ie, M. ARID1A, a factor that promotes formation of SWI/SNF-mediated chromatin remodeling, is a tumor suppressor in gynecologic cancers. *Cancer Res* **71**, 6718-27 (2011).
15. Wu, J.N. & Roberts, C.W. ARID1A mutations in cancer: another epigenetic tumor suppressor? *Cancer Discov* **3**, 35-43 (2013).
16. Kartha, N., Shen, L., Maskin, C., Wallace, M. & Schimenti, J.C. The Chromatin Remodeling Component Arid1a Is a Suppressor of Spontaneous Mammary Tumors in Mice. *Genetics* **203**, 1601-11 (2016).
17. Buenrostro, J.D., Wu, B., Chang, H.Y. & Greenleaf, W.J. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol* **109**, 21.29.1-9 (2015).
18. Ackermann, A.M., Wang, Z., Schug, J., Naji, A. & Kaestner, K.H. Integration of ATAC-seq and RNA-seq identifies human alpha cell and beta cell signature genes. *Mol Metab* **5**, 233-44 (2016).
19. Sengupta, P., Xu, Y., Wang, L., Widom, R. & Smith, B.D. Collagen alpha1(I) gene (COL1A1) is repressed by RFX family. *J Biol Chem* **280**, 21004-14 (2005).
20. Banyard, J., Bao, L. & Zetter, B.R. Type XXIII collagen, a new transmembrane collagen identified in metastatic tumor cells. *J Biol Chem* **278**, 20989-94 (2003).
21. Spivey, K.A. *et al.* Collagen XXIII: a potential biomarker for the detection of primary and recurrent non-small cell lung cancer. *Cancer Epidemiol Biomarkers Prev* **19**, 1362-72 (2010).
22. Banyard, J. *et al.* Collagen XXIII expression is associated with prostate cancer recurrence and distant metastases. *Clin Cancer Res* **13**, 2634-42 (2007).
23. Andre, F. & Cortes, J. Rationale for targeting fibroblast growth factor receptor signaling in breast cancer. *Breast Cancer Res Treat* **150**, 1-8 (2015).
24. Barcus, C.E. *et al.* Elevated collagen-I augments tumor progressive signals, intravasation and metastasis of prolactin-induced estrogen receptor alpha positive mammary tumor cells. *Breast Cancer Res* **19**, 9 (2017).
25. Chai, F. *et al.* Systematically identify key genes in inflammatory and non-inflammatory breast cancer. *Gene* **575**, 600-14 (2016).
26. Chen, S.J., Yuan, W., Lo, S., Trojanowska, M. & Varga, J. Interaction of smad3 with a proximal smad-binding element of the human alpha2(I) procollagen gene promoter required for transcriptional activation by TGF-beta. *J Cell Physiol* **183**, 381-92 (2000).
27. Esbona, K. *et al.* COX-2 modulates mammary tumor progression in response to collagen density. *Breast Cancer Res* **18**, 35 (2016).

28. Loss, L.A. *et al.* Prediction of epigenetically regulated genes in breast cancer cell lines. *BMC Bioinformatics* **11**, 305 (2010).
29. Ibanez de Caceres, I. *et al.* Identification of novel target genes by an epigenetic reactivation screen of renal cancer. *Cancer Res* **66**, 5021-8 (2006).
30. Hayashi, M. *et al.* Identification of the collagen type 1 alpha 1 gene (COL1A1) as a candidate survival-related factor associated with hepatocellular carcinoma. *BMC Cancer* **14**, 108 (2014).
31. Bonazzi, V.F. *et al.* Cross-platform array screening identifies COL1A2, THBS1, TNFRSF10D and UCHL1 as genes frequently silenced by methylation in melanoma. *PLoS One* **6**, e26121 (2011).
32. Li, J., Ding, Y. & Li, A. Identification of COL1A1 and COL1A2 as candidate prognostic factors in gastric cancer. *World J Surg Oncol* **14**, 297 (2016).
33. Katoh, M. Cancer genomics and genetics of FGFR2 (Review). *Int J Oncol* **33**, 233-7 (2008).
34. Haugsten, E.M., Wiedlocha, A., Olsnes, S. & Wesche, J. Roles of fibroblast growth factor receptors in carcinogenesis. *Mol Cancer Res* **8**, 1439-52 (2010).
35. Anisimov, V.N. & Bartke, A. The key role of growth hormone-insulin-IGF-1 signaling in aging and cancer. *Crit Rev Oncol Hematol* **87**, 201-23 (2013).
36. Christopoulos, P.F., Msaouel, P. & Koutsilieris, M. The role of the insulin-like growth factor-1 system in breast cancer. *Mol Cancer* **14**, 43 (2015).
37. Denduluri, S.K. *et al.* Insulin-like growth factor (IGF) signaling in tumorigenesis and the development of cancer drug resistance. *Genes Dis* **2**, 13-25 (2015).
38. Gallagher, E.J. & LeRoith, D. The proliferating role of insulin and insulin-like growth factors in cancer. *Trends Endocrinol Metab* **21**, 610-8 (2010).
39. Tsang, J.Y. *et al.* Nerve growth factor receptor (NGFR): a potential marker for specific molecular subtypes of breast cancer. *J Clin Pathol* **66**, 291-6 (2013).
40. Chan, M.M. & Tahan, S.R. Low-affinity nerve growth factor receptor (P75 NGFR) as a marker of perineural invasion in malignant melanomas. *J Cutan Pathol* **37**, 336-43 (2010).
41. Li, B. *et al.* Nerve growth factor modulates the tumor cells migration in ovarian cancer through the WNT/beta-catenin pathway. *Oncotarget* **7**, 81026-81048 (2016).
42. Yang, Z. *et al.* Epigenetic inactivation and tumor-suppressor behavior of NGFR in human colorectal cancer. *Mol Cancer Res* **13**, 107-19 (2015).
43. Zhou, X. *et al.* Nerve growth factor receptor negates the tumor suppressor p53 as a feedback regulator. *Elife* **5**(2016).
44. Jiang, Y.Z. *et al.* Transcriptome Analysis of Triple-Negative Breast Cancer Reveals an Integrated mRNA-lncRNA Signature with Predictive and Prognostic Value. *Cancer Res* **76**, 2105-14 (2016).
45. Cyr-Depauw, C. *et al.* Chordin-Like 1 Suppresses Bone Morphogenetic Protein 4-Induced Breast Cancer Cell Migration and Invasion. *Mol Cell Biol* **36**, 1509-25 (2016).

46. Pei, Y.F. *et al.* Hypermethylation of the CHRDL1 promoter induces proliferation and metastasis by activating Akt and Erk in gastric cancer. *Oncotarget* **8**, 23155-23166 (2017).
47. Chen, L., Zhuo, D., Chen, J. & Yuan, H. Screening feature genes of lung carcinoma with DNA microarray analysis. *Int J Clin Exp Med* **8**, 12161-71 (2015).
48. Mithani, S.K., Smith, I.M. & Califano, J.A. Use of integrative epigenetic and cytogenetic analyses to identify novel tumor-suppressor genes in malignant melanoma. *Melanoma Res* **21**, 298-307 (2011).
49. Chatterjee, S., Behnam Azad, B. & Nimmagadda, S. The intricate role of CXCR4 in cancer. *Adv Cancer Res* **124**, 31-82 (2014).
50. McMahon, B.J. & Kwaan, H.C. Components of the Plasminogen-Plasmin System as Biologic Markers for Cancer. *Adv Exp Med Biol* **867**, 145-56 (2015).
51. Duffy, M.J., Walsh, S., McDermott, E.W. & Crown, J. Biomarkers in Breast Cancer: Where Are We and Where Are We Going? *Adv Clin Chem* **71**, 1-23 (2015).
52. Jiang, H. *et al.* The role of semaphorin 4D in tumor development and angiogenesis in human breast cancer. *Onco Targets Ther* **9**, 5737-5750 (2016).
53. Yang, Y.H. *et al.* Semaphorin 4D Promotes Skeletal Metastasis in Breast Cancer. *PLoS One* **11**, e0150151 (2016).
54. Chen, Y., Zhang, L., Pan, Y., Ren, X. & Hao, Q. Over-expression of semaphorin4D, hypoxia-inducible factor-1alpha and vascular endothelial growth factor is related to poor prognosis in ovarian epithelial cancer. *Int J Mol Sci* **13**, 13264-74 (2012).
55. Kato, S. *et al.* Semaphorin 4D, a lymphocyte semaphorin, enhances tumor cell motility through binding its receptor, plexinB1, in pancreatic cancer. *Cancer Sci* **102**, 2029-37 (2011).
56. Liu, H. *et al.* Semaphorin 4D expression is associated with a poor clinical outcome in cervical cancer patients. *Microvasc Res* **93**, 1-8 (2014).
57. Wang, J.S. *et al.* Semaphorin 4D and hypoxia-inducible factor-1alpha overexpression is related to prognosis in colorectal carcinoma. *World J Gastroenterol* **21**, 2191-8 (2015).
58. Arellano-Ortiz, A.L. *et al.* DNA Methylation of Cellular Retinoic Acid-Binding Proteins in Cervical Cancer. *Genet Epigenet* **8**, 53-57 (2016).
59. Huang, Y., de la Chapelle, A. & Pellegata, N.S. Hypermethylation, but not LOH, is associated with the low expression of MT1G and CRABP1 in papillary thyroid carcinoma. *Int J Cancer* **104**, 735-44 (2003).
60. Lind, G.E. *et al.* ADAMTS1, CRABP1, and NR3C1 identified as epigenetically deregulated genes in colorectal tumorigenesis. *Cell Oncol* **28**, 259-72 (2006).
61. Miyake, T. *et al.* CRABP1-reduced expression is associated with poorer prognosis in serous and clear cell ovarian adenocarcinoma. *J Cancer Res Clin Oncol* **137**, 715-22 (2011).

62. Tanaka, K. *et al.* Frequent methylation-associated silencing of a candidate tumor-suppressor, CRABP1, in esophageal squamous-cell carcinoma. *Oncogene* **26**, 6456-68 (2007).
63. Liu, R.Z. *et al.* CRABP1 is associated with a poor prognosis in breast cancer: adding to the complexity of breast cancer cell response to retinoic acid. *Mol Cancer* **14**, 129 (2015).
64. Vlotides, G., Eigler, T. & Melmed, S. Pituitary tumor-transforming gene: physiology and implications for tumorigenesis. *Endocr Rev* **28**, 165-86 (2007).
65. Huang, S., Liao, Q., Li, L. & Xin, D. PTTG1 inhibits SMAD3 in prostate cancer cells to promote their proliferation. *Tumour Biol* **35**, 6265-70 (2014).
66. Li, W.H. *et al.* Knockdown of PTTG1 inhibits the growth and invasion of lung adenocarcinoma cells through regulation of TGFB1/SMAD3 signaling. *Int J Immunopathol Pharmacol* **28**, 45-52 (2015).
67. Zhang, G., Zhao, Q., Yu, S., Lin, R. & Yi, X. Pttg1 inhibits TGFbeta signaling in breast cancer cells to promote their growth. *Tumour Biol* **36**, 199-203 (2015).
68. Rudnick, J.A. & Kuperwasser, C. Stromal biomarkers in breast cancer development and progression. *Clin Exp Metastasis* **29**, 663-72 (2012).
69. De Vlieghere, E., Verset, L., Demetter, P., Bracke, M. & De Wever, O. Cancer-associated fibroblasts as target and tool in cancer therapeutics and diagnostics. *Virchows Arch* **467**, 367-82 (2015).
70. Jeon, M. *et al.* Dimerization of EGFR and HER2 induces breast cancer cell motility through STAT1-dependent ACTA2 induction. *Oncotarget* (2016).
71. Allen, S.G. *et al.* Macrophages Enhance Migration in Inflammatory Breast Cancer Cells via RhoC GTPase Signaling. *Sci Rep* **6**, 39190 (2016).
72. Kawata, H. *et al.* RhoC upregulation is correlated with reduced E-cadherin in human breast cancer specimens after chemotherapy and in human breast cancer MCF-7 cells. *Horm Cancer* **5**, 414-23 (2014).
73. Kleer, C.G. *et al.* RhoC-GTPase is a novel tissue biomarker associated with biologically aggressive carcinomas of the breast. *Breast Cancer Res Treat* **93**, 101-10 (2005).
74. Rosenthal, D.T. *et al.* RhoC impacts the metastatic potential and abundance of breast cancer stem cells. *PLoS One* **7**, e40979 (2012).
75. van Golen, K.L., Wu, Z.F., Qiao, X.T., Bao, L. & Merajver, S.D. RhoC GTPase overexpression modulates induction of angiogenic factors in breast cells. *Neoplasia* **2**, 418-25 (2000).
76. van Golen, K.L., Wu, Z.F., Qiao, X.T., Bao, L.W. & Merajver, S.D. RhoC GTPase, a novel transforming oncogene for human mammary epithelial cells that partially recapitulates the inflammatory breast cancer phenotype. *Cancer Res* **60**, 5832-8 (2000).
77. Wu, M., Wu, Z.F., Kumar-Sinha, C., Chinnaiyan, A. & Merajver, S.D. RhoC induces differential expression of genes involved in invasion and metastasis in MCF10A breast cells. *Breast Cancer Res Treat* **84**, 3-12 (2004).

78. Xu, X.D. *et al.* Anti-RhoC siRNAs inhibit the proliferation and invasiveness of breast cancer cells via modulating the KAI1, MMP9, and CXCR4 expression. *Onco Targets Ther* **10**, 1827-1834 (2017).
79. Farkas, S.A., Vymetalkova, V., Vodickova, L., Vodicka, P. & Nilsson, T.K. DNA methylation changes in genes frequently mutated in sporadic colorectal cancer and in the DNA repair and Wnt/beta-catenin signaling pathway genes. *Epigenomics* **6**, 179-91 (2014).
80. Hauer, K. *et al.* DKK2 mediates osteolysis, invasiveness, and metastatic spread in Ewing sarcoma. *Cancer Res* **73**, 967-77 (2013).
81. Hirata, H. *et al.* Wnt antagonist gene DKK2 is epigenetically silenced and inhibits renal cancer progression through apoptotic and cell cycle pathways. *Clin Cancer Res* **15**, 5678-87 (2009).
82. Lin, Y.F. *et al.* Selective Retention of an Inactive Allele of the DKK2 Tumor Suppressor Gene in Hepatocellular Carcinoma. *PLoS Genet* **12**, e1006051 (2016).
83. Xu, W. *et al.* Dickkopf 2 promotes proliferation and invasion via Wnt signaling in prostate cancer. *Mol Med Rep* **14**, 2283-8 (2016).
84. Zhu, J., Zhang, S., Gu, L. & Di, W. Epigenetic silencing of DKK2 and Wnt signal pathway components in human ovarian carcinoma. *Carcinogenesis* **33**, 2334-43 (2012).
85. Jing, J. *et al.* Effect of small nuclear ribonucleoprotein-associated polypeptide N on the proliferation of medulloblastoma cells. *Mol Med Rep* **11**, 3337-43 (2015).
86. Ma, J., Zhang, Z. & Wang, J. Small nuclear ribonucleoprotein associated polypeptide N accelerates cell proliferation in pancreatic adenocarcinoma. *Mol Med Rep* **12**, 6060-4 (2015).
87. Wohrle, S. *et al.* Fibroblast growth factor receptors as novel therapeutic targets in SNF5-deleted malignant rhabdoid tumors. *PLoS One* **8**, e77652 (2013).
88. Ondrusova, L., Vachtenheim, J., Reda, J., Zakova, P. & Benkova, K. MITF-independent pro-survival role of BRG1-containing SWI/SNF complex in melanoma cells. *PLoS One* **8**, e54110 (2013).
89. Sandgren, J. *et al.* Whole Exome- and mRNA-Sequencing of an AT/RT Case Reveals Few Somatic Mutations and Several Deregulated Signalling Pathways in the Context of SMARCB1 Deficiency. *Biomed Res Int* **2015**, 862039 (2015).
90. Lim, M.S. & Jeong, K.W. Role of Flightless-I (Drosophila) homolog in the transcription activation of type I collagen gene mediated by transforming growth factor beta. *Biochem Biophys Res Commun* **454**, 393-8 (2014).
91. Kaeser, M.D., Aslanian, A., Dong, M.Q., Yates, J.R., 3rd & Emerson, B.M. BRD7, a novel PBAF-specific SWI/SNF subunit, is required for target gene activation and repression in embryonic stem cells. *J Biol Chem* **283**, 32254-63 (2008).
92. Bolger, A.M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-20 (2014).

93. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-9 (2012).
94. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-89 (2010).

CHAPTER 4

Induction of *Arid1a* Expression as a Suppressive Mechanism of Mammary Tumor Growth

Nithya Kartha[1,2] and John C. Schimenti[1,2,3]

Affiliations:

[1] Department of Biomedical Sciences

[2] Department of Molecular Biology and Genetics

[3] Center for Vertebrate Genomics

Cornell University, Ithaca, NY 14853, USA

One Sentence Summary: Endogenous activation of intact *Arid1a* allele in heterozygously deleted *Chaos3* MT cell line results in reduced proliferation and colony formation rates.

4.1 Abstract

The chromatin remodeling component *ARID1A* has been recently identified as being altered at high frequencies across a spectrum on human tumor types. It is now recognized as a bona fide tumor suppressor gene, based on experiments conducted in different model systems, both *in vitro* and *in vivo*. Interestingly, *ARID1A* is heterozygously deleted in a substantial fraction of human tumors. In our mouse model for sporadic breast cancer, we have found that the majority of mammary tumors also bear heterozygous deletions of *Arid1a*, and have shown that this monoallelic deletion correlates with expressed transcript levels of the gene. While canonically, tumor suppressor genes (TSGs) are thought to behave in a recessive manner, requiring biallelic loss of function to drive carcinogenesis, it is now becoming clear that certain TSGs can in fact behave in a haploinsufficient manner, with dosage levels playing a critical role in certain types of cancer. In this chapter, I aimed to address the question of whether ARID1A may behave as a haploinsufficient tumor suppressor gene in the context of certain tissue backgrounds and tumor types, particularly focusing on a mammary tumor cell line in which one intact allele of *Arid1a* remains expressed. Using a CRISPR/dCas9-based technique for locus-specific activation of gene expression, I was able to endogenously induce *Arid1a* expression to biallelic levels in these tumor cells, and found that it was sufficient to significantly reduce the rate of proliferation and colony formation, suggestive of the importance of dosage levels in *Arid1a's* tumor suppressive activity.

4.2 Introduction

Sporadic breast cancer (i.e., not associated with inherited neoplasia-driving mutations) accounts for the vast majority of breast cancer cases worldwide (80-85%). Such cancers arise from multiple spontaneous genetic mutations and/or epigenetic modifications, and it is important to be able to differentiate between those that drive cancer formation and those that are a by-product of it. *ARID1A*, encoding an important component of the mammalian SWI/SNF complex, has emerged as one of the most commonly mutated or downregulated genes in diverse cancer types[1-8], including breast cancer[9,10]. ARID1A impacts epigenetic gene regulation by altering chromatin structure around promoters of specific loci in conjunction with its associated SWI/SNF complex components[11,12]. Therefore, its downregulation or mutation in somatic cells can have profound consequences, including inappropriate proliferation[13].

TCGA data has revealed that the majority of cancer-associated alterations in *ARID1A* are inactivating by nature, with either nonsense/frameshift mutations or partial/whole gene deletions detected, based predominantly on tumor type[14]. Most interestingly, a significant number of these tumor mutations/deletions appear to be heterozygous, i.e. affecting a single allele, with detectable protein expression present in ovarian clear cell, gastric, hepatocellular and breast carcinomas[10,15-18]. These observations lead to the hypothesis that *ARID1A* is recurrently mutated/deleted in one allele, but expressed by the remaining allele, thereby suggesting that *ARID1A* levels may facilitate a haploinsufficient effect during carcinogenesis.

While canonically, TSGs were thought to behave according to Knudson's two-hit hypothesis[19], it is now widely accepted that certain TSGs, including TP53 and

PTEN, can behave in a much more potent manner, with even partial inactivation critically contributing to tumor growth/maintenance[20-22]. There are multiple studies that support a haploinsufficient role for *ARID1A*. In breast cancer, two independent studies have found that ARID1A haploinsufficiency correlates significantly with higher risk and poorer prognosis[10,23]. *In vitro*, partial knock-down of ARID1A expression in different cell types results in increased proliferation and colony formation[10,23-25], decreased apoptosis[26] and impaired differentiation[25,27]. *In vivo*, *Arid1a* haploinsufficiency in mice has been found to lead to incomplete embryonic development[27], necessitating the use of conditional knock-out models to study the consequence of complete loss of ARID1A function in specific tissues. This in itself suggests a critical dosage effect of *Arid1a* during development and differentiation, which could potentially translate to carcinogenesis as well.

In my work described in this chapter, I attempt to address this question of whether *Arid1a* may behave as a haploinsufficient TSG using a mammary tumor cell line derived from our lab's mouse model for sporadic breast cancer (the *Chaos3* mouse model), which is heterozygously deleted for *Arid1a*, with the remaining intact allele still expressing it. I have used a CRISPR-based approach to restore expression of *Arid1a* to wild-type (WT) levels via endogenous induction at its promoter. This resulted in a reduced rate of colony formation and proliferation *in vitro*.

4.3 Results

Endogenous induction of Arid1a using Synergistic Activation Mediators

To endogenously induce expression of *Arid1a*, I used a CRISPR/dCAS9 – based transcription activation system called Synergistic Activation Mediators (SAM), developed by Konermann et al.[28]. This method involves the locus-specific activation of gene expression via three main components – (1) a deactivated/dead Cas9 (dCas9) protein fused with a transactivating domain (VP64) (2) MS2 aptamers fused with additional activating co-factors (HSF1 and p65) and (3) a locus-specific small guide RNA (sgRNA) incorporated with MS2 aptamers. Together, when these components are transduced into the target cells, the sgRNA will guide the activating factors (dCas9-VP64 and MS2-p65-HSF1) to the promoter of the targeted gene and induce transcriptional activation (Figure 4-1A). Here, I have used an *in vitro* system consisting of a mammary tumor cell line (23116 MT) derived from the *Chaos3* mouse model which is hemizygously deleted for *Arid1a*, but retains expression of the remaining intact allele.

But first, I needed to evaluate the efficiency of candidate sgRNAs targeting the *Arid1a* promoter locus. I selected 8 sgRNAs spread across the proximal promoter of *Arid1a*, all within 200 base pairs of the transcription start site (TSS) (Figure 4-1B).

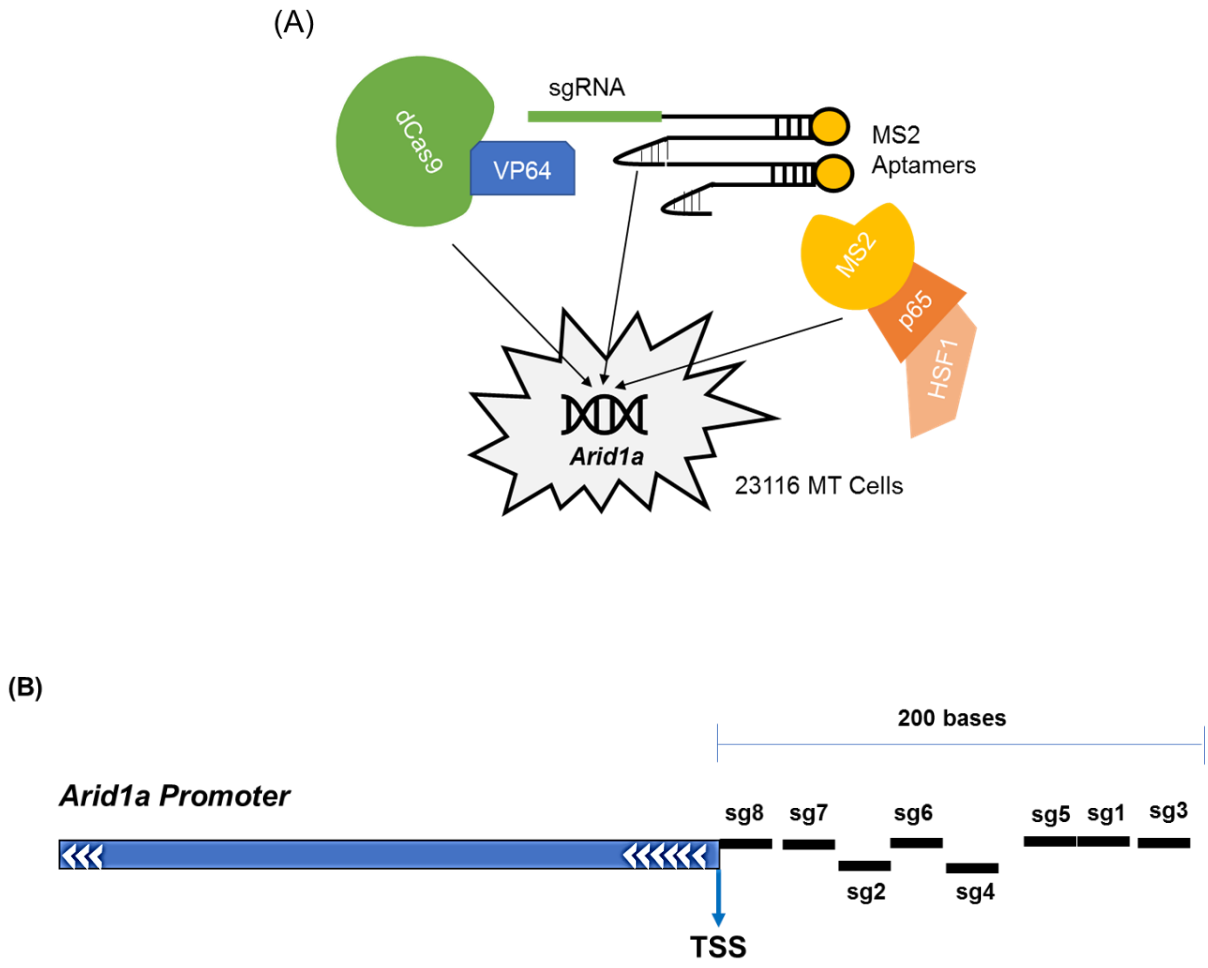


Figure 4-1. *Arid1a* induction using synergistic activation mediators (SAM).

(A) Outline of SAM induction at Arid1a promoter in Chaos3 MT cell line 23116 MT.

(B) Genomic locations on UCSC browser of candidate small guides (sg) tested. Arrows indicate direction of gene transcription.

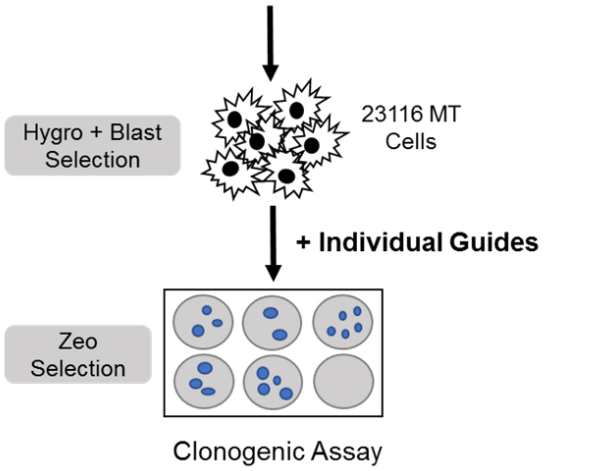
Determining Efficient Candidate Guides for Arid1a Induction

Efficient induction of *Arid1a* expression was determined by testing eight different candidate sgRNAs targeting the *Arid1a* promoter. 23116 MT cells were first co-transduced with the SAM activating factors, and successfully integrated cells were then transduced with individual guides. Following drug selection, the candidate guides were assayed for efficient induction of *Arid1a* by comparing the rate of colony growth (Figure 4-2A). Successful induction on *Arid1a* expression in the tumor cells would expectedly result in either smaller or fewer colonies formed following selective growth. Small guides (sg) 1 and 5 were the most consistently effective at producing smaller/fewer colonies following SAM transduction and selection, relative to the two non-targeting controls (NTCs) (Figure 4-2B).

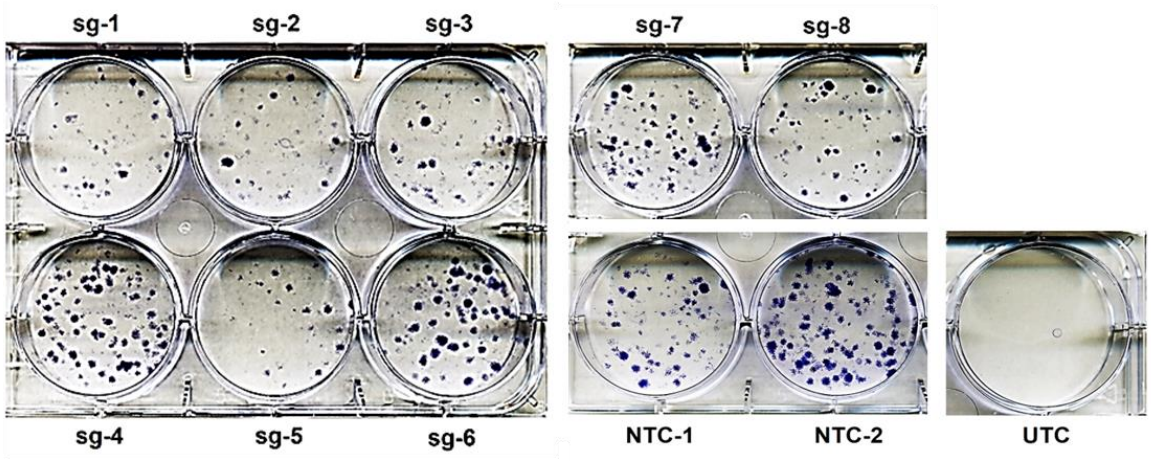
Next, to confirm that the clonogenic assay phenotype was indeed due to activation of *Arid1a* expression, I picked individual colonies from the two representative candidate guides (sg1 and sg5) and the two NTCs and extracted mRNA to determine relative *Arid1a* transcript levels by quantitative real-time PCR (q-RT-PCR). This revealed that of the two candidate guides, only sg5-transduced colonies were upregulated for *Arid1a* expression, on average approximately two-fold relative to the NTCs, whereas the sg1-transduced colonies were not, suggesting that the phenotype observed in that case was due to some off-target effect (Figure 4-2C). Based on these results, I decided to move forward with sg5 as the most efficient candidate guide for endogenous induction of *Arid1a* using SAM.

Figure 4-2. Candidate guides tested to determine induction efficiency.
(A) Schematic outline of SAM transduction and clonogenic assay protocol. (B) Representative clonogenic assay comparing colony formation ability of 23116 MT cells transduced with different sgs; NTC = Non-targeting control; UTC = Untransduced control. (C) q-RT-PCRs quantifying Arid1a transcript levels in induced clonal cell lines and control cell lines; N=number of clonal lines analyzed. Error bars represent SEM.

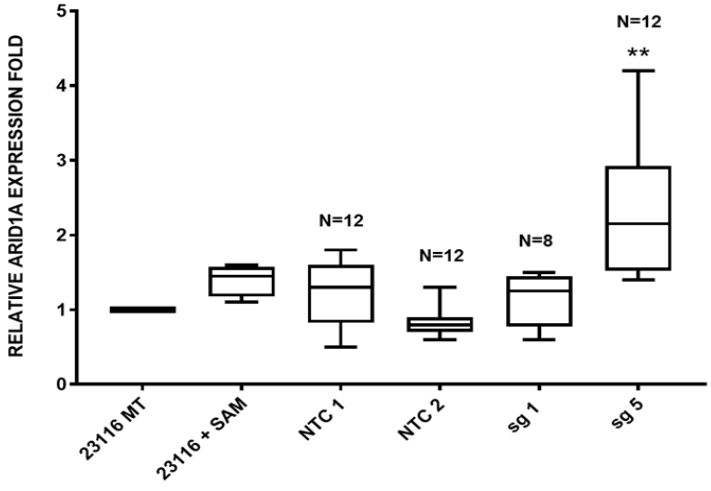
(A)


SAM Vectors (dCas9-VP64 + MS2-p65-HSF1)
Hygro + Blast Selection
23116 MT Cells
+ Individual Guides
Zeo Selection
Clonogenic Assay

(B)


sg-1
sg-2
sg-3
sg-7
sg-8
sg-4
sg-5
sg-6
NTC-1
NTC-2
UTC

(C)


RELATIVE ARID1A EXPRESSION FOLD
0
1
2
3
4
5
N=12
**
N=12
N=12
N=8
23116 MT
23116 + SAM
NTC 1
NTC 2
sg 1
sg 5

Endogenous Induction of Arid1a Expression Significantly Reduces the Rate of Cell Proliferation

Next, I wanted to confirm that the increased expression of *Arid1a* observed in the induced 23116 MT cells actually has an effect on the rate of cell proliferation. For this, I conducted a 5-day growth assay of three representative "induced" cell lines and "uninduced" control cell lines, transduced with sg5 or NTCs, respectively. The induced cell lines showed approximately twice the amount of *Arid1a* expression levels relative to the NTC cell lines (Figure 4-3A). The growth assay showed a significant decrease in the rate of proliferation of all three induced cell lines, compared to the NTC cell lines (Figure 4-3B), thus confirming that induced *Arid1a* expression to WT levels results in a decreased growth rate of cells *in vitro*.

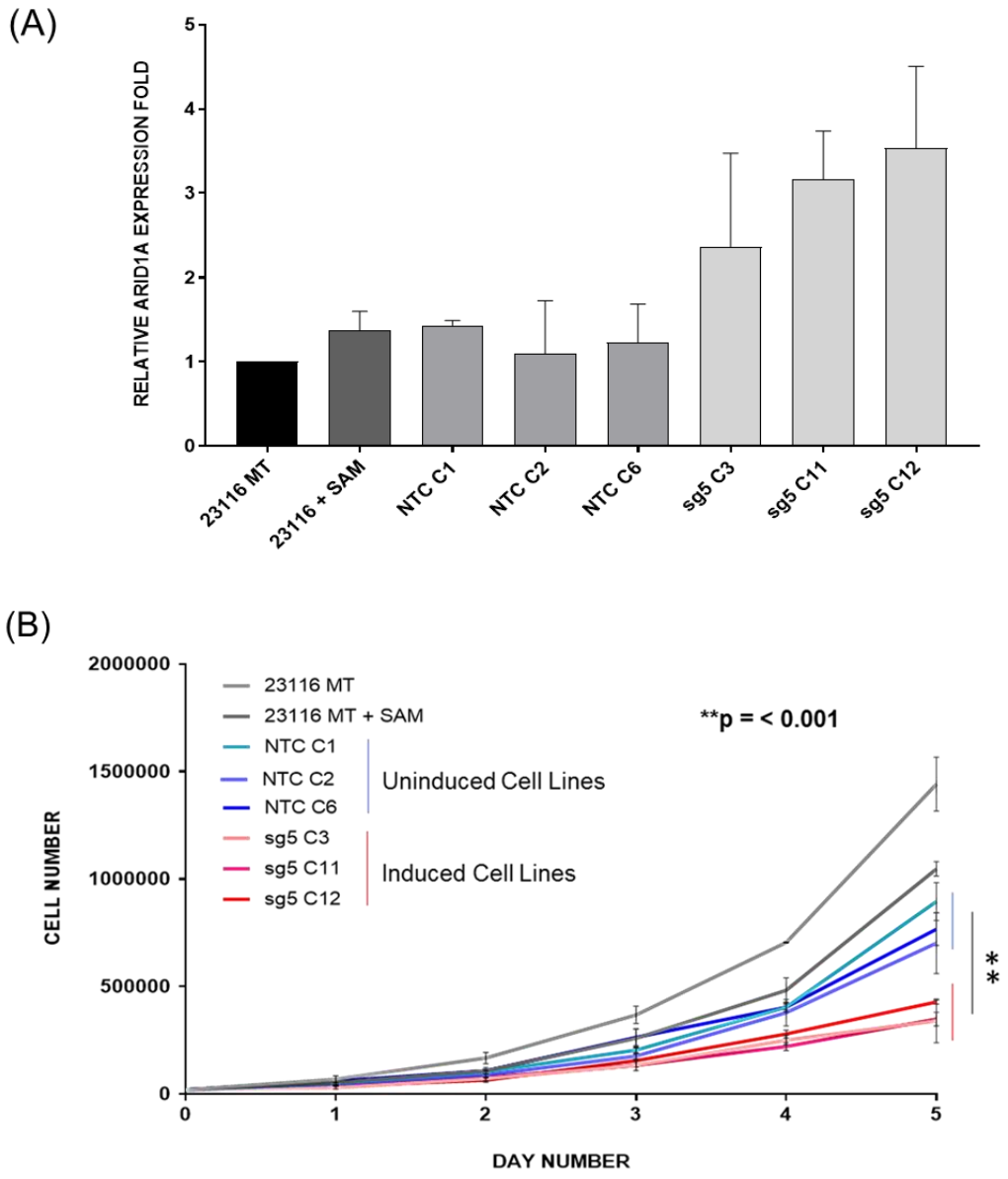


Figure 4-3. Growth assays to determine proliferation rates of induced cell lines. (A) Arid1a mRNA transcript levels of clonal lines assayed for proliferation; Cn = clone number (B) Growth rates of induced and control cell lines over a period of five days. Error bars represent SEM. P value was calculated for averages between the two groups at final data point (Day 5).

Confirming the Specificity of Arid1a induction by SAM

To confirm that the observed reduction in growth in 23116 MT cells following SAM transduction with sg5 was indeed a direct result of endogenous activation of *Arid1a* expression, I used siRNAs targeting *Arid1a* to knock-down its expression in representative "induced" and "uninduced" control cell lines. The efficiency of *Arid1a* knock-down was approximately 40%, assessed via q-RT-PCR (Figure 4-4A). Knocking down *Arid1a* in the induced cells resulted in a significant increase in proliferation, confirming the reversal of the reduced growth phenotype in these cells, while there was no effect on the rate of proliferation in the uninduced control cells (Figure 4-4B). This experiment validates the specificity of endogenous *Arid1a* induction via SAM, using sg5 as a guide.

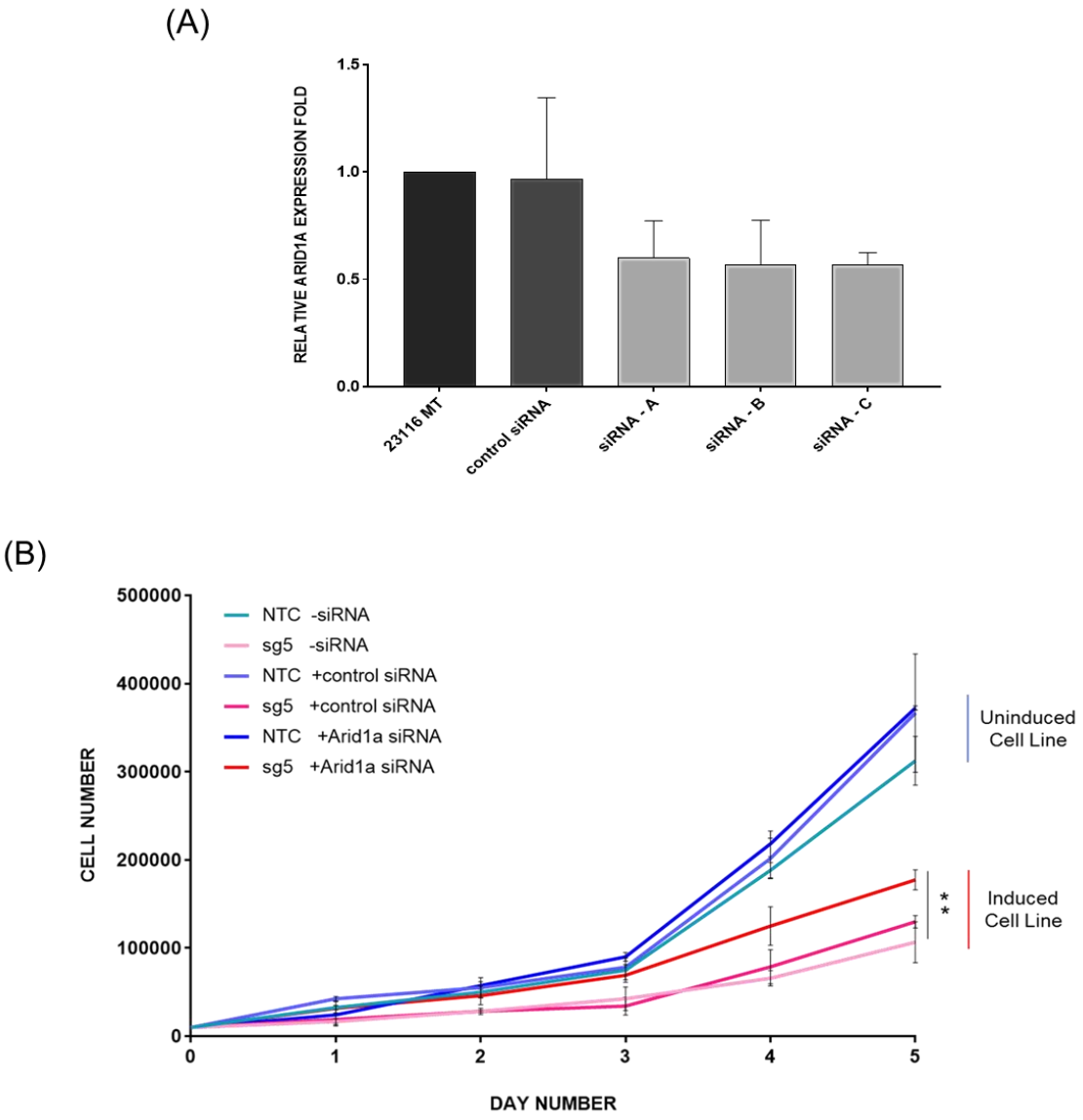


Figure 4-4. siRNA knock-down (KD) of Arid1a in induced clonal cell line to reverse reduced growth phenotype.

Arid1a KD verification by (A) q-RT-PCR and (B) Cell growth assay following Arid1a siRNA KD in induced and control clonal cell lines. Error bars represent SEM. P value was calculated for averages of KD and control cell lines at final data point (Day 5).

4.4 Discussion

In breast cancer studies, partial loss of *ARID1A* expression has been linked with advanced clinical stage and unfavorable outcomes in patients[23,29,30]. Additionally, a significant subset of human breast cancers is heterozygously deleted for *ARID1A*, suggestive of a haploinsufficient tumor suppressive role[9,31,32]. Mammary tumors that developed spontaneously in the *Chaos3* mouse model for sporadic breast cancer were also found to bear mono-allelic deletions of *Arid1a*, and these deletions correlated with reduced *Arid1a* expression levels[33].

To determine whether *Arid1a* dosage was critical for tumor cell growth, I utilized an *in vitro* system consisting of a mammary tumor cell line derived from the *Chaos3* mouse model, in which one copy of *Arid1a* was deleted, and expression levels were less than half relative to wild-type mammary tissue[33]. With the goal of returning *Arid1a* expression to biallelic levels in these tumor cells, I used a CRISPR/dCas9-based activation system to epigenetically induce expression of the remaining intact *Arid1a* allele. Induced *Arid1a* expression levels were confirmed by q-RT-PCR. Phenotypic analyses of tumor cells in with restored *Arid1a* expression showed a decrease in cell proliferation and colony formation, signifying the importance of *Arid1a* dosage in regulating tumor cell growth. Previously, we have shown that ectopic overexpression of *Arid1a* in this same mammary tumor cell line resulted in the alteration of critical oncogenic and tumor suppressive pathways, including TGFß and TP53, and dramatically reduced cell proliferation and tumor growth rates[33]. It is possible that these same pathways are also being altered upon endogenous induction of the remaining *Arid1a* allele to wild-type expression levels, and follow-up

experiments such as whole transcriptome analyses in the *Arid1a*-induced cell lines would be necessary to further characterize this effect.

Therapeutically, this would offer an opportunity for targeting subsets of human cancers that retain a single copy of *ARID1A*, by attempting to decrease tumor growth through endogenous induction of gene expression of this potent tumor suppressor. In my experiments, I utilized a CRISPR/dCas9-based system to endogenously activate *Arid1a* expression and slow down proliferation in mammary tumor cells. Several recent proof-of-principal studies have used similar CRISPR-based techniques to directly alter tumor genomes *in vivo* using mouse models[34-39]. While there are current technical limitations to directly applying such technology in gene therapy for cancer patients, rapid advancements in improving this technology will likely increase the efficiency of delivery of CRISPR- Cas9/dCas9 components, ultimately allowing for the therapeutic alteration of expression for single or multiple cancer driver genes[40-42]. Studies conducted thus far in the year 2017 have successfully used artificial viruses, synthetic nanoparticles and combination viral and non-viral delivery of CRISPR-Cas9 components for efficient gene targeting in both cells and mouse models of disease[43-45]. The ultimate goal will be to help usher in the era of personalized cancer treatment by using advanced technological approaches to (1) identify modifiable genetic/epigenetic alterations in individual patient tumors, (2) characterize the driver alterations and differentiate them from potential passenger alterations and (3) specifically edit cancer genomes by targeting these driver oncogenes and tumor suppressors to regress/halt tumor growth with minimal side-effects.

4.5 Materials and Methods

SAM Plasmids

SAM effector plasmids (dCas9-effector and MS2-effector) and sgRNA backbone plasmid were obtained from Addgene (#61425, # 61426 and #61427). The small guide sequences targeting the *Arid1a* promoter were designed using the bioinformatic tool developed by Feng Zhang's lab (http://sam.genome-engineering.org/database/). Individual sgRNA sequences and non-targeting control sequences (NTCs) were cloned into the empty backbone vector (Addgene #61427) by golden gate assembly, and successful clones were verified by Sanger sequencing.

SAM Lentivirus Production

HEK293T cells were cultured in high-glucose DMEM (Life Technologies), supplemented with 10% fetal bovine serum (Gemini Bio Products), 1% penicillin/streptomycin and sodium pyruvate (Life Technologies). For generating SAM viruses, HEK293T cells were seeded in 10cm dishes at 40-60% confluency, and viral packaging plasmid pMD2.G (2.5ug), viral enveloping plasmid psPAX2 (7.5ug) and each SAM plasmid (10ug) were transfected together into the cells in media not containing any antibiotics. Twelve hours following transfection, viral media was replaced with 10ml fresh media and allowed to incubate for 48 hours at 37C. After this, the first round of virus was collected and stored at 4C. 5ml fresh media was added and cells were incubated for another 24 hours before second round of virus collection. 15 ml of virus was then concentrated by centrifugation in Amicon Ultra filtration columns (EMD Millipore), followed by additional filter sterilization of

concentrated virus using 0.45-um PVDF filters. Virus for each SAM component (dCas9-effector and MS2-effector), and each individual candidate sgRNA was generated independently.

SAM Lentivirus Transduction and Selection

23116 MT cells were cultured in complete medium described above. The day before virus transduction, 250,000 cells were seeded in 35mm cell culture dishes. The next day, freshly prepared and concentrated SAM-effector viruses (dCas9-effector and MS2-effector) were added to the cells and incubated overnight at 37C. The next morning, viral media was replaced with fresh media containing Blasticidin (9ug/ml) and Hygromycin (500ug/ml). Cells were incubated for ~1 week at 37C, under continuous drug selection. Resistant cells were pooled and used to transduce the individual sgRNAs targeting the *Arid1a* promoter region, to determine their efficiencies.

Clonogenic Assays

Pooled 23116 MT cells transduced with SAM-effectors were seeded in 6-well culture dishes at a density of 1000 cells/well. Fresh, concentrated virus generated for 8 sgRNAs and 2 NTCs were added to individual wells. The next morning, viral media was removed and replaced with media containing Zeocin (1mg/ml). Cells were incubated at 37C for ~10 days under drug selection, allowing for the formation of colonies. To visualize the colonies formed, media was removed and each well was rinsed with PBS. 2ml/well fixation solution (3:1 methanol:acetic acid) was added and

cells were incubated at room temperature (RT) for 10mins. Fixation solution was removed and 2ml/well crystal violet (HT90132 Sigma) staining solution was added and cells were incubated at RT for 30mins. Culture dishes were then immersed gently in tap water to rinse of staining solution completely, and then allowed to dry at RT. Colony formation was visualized and documented after scanning whole culture dishes and generating TIFF images.

q-RT-PCR

Total RNA was isolated from cells using the E.Z.N.A. Total RNAKit I (Omega Biotek). A total of 100 ng of RNA was used for cDNA synthesis using the qScript cDNA Supermix Kit (Quantabio). q-RT-PCR analyses was done using Fast SYBR Green Master Mix (Life Technologies) and custom designed primers (Table) for detecting *Arid1a* transcript levels, using *Gapdh* as an endogenous housekeeping control. Assays were run on the CFX96 Touch Real-Time PCR Detection System (Bio-Rad). Each sample was run in triplicate wells, from which mean Ct values were obtained. Relative quantification of gene expression was calculated using the DDCt method. At least two technical replicates were run for each experiment to obtain standard error values.

Cell Proliferation Assays

Representative “induced” and “uninduced” cell lines were seeded in triplicate in 6-well culture dishes at a density of 20,000 cells/well. The next day, cells were washed with PBS and trypsinized, and counted using a hemocytometer. After cell

counts were recorded, the cells were re-seeded in fresh 6-well culture dishes and incubated at 37C. This protocol was repeated for the next four days (total growth assay time = five days). Standard error values were obtained from three experimental replicates.

siRNA Transfection

siRNAs targeting *Arid1a* and a non-targeting scramble control were obtained from Origene (# SR423286, Trilencer-27). siRNAs were first diluted to a stock concentration of 20uM in the duplex buffer provided. The day before transfection, representative "induced" and "uninduced" cell lines were seeded in 6-well culture dishes at a density of 20,000 cells/well and incubated at 37C. The next day, *Arid1a* and control siRNAs were further diluted to a working concentration of 50uM, and then transfection mixtures were set up for a single well as follows – 20ul siTRAN 1.0 (Origene) +120ul Opti-MEM (Life Technologies) + 4ul siRNA (final concentration = 10nM) + 2800ul culture media. Transfection mixtures were incubated at RT for 10 mins before adding it dropwise to cells seeded in 6-well format. Cells were then incubated at 37C.

For validating knock-down efficiency, cells were harvested 48 hours following transfection to assay *Arid1a* mRNA transcript levels by q-RT-PCR. Each siRNA reaction was carried out in triplicate to obtain standard error values.

4.6 References

1. Cajuso, T. *et al.* Exome sequencing reveals frequent inactivating mutations in ARID1A, ARID1B, ARID2 and ARID4A in microsatellite unstable colorectal cancer. *Int J Cancer* **135**, 611-23 (2014).
2. Imielinski, M. *et al.* Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* **150**, 1107-20 (2012).
3. Jones, S. *et al.* Frequent mutations of chromatin remodeling gene ARID1A in ovarian clear cell carcinoma. *Science* **330**, 228-31 (2010).
4. Kandoth, C. *et al.* Integrated genomic characterization of endometrial carcinoma. *Nature* **497**, 67-73 (2013).
5. Liang, H. *et al.* Whole-exome sequencing combined with functional genomics reveals novel candidate driver cancer genes in endometrial cancer. *Genome Res* **22**, 2120-9 (2012).
6. Takeda, T. *et al.* ARID1A gene mutation in ovarian and endometrial cancers (Review). *Oncol Rep* **35**, 607-13 (2016).
7. Waddell, N. *et al.* Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature* **518**, 495-501 (2015).
8. Wang, K. *et al.* Exome sequencing identifies frequent mutation of ARID1A in molecular subtypes of gastric cancer. *Nat Genet* **43**, 1219-23 (2011).
9. Cornen, S. *et al.* Mutations and deletions of ARID1A in breast tumors. *Oncogene* **31**, 4255-6 (2012).
10. Mamo, A. *et al.* An integrated genomic approach identifies ARID1A as a candidate tumor-suppressor gene in breast cancer. *Oncogene* **31**, 2090-100 (2012).
11. Chandler, R.L. *et al.* ARID1a-DNA interactions are required for promoter occupancy by SWI/SNF. *Mol Cell Biol* **33**, 265-80 (2013).
12. Inoue, H. *et al.* Target genes of the largest human SWI/SNF complex subunit control cell growth. *Biochem J* **434**, 83-92 (2011).
13. Romero, O.A. & Sanchez-Cespedes, M. The SWI/SNF genetic blockade: effects in cell differentiation, cancer and developmental diseases. *Oncogene* **33**, 2681-9 (2014).
14. Wu, R.C., Wang, T.L. & Shih Ie, M. The emerging roles of ARID1A in tumor suppression. *Cancer Biol Ther* **15**, 655-64 (2014).
15. Guan, B. *et al.* Mutation and loss of expression of ARID1A in uterine low-grade endometrioid carcinoma. *Am J Surg Pathol* **35**, 625-32 (2011).
16. Wiegand, K.C. *et al.* Loss of BAF250a (ARID1A) is frequent in high-grade endometrial carcinomas. *J Pathol* **224**, 328-33 (2011).
17. Guichard, C. *et al.* Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. *Nat Genet* **44**, 694-8 (2012).
18. Zang, Z.J. *et al.* Exome sequencing of gastric adenocarcinoma identifies recurrent somatic mutations in cell adhesion and chromatin remodeling genes. *Nat Genet* **44**, 570-4 (2012).

19. Knudson, A.G., Jr. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A* **68**, 820-3 (1971).
20. Berger, A.H., Knudson, A.G. & Pandolfi, P.P. A continuum model for tumour suppression. *Nature* **476**, 163-9 (2011).
21. Solimini, N.L. *et al.* Recurrent hemizygous deletions in cancers may optimize proliferative potential. *Science* **337**, 104-9 (2012).
22. Berger, A.H. & Pandolfi, P.P. Haplo-insufficiency: a driving force in cancer. *J Pathol* **223**, 137-46 (2011).
23. Zhang, X. *et al.* Frequent low expression of chromatin remodeling gene ARID1A in breast cancer and its clinical significance. *Cancer Epidemiol* **36**, 288-93 (2012).
24. Huang, J. *et al.* Exome sequencing of hepatitis B virus-associated hepatocellular carcinoma. *Nat Genet* **44**, 1117-21 (2012).
25. Nagl, N.G., Jr., Wang, X., Patsialou, A., Van Scoy, M. & Moran, E. Distinct mammalian SWI/SNF chromatin remodeling complexes with opposing roles in cell-cycle control. *Embo j* **26**, 752-63 (2007).
26. Luo, B. *et al.* Highly parallel identification of essential genes in cancer cells. *Proc Natl Acad Sci U S A* **105**, 20380-5 (2008).
27. Gao, X. *et al.* ES cell pluripotency and germ-layer formation require the SWI/SNF chromatin remodeling component BAF250a. *Proc Natl Acad Sci U S A* **105**, 6656-61 (2008).
28. Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583-8 (2015).
29. Takao, C. *et al.* Downregulation of ARID1A, a component of the SWI/SNF chromatin remodeling complex, in breast cancer. *J Cancer* **8**, 1-8 (2017).
30. Zhao, J., Liu, C. & Zhao, Z. ARID1A: a potential prognostic factor for breast cancer. *Tumour Biol* **35**, 4813-9 (2014).
31. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61-70 (2012).
32. Ciriello, G. *et al.* Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. *Cell* **163**, 506-19 (2015).
33. Kartha, N., Shen, L., Maskin, C., Wallace, M. & Schimenti, J.C. The Chromatin Remodeling Component Arid1a Is a Suppressor of Spontaneous Mammary Tumors in Mice. *Genetics* **203**, 1601-11 (2016).
34. Cortina, C. *et al.* A genome editing approach to study cancer stem cells in human tumors. *EMBO Mol Med* (2017).
35. Gao, M. & Liu, D. CRISPR/Cas9-based Pten Knock-out and Sleeping Beauty Transposon-mediated Nras Knock-in Induces Hepatocellular Carcinoma and Hepatic Lipid Accumulation in Mice. *Cancer Biol Ther*, 0 (2017).
36. Guan, L., Han, Y., Zhu, S. & Lin, J. Application of CRISPR-Cas system in gene therapy: Pre-clinical progress in animal model. *DNA Repair (Amst)* **46**, 1-8 (2016).
37. Roper, J. *et al.* In vivo genome editing and organoid transplantation models of colorectal cancer and metastasis. *Nat Biotechnol* (2017).

38. Wei, L. *et al.* Histone methyltransferase G9a promotes liver cancer development by epigenetic silencing of tumor suppressor gene RARRES3. *J Hepatol* (2017).
39. Xue, W. *et al.* CRISPR-mediated direct mutation of cancer genes in the mouse liver. *Nature* **514**, 380-4 (2014).
40. Sanchez-Rivera, F.J. & Jacks, T. Applications of the CRISPR-Cas9 system in cancer biology. *Nat Rev Cancer* **15**, 387-95 (2015).
41. Sayin, V.I. & Papagiannakopoulos, T. Application of CRISPR-mediated genome engineering in cancer research. *Cancer Lett* **387**, 10-17 (2017).
42. Khan, F.A. *et al.* CRISPR/Cas9 therapeutics: a cure for cancer and other genetic diseases. *Oncotarget* **7**, 52541-52552 (2016).
43. Li, L. *et al.* Artificial Virus Delivers CRISPR-Cas9 System for Genome Editing of Cells in Mice. *ACS Nano* **11**, 95-111 (2017).
44. Miller, J.B. *et al.* Non-Viral CRISPR/Cas Gene Editing In Vitro and In Vivo Enabled by Synthetic Nanoparticle Co-Delivery of Cas9 mRNA and sgRNA. *Angew Chem Int Ed Engl* **56**, 1059-1063 (2017).
45. Yin, H. *et al.* Therapeutic genome editing by combined viral and non-viral delivery of CRISPR system components in vivo. *Nat Biotechnol* **34**, 328-33 (2016).

CHAPTER 5

Summary and Discussion

5.1 Chromatin Remodelers in Cancer

Eukaryotic chromatin consists of highly condensed structural units called nucleosomes, formed by the tight coiling of chromosomal DNA and histone proteins[1]. Chromatin remodeling proteins utilize ATP energy to mobilize nucleosomes, both by direct physical sliding as well as by indirect catalysis of the ejection/insertion of histones, thus mediating the access of DNA-binding proteins to double-stranded DNA in a locus-specific manner[2-4]. Chromatin remodeling proteins thus play important roles in several essential biological processes, including the regulation of cell division, gene expression, DNA repair, and lineage-specification[4,5]. Of these, the SWI/SNF complexes have been the most extensively studied and best characterized in yeast, fruit flies and mammalian model systems[6]. There are two distinct mammalian (m)SWI/SNF complexes – BRG1-associated factor (BAF) and polybromo BRG1-associated factor (PBAF) – that share common core components, but also retain certain exclusive components[4,6]. Recent large-scale genome sequencing studies have revealed that genes encoding mSWI/SNF components are frequently mutated in a variety of human cancers. Of these, the most well studied are the shared components SMARCA4/2 (BRG1/BRM), SMARCB1 (SNF5/BAF47), the PBAF-specific component PBRM1 (BAF180), and the BAF-specific component SMARCF1 (ARID1A/BAF250A)[4,7].

Several studies have characterized the essential role of mSWI/SNF complexes in differentiation of various cell lineages, through its cooperation with tissue-specific transcription factors[8-10]. Since it is thought that there are several hundred variants of the mSWI/SNF complexes based on combinatorial assembly of cell type-specific components, these variant subunits may dictate the binding to lineage-specific factors and thus modulate mSWI/SNF activity at tissue-specific loci[11-13]. This would in part explain the tissue-specific cancer phenotypes of inactivating mutations of mSWI/SNF components observed in mouse models, and the distinct range of cancers associated with mutations in each mSWI/SNF subunit in humans[4].

With respect to the key targets of mSWI/SNF regulation, studies in different model systems have revealed direct interactions with TSGs such as RB and TP53[14-17], suggestively in a functionally redundant manner, as well as with oncogenes such as MYC[18-20]. They have also been found to interact with nuclear-hormone receptors, and stem cell maintenance/differentiation factors[21-25]. Thus, mSWI/SNF modulates critical growth and signaling pathways that are tightly linked to the development of cancer.

5.2 *ARID1A* Mutations in Human Cancer

Recent human cancer genome sequencing studies have identified frequent inactivating *ARID1A* mutations in ovarian[26,27], endometrial[28-31], esophageal[32-34], gastric[35-38], colorectal[39-43], liver[44-50], pancreatic[51-54], breast[40,55-59], bladder[60,61], melanoma[2,62], lymphoma[63-66], neuroblastoma[67,68] and lung cancers[2,40,69-71]. Reduced ARID1A expression has also been detected at the protein level in certain precancerous

lesions, suggesting a critical role for loss of *ARID1A* function in tumor initiation[26,30,58,72-80].

In addition to the correlative genomic data implicating *ARID1A* as a tumor suppressor in human cancers, functional proof for its role in carcinogenesis has been elucidated more recently with the help of multiple different mouse models[25,81-85]. Due to the fact that *Arid1a* is heterozygous lethal in mice[25], *in vivo* analyses of function necessitate conditionally knocking it out in a tissue-specific manner. To date, this experiment has not been done in the mammary gland in the context of breast carcinogenesis, although it is now known that *ARID1A* is deleted in a significant fraction of human mammary tumors, and that low ARID1A expression in tumors of breast cancer patients correlates significantly with poorer prognosis and overall survival[55,57,86,87].

In my research, I have identified *Arid1a* as being recurrently deleted in mammary tumors that arise spontaneously in the *Chaos3* mouse model for breast cancer, and have shown that the loss of a single allele of *Arid1a* is required for maintaining proliferation and growth of cancer cells derived from this model. This is the first mouse model to characterize *Arid1a* as a tumor suppressor gene in sporadic mammary tumors, thus adding to the mounting *in vivo* evidence of the importance of *Arid1a* function in carcinogenesis.

5.3 Mechanisms of *ARID1A* Tumor Suppression

Exome sequencing studies have not only shed light on the frequency of *ARID1A* mutations in human cancers, but have also revealed cooccurring/mutually

exclusive mutations in other cancer pathways, suggestive of potential tumor suppressive mechanisms.

In ovarian cancer, inactivating *ARID1A* mutations commonly coexist with activating phosphatidylinositol 3-kinase (PI3K) mutations[88-91]. A recent study using a mouse model in which *ARID1A* was conditionally deleted in ovarian surface epithelial cells showed that cooccurring mutations in the gene encoding the catalytic subunit of PI3K (PIK3CA) were essential for the formation of ovarian clear cell carcinomas *in vivo*[81]. They further demonstrated that these two pathways converge on carcinogenic cytokine signaling, suggesting a protective role for *ARID1A* against tumor formation driven by inflammation.

In endometrial[91,92], gastric[93,94], esophageal[74] and breast[55] carcinomas, inactivating *ARID1A* and *TP53* mutations frequently occur in a mutually exclusive manner. A study using an ovarian cancer model showed that ARID1A can physically interact with TP53 and help guide p53-mediated transcriptional regulation in association with the mSWI/SNF complex[17]. My thesis work corroborates this model in breast cancer, where we found that the p53 pathway is significantly activated upon overexpressing *Arid1a* in mammary tumor cells, and its tumor suppressive activity appears to be dependent on the presence of functional wild-type p53[95].

Another interesting study showed that inactivating *Arid1a* following induced injury in the liver and outer ear of a mouse model promotes tissue regeneration by curbing chromatin access of specific transcription factors that promote differentiation and suppress proliferation, thus revealing a role for ARID1A in maintaining normal organ structure and function[85].

Mechanistically, mSWI/SNF complexes have been found to function at both promoters and enhancers of genes, depending on the cell/tissue type[96-98]. Since the ARID domain-DNA binding is essential for mSWI/SNF activity[99], chromatin immunoprecipitation (ChIP) of ARID1A would be one way in which to determine direct binding patterns of mSWI/SNF in a specific biological context. One study conducted ChIP-sequencing of ARID1A in a human liver cancer cell line (HepG2), and concluded that in this context *ARID1A* regulates expression primarily at the transcriptional start sites of genes[100]. A different study conducted ChIP-sequencing in a human colorectal cancer cell line (HCT116) with and without *ARID1A*, and showed that in this context ARID1A normally targets mSWI/SNF complexes to enhancers, where they coordinate with other transcription factors to regulate gene activation[84].

Taken together, these important studies have helped elucidate some possible mechanisms by which *ARID1A* functions as a tumor suppressor in different biological contexts, while also revealing how much more we still need to understand to better characterize its role in normal development and carcinogenesis.

5.4 A Potential Haploinsufficient Role for *ARID1A* Tumor Suppression

The striking observation that ARID1A is often heterozygously deleted/mutated in a spectrum of cancers, including breast[57,86], ovarian[101], pancreatic[52,54] and renal[102] carcinomas, suggests that loss of function of a single allele may be sufficient to contribute to carcinogenesis.

Studies using models in which *Arid1a* was monoallelically and biallelically deleted in different cell types have revealed the apparent lineage-specific effect of

Arid1a haploinsufficiency in carcinogenesis. Biallelic deletions of *Arid1a* in mouse ovarian epithelial cells were necessary to drive formation of ovarian clear cell carcinomas, suggesting that in this model, *Arid1a* likely follows the canonical 'two-hit' model of tumor suppression[81]. However, in comparing HCT116 isogenic cell lines bearing monoallelic and biallelic deletions of *ARID1A*, partial loss of *ARID1A* function was found to affect enhancer activity and gene transcription, suggestive of a haploinsufficient effect in this model[84].

In my research, I have attempted to approach this question using a mammary tumor cell line derived from the *Chaos3* mouse model for sporadic breast cancer, in which only one allele of *Arid1a* is intact and expressed. I have utilized a CRISPR/dCas9-based system to study the effect of endogenously restore *Arid1a* expression to biallelic levels in these tumor cells. Through these experiments, I have shown that restoring *Arid1a* expression to "wild-type" levels is sufficient to significantly reduce cancer cell proliferation and colony formation rates *in vitro*.

5.5 Therapeutic Intervention in *ARID1A* Mutant Cancers

Since mSWI/SNF components are altered in nearly 20% of all human cancers, identifying mechanisms by which we can therapeutically target these cancers would be extremely important and beneficial. Recent studies have shown that in some models mSWI/SNF-mutant cancers depend on residual mSWI/SNF complexes for malignant growth, while in others they rely on co-dependent oncogenic/tumor suppressive pathways[7]. Multiple synthetic lethal screens have been conducted to identify such

interactions which may be exploited for targeted therapy of mSWI/SNF-driven cancers[7,103].

With respect to cancers in which *ARID1A* is altered/mutated, associated gene/pathway vulnerabilities identified thus far are ARID1B, PIK3/AKT, EZH2, YES1, PARP and ATR[104-108]. These studies use a variety of human cancer cell lines as well as *in vivo* mouse models, all of which are based on inactivating *ARID1A* point mutations.

As we know that there is a significant subset of tumors that are heterozygously deleted for *ARID1A*, but still retain low levels of expression from the remaining allele, a compelling therapeutic avenue to explore would be in trying to activate the remaining intact allele in these tumors. To date, the activation of TSGs as a method of clinical cancer treatment has been mainly through indirect and indiscriminate methods, by using so-called "Epi-drugs" (epigenetic drugs) that inhibit epigenetic enzymes – mainly DNA methyltransferase inhibitors (DNMTis) and histone deacetylase inhibitors (HDACis) – and result in gene upregulation[109]. Current FDA approved DNMTis (eg. azacitidine and decitabine) and HDACis (eg. vorinostat and romidepsin), are only effective on TSGs that are silenced by these epigenetic mechanisms, but due to their lack of target selectivity, can cause detrimental genome wide off-target effects, such as the upregulation of prometastatic genes[109-111].

Single gene cancer therapy has been a relatively unexplored method of treatment up till this point. One study conducted nearly ten years ago explored this possibility by administering wild-type TP53 (in combination with standard chemotherapy) using replication-deficient adenoviral vectors, to a group of patients

with primary stage III ovarian cancers that were mutant for p53[112]. While this trial proved unsuccessful, amongst the main reasons attributed to its failure were (a) poor infective efficiency of the vectors used for transmission of the gene (b) the potential dominant-negative effect of mutant p53 that was present in these tumors[113]. The first issue represents a fixable technical problem, and current adenoviral vectors are greatly improved in their efficacy of delivery[114-116]. The second issue is specific to mutant TP53 function, and would not necessarily be relevant to other tumor suppressor genes. Thus, this is still an important therapeutic avenue that could be revisited and explored, as we gain more knowledge of genes that have potent tumor suppressive activity on their own, such as ARID1A.

5.6 References

1. Yaniv, M. Chromatin remodeling: from transcription to cancer. *Cancer Genet* **207**, 352-7 (2014).
2. Oike, T., Ogiwara, H., Nakano, T., Yokota, J. & Kohno, T. Inactivating mutations in SWI/SNF chromatin remodeling genes in human cancer. *Jpn J Clin Oncol* **43**, 849-55 (2013).
3. Saha, A., Wittmeyer, J. & Cairns, B.R. Chromatin remodelling: the industrial revolution of DNA around histones. *Nat Rev Mol Cell Biol* **7**, 437-47 (2006).
4. Wilson, B.G. & Roberts, C.W. SWI/SNF nucleosome remodellers and cancer. *Nat Rev Cancer* **11**, 481-92 (2011).
5. Euskirchen, G., Auerbach, R.K. & Snyder, M. SWI/SNF chromatin-remodeling factors: multiscale analyses and diverse functions. *J Biol Chem* **287**, 30897-905 (2012).
6. Kadoch, C. & Crabtree, G.R. Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics. *Sci Adv* **1**, e1500447 (2015).
7. St Pierre, R. & Kadoch, C. Mammalian SWI/SNF complexes in cancer: emerging therapeutic opportunities. *Curr Opin Genet Dev* **42**, 56-67 (2017).
8. de la Serna, I.L., Ohkawa, Y. & Imbalzano, A.N. Chromatin remodelling in mammalian differentiation: lessons from ATP-dependent remodellers. *Nat Rev Genet* **7**, 461-73 (2006).

9. Flowers, S., Nagl, N.G., Jr., Beck, G.R., Jr. & Moran, E. Antagonistic roles for BRM and BRG1 SWI/SNF complexes in differentiation. *J Biol Chem* **284**, 10067-75 (2009).
10. Young, D.W. *et al.* SWI/SNF chromatin remodeling complex is obligatory for BMP2-induced, Runx2-dependent skeletal gene expression that controls osteoblast differentiation. *J Cell Biochem* **94**, 720-30 (2005).
11. Wu, J.I. *et al.* Regulation of dendritic development by neuron-specific chromatin remodeling complexes. *Neuron* **56**, 94-108 (2007).
12. Wu, J.I., Lessard, J. & Crabtree, G.R. Understanding the words of chromatin regulation. *Cell* **136**, 200-6 (2009).
13. Lickert, H. *et al.* Baf60c is essential for function of BAF chromatin remodelling complexes in heart development. *Nature* **432**, 107-12 (2004).
14. Trouche, D., Le Chalony, C., Muchardt, C., Yaniv, M. & Kouzarides, T. RB and hbrm cooperate to repress the activation functions of E2F1. *Proc Natl Acad Sci U S A* **94**, 11268-73 (1997).
15. Isakoff, M.S. *et al.* Inactivation of the Snf5 tumor suppressor stimulates cell cycle progression and cooperates with p53 loss in oncogenic transformation. *Proc Natl Acad Sci U S A* **102**, 17745-50 (2005).
16. Nagl, N.G., Jr. *et al.* The p270 (ARID1A/SMARCF1) subunit of mammalian SWI/SNF-related complexes is essential for normal cell cycle arrest. *Cancer Res* **65**, 9236-44 (2005).
17. Guan, B., Wang, T.L. & Shih Ie, M. ARID1A, a factor that promotes formation of SWI/SNF-mediated chromatin remodeling, is a tumor suppressor in gynecologic cancers. *Cancer Res* **71**, 6718-27 (2011).
18. Nagl, N.G., Jr., Zweitzig, D.R., Thimmapaya, B., Beck, G.R., Jr. & Moran, E. The c-myc gene is a direct target of mammalian SWI/SNF-related complexes during differentiation-associated cell cycle arrest. *Cancer Res* **66**, 1289-93 (2006).
19. Cheng, S.W. *et al.* c-MYC interacts with INI1/hSNF5 and requires the SWI/SNF complex for transactivation function. *Nat Genet* **22**, 102-5 (1999).
20. Nagl, N.G., Jr., Wang, X., Patsialou, A., Van Scoy, M. & Moran, E. Distinct mammalian SWI/SNF chromatin remodeling complexes with opposing roles in cell-cycle control. *Embo j* **26**, 752-63 (2007).
21. Kidder, B.L., Palmer, S. & Knott, J.G. SWI/SNF-Brg1 regulates self-renewal and occupies core pluripotency-related genes in embryonic stem cells. *Stem Cells* **27**, 317-28 (2009).
22. Ho, L. *et al.* An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. *Proc Natl Acad Sci U S A* **106**, 5187-91 (2009).
23. Yan, Z. *et al.* BAF250B-associated SWI/SNF chromatin-remodeling complex is required to maintain undifferentiated mouse embryonic stem cells. *Stem Cells* **26**, 1155-65 (2008).
24. Trotter, K.W. & Archer, T.K. The BRG1 transcriptional coregulator. *Nucl Recept Signal* **6**, e004 (2008).

25. Gao, X. *et al.* ES cell pluripotency and germ-layer formation require the SWI/SNF chromatin remodeling component BAF250a. *Proc Natl Acad Sci U S A* **105**, 6656-61 (2008).
26. Itamochi, H. *et al.* Loss of ARID1A expression is associated with poor prognosis in patients with stage I/II clear cell carcinoma of the ovary. *Int J Clin Oncol* **20**, 967-73 (2015).
27. Coatham, M. *et al.* Concurrent ARID1A and ARID1B inactivation in endometrial and ovarian dedifferentiated carcinomas. *Mod Pathol* **29**, 1586-1593 (2016).
28. Wiegand, K.C. *et al.* Loss of BAF250a (ARID1A) is frequent in high-grade endometrial carcinomas. *J Pathol* **224**, 328-33 (2011).
29. Liang, H. *et al.* Whole-exome sequencing combined with functional genomics reveals novel candidate driver cancer genes in endometrial cancer. *Genome Res* **22**, 2120-9 (2012).
30. Mao, T.L. *et al.* Loss of ARID1A expression correlates with stages of tumor progression in uterine endometrioid carcinoma. *Am J Surg Pathol* **37**, 1342-8 (2013).
31. Guan, B. *et al.* Mutation and loss of expression of ARID1A in uterine low-grade endometrioid carcinoma. *Am J Surg Pathol* **35**, 625-32 (2011).
32. Ozawa, Y. *et al.* Decreased expression of ARID1A contributes to infiltrative growth of esophageal squamous cell carcinoma. *Tohoku J Exp Med* **235**, 185-91 (2015).
33. Nakazato, H. *et al.* Early-Stage Induction of SWI/SNF Mutations during Esophageal Squamous Cell Carcinogenesis. *PLoS One* **11**, e0147372 (2016).
34. Dulak, A.M. *et al.* Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. *Nat Genet* **45**, 478-86 (2013).
35. Aso, T., Uozaki, H., Morita, S., Kumagai, A. & Watanabe, M. Loss of ARID1A, ARID1B, and ARID2 Expression During Progression of Gastric Cancer. *Anticancer Res* **35**, 6819-27 (2015).
36. Takeshima, H. *et al.* Frequent involvement of chromatin remodeler alterations in gastric field cancerization. *Cancer Lett* **357**, 328-38 (2015).
37. Abe, H. *et al.* ARID1A expression loss in gastric cancer: pathway-dependent roles with and without Epstein-Barr virus infection and microsatellite instability. *Virchows Arch* **461**, 367-77 (2012).
38. Chong, I.Y. *et al.* The genomic landscape of oesophagogastric junctional adenocarcinoma. *J Pathol* **231**, 301-10 (2013).
39. Torabi, K. *et al.* Patterns of somatic uniparental disomy identify novel tumor suppressor genes in colorectal cancer. *Carcinogenesis* **36**, 1103-10 (2015).
40. Jones, S. *et al.* Somatic mutations in the chromatin remodeling gene ARID1A occur in several tumor types. *Hum Mutat* **33**, 100-3 (2012).
41. Chou, A. *et al.* Loss of ARID1A expression in colorectal carcinoma is strongly associated with mismatch repair deficiency. *Hum Pathol* **45**, 1697-703 (2014).

42. Cajuso, T. *et al.* Exome sequencing reveals frequent inactivating mutations in ARID1A, ARID1B, ARID2 and ARID4A in microsatellite unstable colorectal cancer. *Int J Cancer* **135**, 611-23 (2014).
43. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330-7 (2012).
44. Zucman-Rossi, J., Villanueva, A., Nault, J.C. & Llovet, J.M. Genetic Landscape and Biomarkers of Hepatocellular Carcinoma. *Gastroenterology* **149**, 1226-1239.e4 (2015).
45. Zou, S. *et al.* Mutational landscape of intrahepatic cholangiocarcinoma. *Nat Commun* **5**, 5696 (2014).
46. Ruzzenente, A. *et al.* Cholangiocarcinoma Heterogeneity Revealed by Multigene Mutational Profiling: Clinical and Prognostic Relevance in Surgically Resected Patients. *Ann Surg Oncol* **23**, 1699-707 (2016).
47. Jiao, Y. *et al.* Exome sequencing identifies frequent inactivating mutations in BAP1, ARID1A and PBRM1 in intrahepatic cholangiocarcinomas. *Nat Genet* **45**, 1470-3 (2013).
48. Ito, T. *et al.* Genomic and transcriptional alterations of cholangiocarcinoma. *J Hepatobiliary Pancreat Sci* **21**, 380-7 (2014).
49. Guichard, C. *et al.* Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. *Nat Genet* **44**, 694-8 (2012).
50. Fujimoto, A. *et al.* Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nat Genet* **44**, 760-4 (2012).
51. Waddell, N. *et al.* Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature* **518**, 495-501 (2015).
52. Shain, A.H. *et al.* Convergent structural alterations define SWItch/Sucrose NonFermentable (SWI/SNF) chromatin remodeler as a central tumor suppressive complex in pancreatic cancer. *Proc Natl Acad Sci U S A* **109**, E252-9 (2012).
53. Jiao, Y. *et al.* Whole-exome sequencing of pancreatic neoplasms with acinar differentiation. *J Pathol* **232**, 428-35 (2014).
54. Birnbaum, D.J. *et al.* Genome profiling of pancreatic adenocarcinoma. *Genes Chromosomes Cancer* **50**, 456-65 (2011).
55. Zhang, X. *et al.* Frequent low expression of chromatin remodeling gene ARID1A in breast cancer and its clinical significance. *Cancer Epidemiol* **36**, 288-93 (2012).
56. Meric-Bernstam, F. *et al.* Concordance of genomic alterations between primary and recurrent breast cancer. *Mol Cancer Ther* **13**, 1382-9 (2014).
57. Mamo, A. *et al.* An integrated genomic approach identifies ARID1A as a candidate tumor-suppressor gene in breast cancer. *Oncogene* **31**, 2090-100 (2012).
58. Luchini, C. *et al.* Prognostic role and implications of mutation status of tumor suppressor gene ARID1A in cancer: a systematic review and meta-analysis. *Oncotarget* **6**, 39088-97 (2015).

59. Huang, J., Zhao, Y.L., Li, Y., Fletcher, J.A. & Xiao, S. Genomic and functional evidence for an ARID1A tumor suppressor role. *Genes Chromosomes Cancer* **46**, 745-50 (2007).
60. Gui, Y. *et al.* Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. *Nat Genet* **43**, 875-8 (2011).
61. Balbas-Martinez, C. *et al.* ARID1A alterations are associated with FGFR3-wild type, poor-prognosis, urothelial bladder tumors. *PLoS One* **8**, e62483 (2013).
62. van de Nes, J. *et al.* Targeted next generation sequencing reveals unique mutation profile of primary melanocytic tumors of the central nervous system. *J Neurooncol* **127**, 435-44 (2016).
63. Nastase, A. *et al.* Genomic and proteomic characterization of ARID1A chromatin remodeller in ampullary tumors. *Am J Cancer Res* **7**, 484-502 (2017).
64. Lunning, M.A. & Green, M.R. Mutation of chromatin modifiers; an emerging hallmark of germinal center B-cell lymphomas. *Blood Cancer J* **5**, e361 (2015).
65. Love, C. *et al.* The genetic landscape of mutations in Burkitt lymphoma. *Nat Genet* **44**, 1321-5 (2012).
66. Giulino-Roth, L. *et al.* Targeted genomic sequencing of pediatric Burkitt lymphoma identifies recurrent alterations in antiapoptotic and chromatin-remodeling genes. *Blood* **120**, 5181-4 (2012).
67. Takeuchi, T. *et al.* Expression of SMARCF1, a truncated form of SWI1, in neuroblastoma. *Am J Pathol* **158**, 663-72 (2001).
68. Sausen, M. *et al.* Integrated genomic analyses identify ARID1A and ARID1B alterations in the childhood cancer neuroblastoma. *Nat Genet* **45**, 12-7 (2013).
69. Imielinski, M. *et al.* Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* **150**, 1107-20 (2012).
70. Hao, C. *et al.* Gene mutations in primary tumors and corresponding patient-derived xenografts derived from non-small cell lung cancer. *Cancer Lett* **357**, 179-85 (2015).
71. Fernandez-Cuesta, L. *et al.* Frequent mutations in chromatin-remodelling genes in pulmonary carcinoids. *Nat Commun* **5**, 3518 (2014).
72. Samartzis, E.P. *et al.* Loss of ARID1A/BAF250a-expression in endometriosis: a biomarker for risk of carcinogenic transformation? *Mod Pathol* **25**, 885-92 (2012).
73. Xiao, W., Awadallah, A. & Xin, W. Loss of ARID1A/BAF250a expression in ovarian endometriosis and clear cell carcinoma. *Int J Clin Exp Pathol* **5**, 642-50 (2012).
74. Streppel, M.M. *et al.* Next-generation sequencing of endoscopic biopsies identifies ARID1A as a tumor-suppressor gene in Barrett's esophagus. *Oncogene* **33**, 347-57 (2014).
75. Kim, Y.B., Ham, I.H., Hur, H. & Lee, D. Various ARID1A expression patterns and their clinical significance in gastric cancers. *Hum Pathol* **49**, 61-70 (2016).

76. Lee, S.Y. *et al.* Loss of AT-rich interactive domain 1A expression in gastrointestinal malignancies. *Oncology* **88**, 234-40 (2015).
77. Park, J.H. *et al.* Decreased ARID1A expression correlates with poor prognosis of clear cell renal cell carcinoma. *Hum Pathol* **46**, 454-60 (2015).
78. Wei, X.L. *et al.* Clinicopathologic and prognostic relevance of ARID1A protein loss in colorectal cancer. *World J Gastroenterol* **20**, 18404-12 (2014).
79. Ye, S. *et al.* Clinicopathologic Significance of HNF-1beta, AIRD1A, and PIK3CA Expression in Ovarian Clear Cell Carcinoma: A Tissue Microarray Study of 130 Cases. *Medicine (Baltimore)* **95**, e3003 (2016).
80. Zhao, J. *et al.* The Clinicopathologic Significance of BAF250a (ARID1A) Expression in Hepatocellular Carcinoma. *Pathol Oncol Res* **22**, 453-9 (2016).
81. Chandler, R.L. *et al.* Coexistent ARID1A-PIK3CA mutations promote ovarian clear-cell tumorigenesis through pro-tumorigenic inflammatory cytokine signalling. *Nat Commun* **6**, 6118 (2015).
82. Fang, J.Z. *et al.* Hepatocyte-Specific Arid1a Deficiency Initiates Mouse Steatohepatitis and Hepatocellular Carcinoma. *PLoS One* **10**, e0143042 (2015).
83. Guan, B. *et al.* Roles of deletion of Arid1a, a tumor suppressor, in mouse ovarian tumorigenesis. *J Natl Cancer Inst* **106**(2014).
84. Mathur, R. *et al.* ARID1A loss impairs enhancer-mediated gene regulation and drives colon cancer in mice. *Nat Genet* **49**, 296-302 (2017).
85. Sun, X. *et al.* Suppression of the SWI/SNF Component Arid1a Promotes Mammalian Regeneration. *Cell Stem Cell* **18**, 456-66 (2016).
86. Cornen, S. *et al.* Mutations and deletions of ARID1A in breast tumors. *Oncogene* **31**, 4255-6 (2012).
87. Zhao, J., Liu, C. & Zhao, Z. ARID1A: a potential prognostic factor for breast cancer. *Tumour Biol* **35**, 4813-9 (2014).
88. Jones, S. *et al.* Frequent mutations of chromatin remodeling gene ARID1A in ovarian clear cell carcinoma. *Science* **330**, 228-31 (2010).
89. Huang, H.N., Lin, M.C., Huang, W.C., Chiang, Y.C. & Kuo, K.T. Loss of ARID1A expression and its relationship with PI3K-Akt pathway alterations and ZNF217 amplification in ovarian clear cell carcinoma. *Mod Pathol* **27**, 983-90 (2014).
90. Yamamoto, S., Tsuda, H., Takano, M., Tamai, S. & Matsubara, O. Loss of ARID1A protein expression occurs as an early event in ovarian clear-cell carcinoma development and frequently coexists with PIK3CA mutations. *Mod Pathol* **25**, 615-24 (2012).
91. Bosse, T. *et al.* Loss of ARID1A expression and its relationship with PI3K-Akt pathway alterations, TP53 and microsatellite instability in endometrial cancer. *Mod Pathol* **26**, 1525-35 (2013).
92. Allo, G. *et al.* ARID1A loss correlates with mismatch repair deficiency and intact p53 expression in high-grade endometrial carcinomas. *Mod Pathol* **27**, 255-61 (2014).
93. Wang, K. *et al.* Exome sequencing identifies frequent mutation of ARID1A in molecular subtypes of gastric cancer. *Nat Genet* **43**, 1219-23 (2011).

94. Zang, Z.J. *et al.* Exome sequencing of gastric adenocarcinoma identifies recurrent somatic mutations in cell adhesion and chromatin remodeling genes. *Nat Genet* **44**, 570-4 (2012).
95. Kartha, N., Shen, L., Maskin, C., Wallace, M. & Schimenti, J.C. The Chromatin Remodeling Component Arid1a Is a Suppressor of Spontaneous Mammary Tumors in Mice. *Genetics* **203**, 1601-11 (2016).
96. Euskirchen, G.M. *et al.* Diverse roles and interactions of the SWI/SNF chromatin remodeling complex revealed using global approaches. *PLoS Genet* **7**, e1002008 (2011).
97. Tolstorukov, M.Y. *et al.* Swi/Snf chromatin remodeling/tumor suppressor complex establishes nucleosome occupancy at target promoters. *Proc Natl Acad Sci U S A* **110**, 10165-70 (2013).
98. Alexander, J.M. *et al.* Brg1 modulates enhancer activation in mesoderm lineage commitment. *Development* **142**, 1418-30 (2015).
99. Chandler, R.L. *et al.* ARID1a-DNA interactions are required for promoter occupancy by SWI/SNF. *Mol Cell Biol* **33**, 265-80 (2013).
100. Raab, J.R., Resnick, S. & Magnuson, T. Genome-Wide Transcriptional Regulation Mediated by Biochemically Distinct SWI/SNF Complexes. *PLoS Genet* **11**, e1005748 (2015).
101. Kandoth, C. *et al.* Integrated genomic characterization of endometrial carcinoma. *Nature* **497**, 67-73 (2013).
102. Lichner, Z. *et al.* The chromatin remodeling gene ARID1A is a new prognostic marker in clear cell renal cell carcinoma. *Am J Pathol* **182**, 1163-70 (2013).
103. Hohmann, A.F. & Vakoc, C.R. A rationale to target the SWI/SNF complex for cancer therapy. *Trends Genet* **30**, 356-63 (2014).
104. Bitler, B.G. *et al.* Synthetic lethality by targeting EZH2 methyltransferase activity in ARID1A-mutated cancers. *Nat Med* **21**, 231-8 (2015).
105. Bitler, B.G., Aird, K.M. & Zhang, R. Epigenetic synthetic lethality in ovarian clear cell carcinoma: EZH2 and ARID1A mutations. *Mol Cell Oncol* **3**, e1032476 (2016).
106. Miller, R.E. *et al.* Synthetic Lethal Targeting of ARID1A-Mutant Ovarian Clear Cell Tumors with Dasatinib. *Mol Cancer Ther* **15**, 1472-84 (2016).
107. Samartzis, E.P. *et al.* Loss of ARID1A expression sensitizes cancer cells to PI3K- and AKT-inhibition. *Oncotarget* **5**, 5295-303 (2014).
108. Williamson, C.T. *et al.* ATR inhibitors as a synthetic lethal therapy for tumours deficient in ARID1A. *Nat Commun* **7**, 13837 (2016).
109. Falahi, F., Sgro, A. & Blancafort, P. Epigenome engineering in cancer: fairytale or a realistic path to the clinic? *Front Oncol* **5**, 22 (2015).
110. Duijkers, F.A. *et al.* Epigenetic drug combination induces genome-wide demethylation and altered gene expression in neuro-ectodermal tumor-derived cell lines. *Cell Oncol (Dordr)* **36**, 351-62 (2013).
111. Yu, Y. *et al.* Epigenetic drugs can stimulate metastasis through enhanced expression of the pro-metastatic Ezrin gene. *PLoS One* **5**, e12710 (2010).
112. Haupt, S. & Haupt, Y. Manipulation of the tumor suppressor p53 for potentiating cancer therapy. *Semin Cancer Biol* **14**, 244-52 (2004).

113. Zeimet, A.G. & Marth, C. Why did p53 gene therapy fail in ovarian cancer? *Lancet Oncol* **4**, 415-22 (2003).
114. Del Papa, J. & Parks, R.J. Adenoviral Vectors Armed with Cell Fusion-Inducing Proteins as Anti-Cancer Agents. *Viruses* **9**(2017).
115. Lukashev, A.N. & Zamyatnin, A.A., Jr. Viral Vectors for Gene Therapy: Current State and Clinical Perspectives. *Biochemistry (Mosc)* **81**, 700-8 (2016).
116. Machitani, M. Development of Novel Genetically Engineered Adenoviruses Based on Functional Analyses of Adenovirus-encoded Small RNAs. *Yakugaku Zasshi* **136**, 1509-1515 (2016).

Appendix A

This appendix represents my data from the following manuscript that has been submitted for review:

"*Talin 1* (*Tln1*) and other modifier loci that influence cancer type in a mouse model of sporadic breast cancer."

Marsha D. Wallace[1,4], Nithya Kartha[1], Pavel Korniliev[2], Teresa L. Southard[3], Jason G. Mezey[2], Lishuang Shen[1,5], John C. Schimenti[1]

Affiliations:

[1] Dept. of Biomedical Sciences, Cornell University, Ithaca, NY

[2] Dept. of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY

[3] Section of Anatomic Pathology, Cornell University, Ithaca, NY

[4] Nuffield Department of Medicine, Ludwig Institute for Cancer Research Ltd, University of Oxford, Oxford, UK.

[5] Center for Personalized Medicine, Children's Hospital Los Angeles, Los Angeles, California

Author Contributions:

M.D.W., N.K. and J.C.S designed experiments; experiments were conducted by M.D.W and N.K.; P.K. and L.S. performed QTL analyses; T.L.S. conducted histopathology; J.G.M. oversaw QTL analyses; J.C.S. oversaw the *Chaos3* project.

Abstract

It is estimated that ~25% of breast cancer cases have a familial basis, but the susceptibility genes underlying heritable predisposition remain largely unknown. Identification of genomic variants contributing to cancer susceptibility is complicated by the breadth of genetic diversity between individuals and populations, differences in the biological subtypes of breast cancers, and differing environmental factors. These issues can be circumvented using mouse cancer models with defined genetic backgrounds in a controlled environment. Here we utilize the *Chaos3* mouse model which, by virtue of a point mutation in the *Mcm4* DNA replication gene that leads to genomic instability, causes spontaneous mammary tumors in the C3HeB/FeJ (C3H) strain background, but other cancer types in the C57BL/6J (B6) strain background. Studies of F1 and F2 (B6xC3H) -$Mcm4^{Chaos3/Chaos3}$ female mice revealed a spectrum of cancer types driven by the $Mcm4^{Chaos3}$ mutation, and genotyping of 189 F2 animals indicated that the genetic architecture of genetic susceptibility and resistance to mammary tumorigenesis and other tumor types is complex. However, one susceptibility locus contained *Tln1*, a gene involved in integrin activation and focal adhesion to the extracellular matrix. Remarkably, we identified a mutation (Glu1910Lys) in the C3H allele of *Tln1* that arose in our stock line of C3H-$Mcm4^{Chaos3}$ mice. Subsequent breeding studies revealed that this *Tln1* mutation, which we hypothesize to be either a hypermorphic or a gain-of-function allele, increased the proportion of C3H-*Chaos3* mice that developed mammary tumors. These results implicate *Tln1* as a breast cancer modifier.

Gene Expression Alterations in *Tln1* Mutant Mammary Cells

To determine if the *Tln1* mutation impacts gene expression in a manner that might explain the heightened mammary tumor susceptibility, I conducted RNA-seq to identify differentially expressed (DE) genes in *Tln1* WT vs Tln1K mammary glands of 3-month-old nulliparous females. This revealed 64 DE genes (FPKM>5; log2 >1) (Table. A1). Six of the most highly DE genes (log2 >3) were selected for validation by qRT-PCR, and five were confirmed to be markedly upregulated (Fig. A1A). The three genes that were most most highly upregulated in mammary glands (*Spp1*, *Cck* and *Ucp1*) were either unchanged or upregulated to a smaller degree (*Spp1*) in tail and gonadal adipose tissue (Fig. A1B, C, respectively).

Of the highly upregulated genes, the two that have plausible roles in tumorigenesis are *Spp1* and *Ucp1*. *SPP1* (osteopontin) has long been known as a marker for epithelial cell transformation, and has been shown in several studies to increase tumorigenicity and metastasis in multiple cancer types, including breast cancer[1,2]. The SPP1 protein contains an integrin recognition motif[3], through which it helps control cell adhesion and motility[4]. UCP1 (thermogenin) is a member of the family of uncoupling mitochondrial proteins that is highly expressed in brown adipose tissue, and has been shown to play a role in "fat-browning" of white adipose tissue in phenomenon of cancer-associated cachexia[5]. A previous study has shown that overexpressing UCP1 in cancer-associated fibroblasts increases the release of ATP-rich vesicles extracellularly, thus stimulating the growth of neighboring epithelial cancer cells in a paracrine manner[6].

Table A-A1. Differentially expressed genes in *Tln1* mutant mammary gland cells.

GENE SYMBOL	log2 (fold_change)	GENE NAME
Ucp1	6.5	uncoupling protein 1 (mitochondrial, proton carrier)
Lalba	4.8	Lactalbumin, alpha
Fabp3	3.9	fatty acid binding protein 3, muscle and heart
Foxi1	3.7	Forkhead Box I1
Cck	3.6	Cholecystokinin
Spp1	3.6	Secreted Phosphoprotein 1
Fam25c	2.9	family with sequence similarity 25, member C
Clic6	2.3	chloride intracellular channel 6
1810010D01Rik	2.3	
S100a8	2.2	S100 calcium binding protein A8 (calgranulin A)
Csn1s2a	2.1	casein alpha s2-like A
S100a9	2.1	S100 calcium binding protein A9 (calgranulin B)
Ltf	1.9	lactotransferrin
Ngp	1.8	neutrophilic granule protein
Btn1a1	1.7	butyrophilin, subfamily 1, member A1
Csn2	1.6	casein beta
Mid1	1.6	midline 1
Anxa8	1.6	annexin A8
Mest	1.4	mesoderm specific transcript
Slc39a8	1.4	solute carrier family 39 (metal ion transporter), member 8
Vnn1	1.3	vanin 1
Scnn1b	1.3	sodium channel, nonvoltage-gated 1 beta
Tspan1	1.3	tetraspanin 1
Pdk4	1.3	pyruvate dehydrogenase kinase, isoenzyme 4
Tmem56	1.1	transmembrane protein 56
Slc12a2	1.1	solute carrier family 12, member 2
Clca2	1.1	chloride channel calcium activated 2
Slc5a5	1.1	solute carrier family 5 (sodium iodide symporter), member 5
Tmprss2	1.1	transmembrane protease, serine 2
Agps	1.1	alkylglycerone phosphate synthase
Hey1	1.1	hairy/enhancer-of-split related with YRPW motif
Col9a1	1.1	collagen, type IX, alpha 1
Rtp4	-1	receptor transporter protein 4
Agtr1a	-1.1	angiotensin II receptor, type 1a
Oas1a	-1.2	2'-5' oligoadenylate synthetase 1A
Thrsp	-1.2	thyroid hormone responsive
Arc	-1.2	activity regulated cytoskeletal-associated protein
Bst2	-1.2	bone marrow stromal cell antigen 2
Egr2	-1.2	early growth response 2
Crabp1	-1.3	cellular retinoic acid binding protein I
Cmpk2	-1.3	cytidine monophosphate (UMP-CMP) kinase 2, mitochondria
Usp18	-1.3	ubiquitin specific peptidase 18
Lrrc15	-1.4	leucine rich repeat containing 15
Dhx58	-1.4	DEXH (Asp-Glu-X-His) box polypeptide 58
Mx2	-1.4	MX dynamin-like GTPase 2
Igfbp2	-1.5	insulin-like growth factor binding protein 2
Zbp1	-1.6	Z-DNA binding protein 1
Oas2	-1.6	2'-5' oligoadenylate synthetase 2
Gzmb	-1.6	granzyme B
Calca	-1.6	calcitonin/calcitonin-related polypeptide, alpha
Isg20	-1.6	interferon-stimulated protein
Oasl2	-1.7	2'-5' oligoadenylate synthetase-like 2
Mx1	-1.7	MX dynamin-like GTPase 1
Ifit1	-1.7	interferon-induced protein with tetratricopeptide repeats 1
Defb1	-1.7	defensin beta 1
Irf7	-2.1	interferon regulatory factor 7
Ifit3	-2.1	interferon-induced protein with tetratricopeptide repeats 3
I830012O16Rik	-2.1	
Wfdc12	-2.3	WAP four-disulfide core domain 12
Oasl1	-2.3	2'-5' oligoadenylate synthetase-like 1
Oas3	-2.3	2'-5' oligoadenylate synthetase 3
Isg15	-2.4	Ubiquitin-Like Modifier
Sct	-2.8	secretin
Pnmt	-2.9	Phenylethanolamine N-Methyltransferase

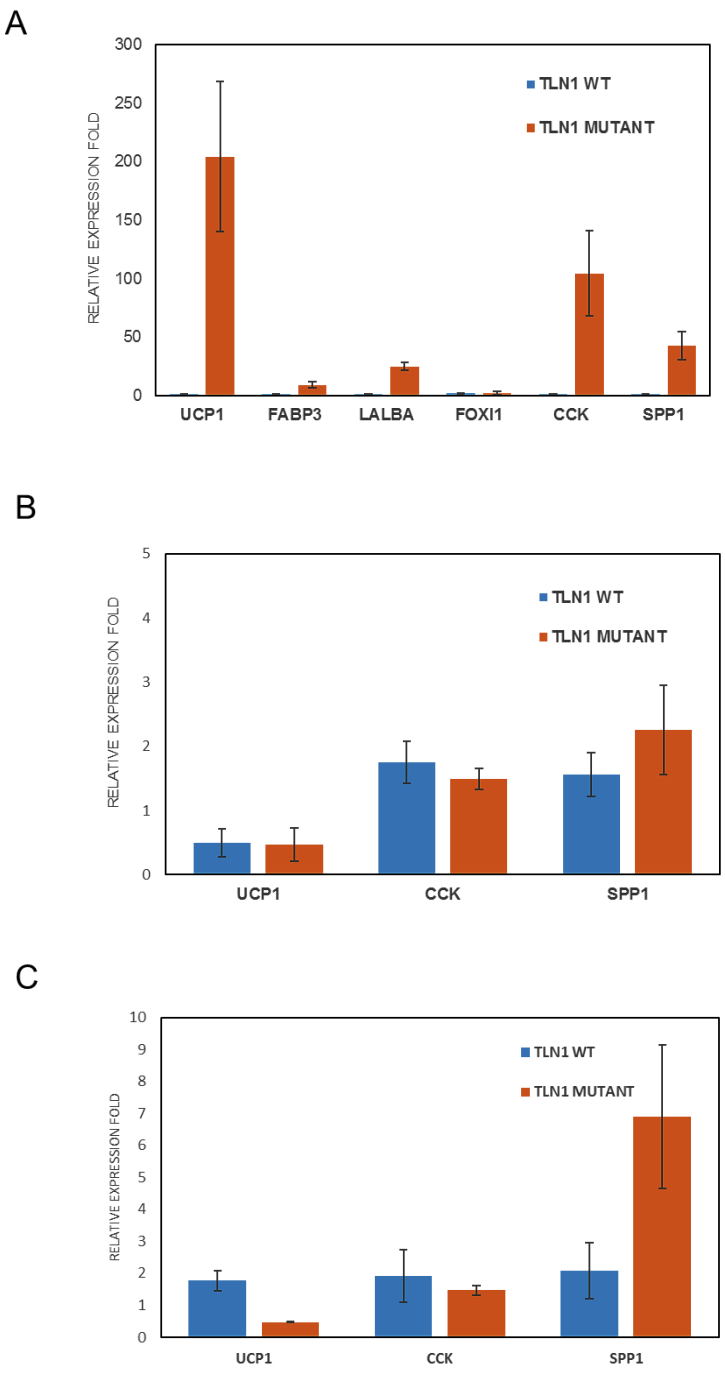


Figure A-A1. Top DE genes with roles in tumorigenesis. (A) qRT-PCR validation of the most highly upregulated genes in Tln1 mutant mammary cells. qRT-PCR to assay expression of top three upregulated genes in (B) tail and (C) gonadal adipose tissue controls.

References

1. Rittling, S.R. & Chambers, A.F. Role of osteopontin in tumour progression. *Br J Cancer* **90**, 1877-81 (2004).
2. Rodrigues, L.R., Teixeira, J.A., Schmitt, F.L., Paulsson, M. & Lindmark-Mansson, H. The role of osteopontin in tumor progression and metastasis in breast cancer. *Cancer Epidemiol Biomarkers Prev* **16**, 1087-97 (2007).
3. Bayless, K.J., Meininger, G.A., Scholtz, J.M. & Davis, G.E. Osteopontin is a ligand for the alpha4beta1 integrin. *J Cell Sci* **111 (Pt 9)**, 1165-74 (1998).
4. Kim, J. *et al.* Secreted phosphoprotein 1 binds integrins to initiate multiple cell signaling pathways, including FRAP1/mTOR, to support attachment and force-generated migration of trophectoderm cells. *Matrix Biol* **29**, 369-82 (2010).
5. Petruzzelli, M. *et al.* A switch from white to brown fat increases energy expenditure in cancer-associated cachexia. *Cell Metab* **20**, 433-47 (2014).
6. Sanchez-Alvarez, R. *et al.* Mitochondrial dysfunction in breast cancer cells prevents tumor growth: understanding chemoprevention with metformin. *Cell Cycle* **12**, 172-82 (2013).

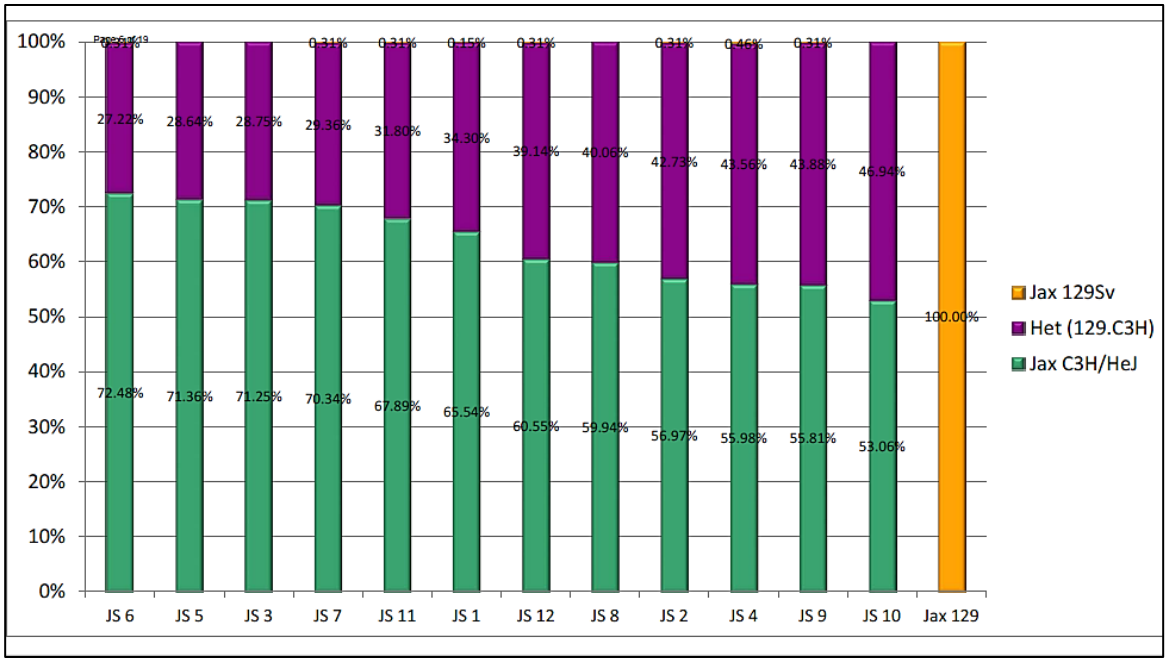
Appendix B

This appendix represents my work in generating an *Arid1a* conditional knock out (CKO) mouse model bearing the *Chaos3* mutation, with the purpose of determining whether loss of *Arid1a* is a driver of mammary carcinogenesis in this model.

To assess whether conditional deletions (homozygous and/or heterozygous) of *Arid1a* in the mammary gland is sufficient to drive tumor formation at a faster rate in the *Chaos3* background, we utilized a mouse model in which exon eight of *Arid1a* is floxed by two LoxP sites, which has been previously verified to result in a loss-of-function allele[1]. The floxed *Arid1a* allele was made congenic in the C3Heb/FeJ mouse strain bearing the Mcm4^{Chaos3} allele through two rounds of speed congenics, followed by selection of the best male breeders for subsequent matings (Figures A-B1). I then bred nulliparous female mice of the relevant genotypes, either with or without the *Talin1* point mutation, setting up two independent cancer screens (Figure A-B2). These mice also contained a MMTV-Cre transgene necessary for catalyzing the *Arid1a* deletions, that I bred into the original C3H-*Chaos3* strain background, using a model that had been previously described as a "less efficient" Cre[2] to avoid any potential problems arising from "leaky" Cre expression.

Following the development of mammary tumors in some of these mice, we assayed samples of the tumor tissues to confirm the deletion of *Arid1a*. Unfortunately, we found that none of the tumors had the expected *Arid1a* deletions, and attributed this to the unsuccessful expression of the MMTV-Cre allele. We are currently in the process of repeating this experiment using a pre-packaged Cre recombinant adeno virus, which we will directly inject into the mammary glands for immediate deletion of the targeted floxed *Arid1a* allele. We have tested this Cre in mammary tumor cells generated from the unsuccessfully edited tumors mentioned above, bearing intact LoxP sites. These preliminary experiments have shown promising results indicating a more powerful Cre delivery system.

(A)



(B)

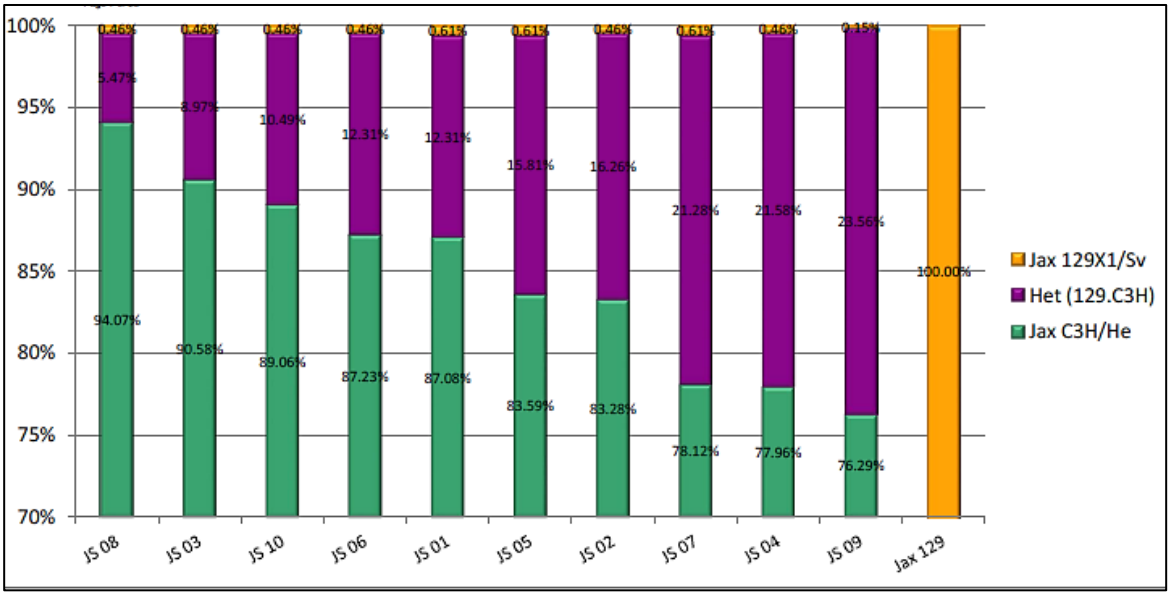


Figure A-B1: Genotyping results from two rounds of speed congenics. Speed congenics results representing (A) first round and (B) second round of enrichment for the C3Heb/FeJ genotype bearing the "floxed" *Arid1a* allele. Male breeders were selected for subsequent matings as follows – Round #1: JS6, JS 5 and JS 3; Round #2: JS 08, JS 03 and JS 10.

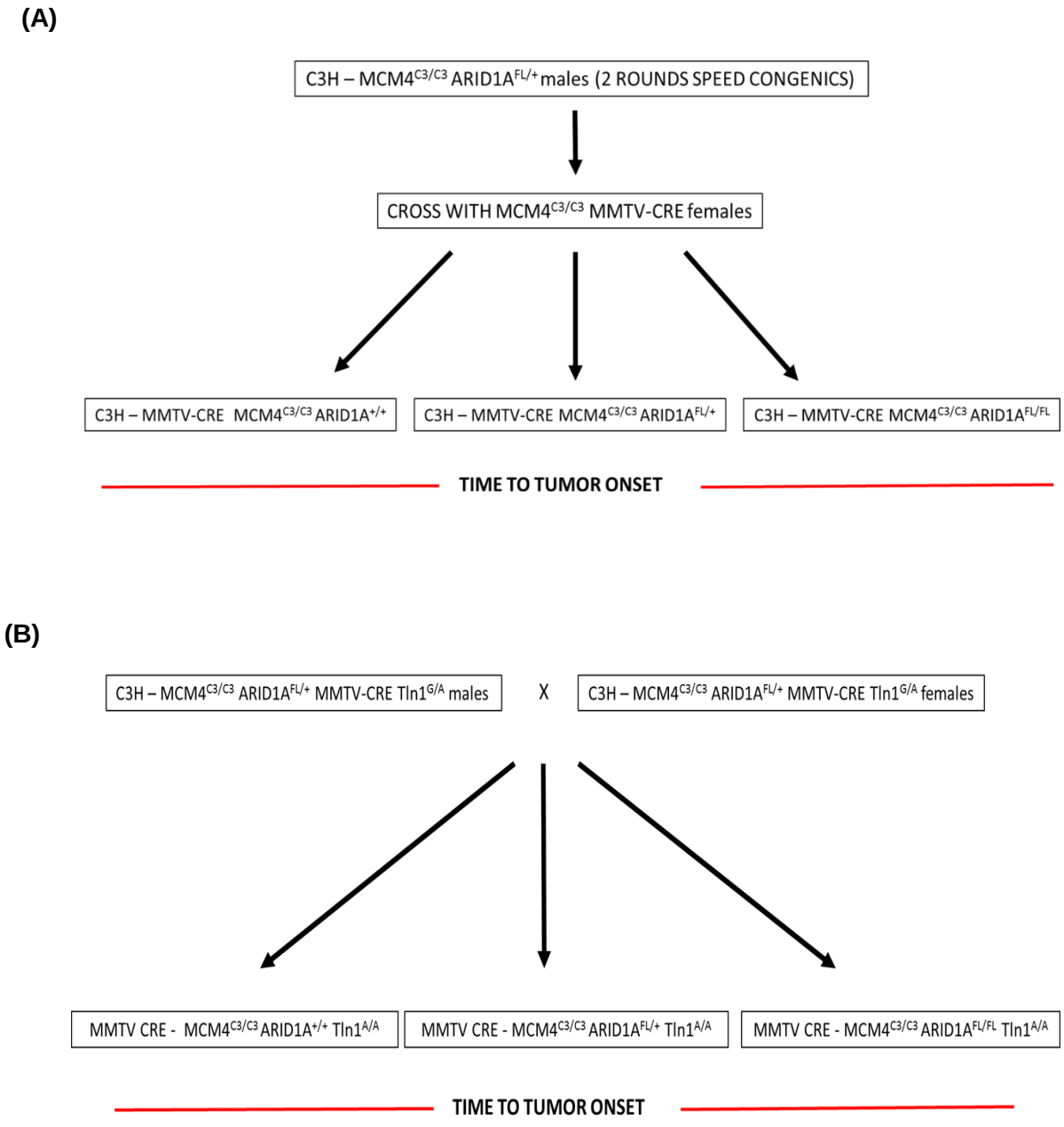


Figure A-B2: Schematic matings for *Arid1a* CKO cancer screens. Two independent cancer screens were set up with mice that were (A) wild-type for *Talin1* mutation or (B) *Talin1* mutant. All mice contained the MMTV-Cre allele, with either wild-type, heterozygous or homozygous alleles of floxed *Arid1a*. N=15 for each genotype in both screens.

References:

1. Gao, X. *et al.* ES cell pluripotency and germ-layer formation require the SWI/SNF chromatin remodeling component BAF250a. *Proc Natl Acad Sci U S A* **105**, 6656-61 (2008).
2. Cheng, L. *et al.* Rb inactivation accelerates neoplastic growth and substitutes for recurrent amplification of cIAP1, cIAP2 and Yap1 in sporadic mammary carcinoma associated with p53 deficiency. *Oncogene* **29**, 5700-11 (2010).