Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Sara Flano April 7, 2023

Defining the cis-regulatory sequence elements regulated by BAF complexes containing different ARID1 variants

by

Sara Flano

Dr David U. Gorkin

Adviser

Biology

Dr David U. Gorkin

Adviser

Dr Arri Eisen

Committee Member

Dr Andreas Fritz

Committee Member

Dr Adam Gracz

Committee Member

2023

Defining the cis-regulatory sequence elements regulated by BAF complexes containing different ARID1 variants

By

Sara Flano

Dr David U. Gorkin

Adviser

An abstract of

a thesis submitted to the Faculty of Emory College of Arts and Sciences

of Emory University in partial fulfillment

of the requirements of the degree of

Bachelor of Science with Honors

Biology

2023

Abstract

Defining the cis-regulatory sequence elements regulated by BAF complexes containing different ARID1 variants

By Sara Flano

BRG1/BRM-associated factor (BAF) complexes regulate gene expression by repositioning the histone proteins that package DNA inside the cell's nucleus. This process, known as chromatin remodeling, regulates gene expression by controlling DNA accessibility to regulatory proteins known as transcription factors (TFs). TFs bind to specific DNA sequences, which are often grouped into larger cis-regulatory sequence elements (cREs) where the binding of multiple TFs influences the transcription of one or more target genes. One subunit of BAF complexes is an AT-rich interaction domain 1 (ARID1) protein. Mammalian genomes encode for two different ARID1 proteins, known as ARID1A and ARID1B. Mutations in genes that encode BAF complex subunits can cause a variety of diseases. Interestingly, mutations in both ARID1A and ARID1B cause disease in humans, but the diseases that are associated with each variant differ: ARID1A is the most frequently mutated BAF subunit in cancer, whereas ARID1B is the BAF subunit most frequently mutated in neurodevelopmental disorders. While the importance of BAF complexes and their ARID1 subunits in human health has been well established, the specific cREs and target genes they regulate remain undefined. Based on the difference in disease involvement between ARID1A and ARID1B, I hypothesize that BAF complexes containing different ARID1 subunits regulate chromatin accessibility at different sets of cREs, which in turn regulate the expression of different target genes. To identify the

regulatory targets of BAF complexes containing ARID1A/B, I assessed changes in chromatin accessibility in Neuro-2A cells treated with an ATPase inhibitor or *Arid1a*-targeting DsiRNAs. My results produce novel and foundational data that pave the way for further characterization of the cREs and target genes that may be differentially regulated by ARID1A- and ARID1B-containing BAF complexes. These investigations will likely contribute to a better understanding of associated neurodevelopmental disorders, and to exploring potential therapeutic targets for associated cancers.

Defining the cis-regulatory sequence elements regulated by BAF complexes containing different ARID1 variants

By

Sara Flano

Dr David U. Gorkin

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences

of Emory University in partial fulfillment

of the requirements of the degree of

Bachelor of Science with Honors

Biology

2023

Acknowledgements

First off, thank you to Dr Gorkin for being an incredible mentor throughout my undergraduate research career. Thank you for always believing in me and encouraging me throughout this process. I am so grateful for all the experiences I have had in your lab, where I have learned so much, and grown as both a person and a scientist. Thank you to my committee members, Dr Arri Eisen, Dr Andreas Fritz, and Dr Adam Gracz for their time and insight. A special thank you to Dr Eisen and Dr Fritz, whose classes inspired me to get involved in research, and sparked a curiosity in development and epigenetics that has profoundly shaped the way I see the world.

Thank you to all the members of the Gorkin Lab. In particular, thank you to Susan for your constant support, advice, and warmth. You have taught me so much, from the joy of cell culture to the magic of having a lab notebook/scrapbook. Thank you to Alex for always taking the time to answer my many questions, teach me something new, or help me figure something out. Thank you, especially, to Sona for being my partner in the lab. I couldn't have gotten through the ups and downs of research without you!

Thank you to my family, especially my Dad, for always supporting me, encouraging me, and keeping me grounded. Thank you to my cat, Mia, for sitting with me and comforting me during the many hours I spent writing this thesis. Lastly, thank you to Chris for being my rock throughout this process: celebrating with me when I succeeded, and comforting me then pushing me back onto my feet when things didn't work. I truly could not have accomplished this without your never-ending support and encouragement.

Table of Contents

List of Figures

Chapter 1: Introduction

1.1: Gene Expression Regulation through Chromatin Remodeling.

The organization of chromatin plays an essential role in regulating gene expression via epigenetic mechanisms that alter transcriptional activity without altering the DNA sequence itself. Organization of the genome begins at its most basic level with the nucleosome, which contains approximately 147 base pairs of DNA wrapped around a histone octamer. Modifications of these histone proteins, as well as their addition, removal, or repositioning create reversible chromatin states that influence gene transcriptional activity and provide a key epigenomic mechanism for regulating gene expression (Allis and Jenuwein 2016). Chromatin accessibility is regulated by sliding aside or eviction of histones to make the underlying DNA accessible to DNA-binding proteins such as Transcription Factors (TFs). TFs are regulatory proteins that bind to specific DNA sequences and recruit cofactors to help initiate or enhance transcription. These DNA sequences are often grouped into cis-regulatory sequence elements (cREs), which are regions of the genome whose accessibility to the binding of multiple TFs ultimately regulates the transcription of one or more target genes. Well-known examples of cREs are promoters, located next to their target gene's transcription start site, and enhancers, which can be found further away from the genes they regulate (Allis and Jenuwein 2016, Gasperini, Tome et al. 2020).

The property of chromatin accessibility is characteristic of cREs and is often used as a proxy to measure the activity level of a given cRE in a particular cell type or context (Allis and Jenuwein 2016, Gasperini, Tome et al. 2020). Chromatin remodelers are protein complexes that establish and maintain chromatin accessibility by repositioning nucleosomes. This work is essential to developmental processes, maintaining cell fates, and allowing for changes in gene expression in response to extracellular and intracellular signals. When chromatin accessibility is

incorrectly regulated, diseases and disorders such as cancer and neurodevelopmental disorders can occur (Allis and Jenuwein 2016, Ellegood, Petkova et al. 2021).

1.2: BAF Complexes in Gene Regulation and Disease

BRG1/BRM-associated factor (BAF) complexes, also known as SWItch/Sucrose Non-Fermentable (SWI/SNF) complexes, are a widely-expressed family of chromatin remodelers (Figure 1A). BAF complexes are known to play key roles in cell differentiation, neural development, and regulation of gene expression within mammalian cells (Kadoch and Crabtree 2015). These complexes are thought to reposition nucleosomes by first forming a clamp around the nucleosome, and then leveraging the SMARCA4/2 ATPase subunit to undergo ATP hydrolysis. This creates a conformational change in the complex, which in turn causes the nucleosome to shift its position on the DNA (Figure 1B) (Ronan, Wu et al. 2013, Mashtalir, D'Avino et al. 2018, Schick, Rendeiro et al. 2019). BAF complexes are recruited to regions of the genome by a combination of protein-protein interactions with TFs, and direct binding to histone modifications and other regional chromatin architectural features via “reader” domains in BAF subunit proteins (Kadoch and Crabtree 2015). Each BAF complex is constructed from 12 to 15 proteins, which come from a pool of approximately 29 potential subunits, some of which have one or more variants (e.g. ARID1A and ARID1B) (Raab, Resnick et al. 2015). This produces a combinatorically diverse family of complexes that differ in their composition, function, and prevalence within different cell types and developmental stages (Moffat, Jung et al. 2021, Bogershausen and Wollnik 2018).

Proper function of the BAF complexes is essential to successful gene regulation, and mutation of its subunits can have profound effects. In fact, over 20% of human cancers have

somatic mutations in at least one BAF subunit (Moffat, Smith et al. 2022, Schick, Rendeiro et al. 2019). In addition, mutations in BAF subunits are known to cause a series of syndromic neurodevelopmental phenotypes called “SWI/SNF-related intellectual disability disorders (SSRIDDs)” (Bogershausen and Wollnik 2018). Mutations in some of these same subunits are further implicated in non-syndromic intellectual disability (ID) and autism spectrum disorders (ASD) (Mashtalir, D'Avino et al. 2018). Notably, in many cases the BAF subunit mutations that cause SSRIDDs lead to loss of function and are dominant, suggesting that cells in development are sensitive to the dosage of functional BAF complexes (Bogershausen and Wollnik 2018). This makes research into the BAF complex and all its subunits incredibly important to both understanding human genetics and neural development, as well as clinical implications in oncology and neurological disease.

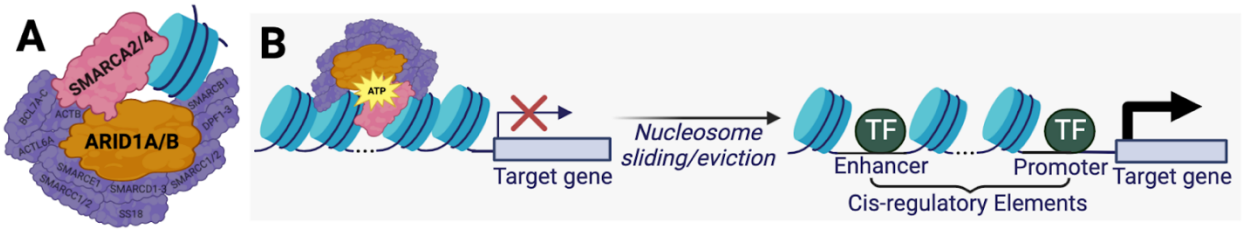

Figure 1: A) Diagram of the canonical BAF complex bound to a nucleosome. B) Model of the BAF complex’s role as a chromatin remodeler.

1.3: ARID Subunits of the cBAF Complex

The BAF complex family is commonly divided into three groups based on their subunit composition: canonical BAF (cBAF), polybromodomain BAF (pBAF), and non-canonical BAF (ncBAF). Canonical BAF (cBAF) complexes are characterized by the inclusion of an AT-rich interaction domain 1 (ARID1) protein, also called BAF250 (Figure 1A) (Mashtalir, D'Avino et al. 2018). Mammalian genomes contain two genes that code for ARID1 protein variants known as ARID1A and ARID1B. These two ARID1 variants are mutually exclusive in that a single

BAF complex includes one or the other, but never both. The proteins share approximately 60% sequence identity, with conserved ARID domain and C-terminus. The ARID domain is a DNA binding domain, and the C-terminus has three BAF core subunit and ATPase binding domains (Moffat, Smith et al. 2022). The exact roles of ARID1 proteins within BAF are still being determined, although they are considered core components of cBAF (Bogershausen and Wollnik 2018) as they are thought to form the bridge that links one prong of the clamp to the other and play a role in BAF complex recruitment and binding to DNA (Figure 1A) (Mashtalir, D'Avino et al. 2018, Schick, Rendeiro et al. 2019). Mutations in the ARID subunits are generally heterozygous truncating, loss-of-function, or intragenic/whole gene deletions, with few exceptions of missense mutations in *ARID1B* (Bogershausen and Wollnik 2018). These mutations can occur anywhere in the gene/protein, lead to highly heterogenous phenotypes, and disrupt complex assembly and function without necessarily disrupting localization (Celen, Chuang et al. 2017, Moffat, Jung et al. 2021). The ARID subunits are among the most frequently mutated BAF subunits in human disease (Mashtalir, D'Avino et al. 2018), making understanding their function within the BAF complex and gene regulation essential.

Most cells express both *ARID1A* and *ARID1B*, and many regions of the genome bound by BAF complexes are bound by both ARID1A- and ARID1B-containing complexes. However, previous research has found that the binding locations of BAF complexes containing these different ARID1 proteins do not completely overlap. Some regions of the genome preferentially recruit ARID1A- or ARID1B-containing BAF complexes, and switching from ARID1A to ARID1B complexes or vice versa during different stages of development is necessary for proper proliferation and differentiation (Jung, Moffat et al. 2017, Pagliaroli, Porazzi et al. 2021). ARID1A/B also exhibit partial functional redundancy as shown by partial compensation and

increased incorporation of the paralog when one is knocked out (Schick, Rendeiro et al. 2019, Jimenez, Antonelli et al. 2022). However, in some contexts, ARID1B exhibits more repressive effects on gene expression while ARID1A activates (Raab, Resnick et al. 2015, Pagliaroli, Porazzi et al. 2021). These findings suggest that the function of BAF complexes containing ARID1A is partially distinct from the function of BAF complexes containing ARID1B. This divergence in function between ARID1A and ARID1B is further supported by the subunits differing roles in human disease. Mutations in *ARID1A* and *ARID1B* are known to cause or contribute to several classes of human disease. However, the diseases where mutations in each variant are most commonly found differ (Mashtalir, D'Avino et al. 2018). *ARID1A* is the BAF subunit that is most frequently mutated in cancer, including colon, rectal, pancreatic, gastric, and ovarian cancer (Kadoch and Crabtree 2015, Mashtalir, D'Avino et al. 2018). On the other hand, *ARID1B* is much less frequently mutated in cancer. Rather, *ARID1B* is the most frequently mutated BAF subunit in neurodevelopmental disorders, and its mutations are the most common cause of SSRIDDs (Kadoch and Crabtree 2015). *ARID1B* is among the most common single gene mutations found in non-syndromic intellectual disability and Coffin-Siris syndrome, as well as one of the most commonly mutated genes in autism spectrum disorder (Celen, Chuang et al. 2017, Pagliaroli, Porazzi et al. 2021, Bogershausen and Wollnik 2018).

Despite the essential role of BAF complexes and their ARID1 subunits in the regulation of gene expression and maintenance of human health, we do not know what specific cREs they regulate. I hypothesize BAF complexes containing different ARID1 subunits regulate different sets of cREs and target genes. I anticipate that their identification will lead to a better understanding of how mutations in *ARID1A/B* contribute to different diseased states.

In this work, I set out to characterize the cREs and target genes that may be differentially

regulated by ARID1A- and ARID1B-containing BAF complexes by first defining the chromatin regions that depend on all BAF complexes for accessibility and then identifying the potential subsets of regions that rely specifically on ARID1A- and/or ARID1B-containing BAF complexes.

Chapter 2: Methods

2.1: Cell Culture

The mouse neuroblast Neuro-2A (N2A) cell line was used in all experiments. These adherent cells are fit-for-purpose as they are undifferentiated precursors to tissues of the central nervous system, and allow the BAF complex, and ARID1 subunits, to be studied within both developmental and neuron characteristics that grow well in laboratory conditions.

N2A cells were cultured in 20mL of complete media in T75 flasks, seeding flasks with 2 million cells and passaging cells at approximately 80% confluency (Figure 1). All cell collections were performed using trypsinization, treating cells with 25% trypsin EDTA for approximately one minute.

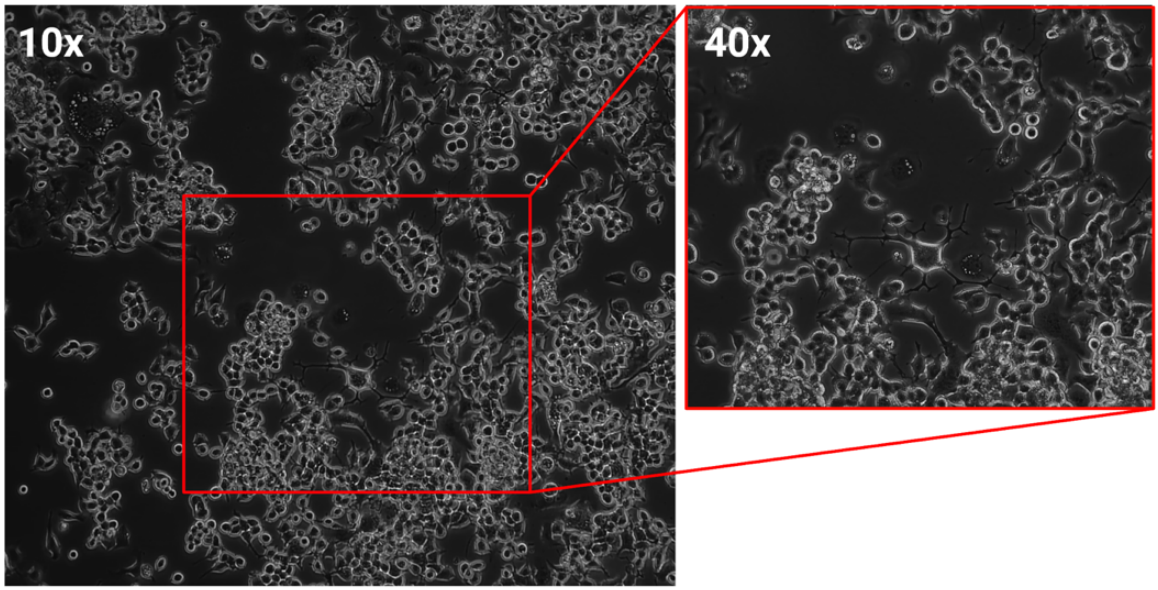

Figure 1: PX+7 N2A cells at approximately 80% confluency, before passage.

2.2: ATPase Inhibitor Treatment

150,000 N2A cells (approximately 15% confluency) were plated per well in 12-well plates in 1mL of complete media. Dilutions of 1mM, 100uM, and 10uM BRM014 were prepared

from 10mM BRM014 stock using DMSO. Treatment wells were treated with 10uL of the corresponding BRM014 dilution 24 hours after seeding. This created final concentrations and treatments of BRM014 at 10uM, 1uM, or 0.1uM. The control wells were not-treated cells and 10uL DMSO treated cells. All wells were collected 24 hours after treatment by trypsinization, snap-frozen in liquid nitrogen, and stored at -80ºC.

2.3: DsiRNA Transfection

2.3A: DsiRNA Selection

All DsiRNAs were ordered from Integrated DNA Technologies' (IDT) TriFECTa RNAi Kit. The control DsiRNAs, a non-targeting DsiRNA and HPRT1 siRNA, were validated by IDT (TriFECTa Kit, IDT). Using their proprietary algorithm, IDT designed three possible DsiRNAs each for *Arid1a* and *Arid1b* (Table 1).

DsiRNA	Sequence
Arid1a #1	CTCTTAGTCTGACACTCTATTTGGTAC
Arid1a #2	TCGTAACCATTGAACTATTCGAAGGAT
Arid1a #3	TTTCTAATGGAGTTTTCTATAACAATG
Arid1b #1	GAATCAGACTGTGAGATAAACCATG
Arid1b #2	CATTGGTAACTTGATAAGCTTCCTA
Arid1b #3	AGATTACCTCAAAAGATATTGTTAC

Table 1: Sequence for all *Arid1a*- and *Arid*1b-targeting siRNAs.

2.3B: DsiRNA Treatment

200,000 N2A cells were seeded into each well of 6-well plates in 2.5uL of transfection media. 24 hours after seeding, cells were treated with RNAi duplexes. The duplexes were prepared by diluting 20uM DsiRNA stocks (IDT) and the Lipofectamine RNAiMAX (ThermoFisher Scientific, 13778075) in Opti-MEM™ I Reduced-Serum Medium (ThermoFisher Scientific, 31985070) and allowed to incubate for 10 minutes. Cells were treated with the

DsiRNAs for 30 hours. The concentration of all DsiRNAs was 10nM for all experiments, and 7.5uL of LipoFECTamine was used per well. Cells were then harvested through trypsinization, and then used fresh for RNA isolation or snap-frozen in liquid nitrogen for ATAC-seq.

2.3C: RNA Isolation and cDNA Synthesis

All RNA isolation was performed on fresh cells using the RNeasy MinElute Kits (Qiagen, 74204) and on-column DNase digestion and manufacturer's instructions. Quality of RNA was assessed using RNA HS ScreenTapes (Agilent, 5067-5579) and stored at -20ºC. cDNA was created using SuperScript IV VILO Master Mix with ezDNAse Enzyme (ThermoFisher Scientific, 11766050) following manufacturer's protocol and stored at -20ºC.

2.3D: TaqMan qPCR

To validate gene knockdown TaqMan qPCR was performed on all samples using TaqMan™ Fast Advanced Master Mix (ThermoFisher Scientific, 4444556) and TaqMan Gene Expression Assays (ThermoFisher Scientific) following manufacturer's protocols for 96-well Fast Plates (ThermoFisher Scientific). TaqMan primer assays were selected and used for all qPCR transfection validation protocols due to their sensitivity and high signal-to-noise ratio. The TaqMan assays were selected from the recommended assays for each gene by the ThermoFisher TaqMan Assay Search Wizard. *Gapdh* expression was used as a qPCR control and knockdown reference gene for all experiments. Knockdown was calculated by subtracting Ct of experimental gene assay by the Ct of the reference gene (*Gapdh*) for each technical replicate of all samples to find the ΔCt value. The ΔCt value for each technical replicate was averaged. The average ΔCt value of the non-targeting sample was subtracted from each targeting siRNA sample's ΔCt value to find the ΔΔCt value. To normalize for gene expression, $2^{-\Delta\Delta Ct}$ was calculated and multiplied by 100 to find percent gene expression for the knocked-down gene. To find percent knockdown

of the experimental gene, percent gene expression for each sample was subtracted by 100. For samples treated with *Arid1a*- or *Arid1b*-targeting siRNAs, knockdown for both *Arid1a* and *Arid1b* was calculated to determine off-target effects of all siRNAs.

2.4: ATAC-seq Library Preparation and Sequencing Analysis

Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) is a commonly used technique used to identify areas of accessible chromatin, using hyperactive Tn5 transposase to cut out open areas of the genome for sequencing and identification. ATAC-seq was performed on snap-frozen cells, size selection was performed with SPRIselect beads (B23318, Beckman Coulter), and all index primers are xGen™ CDI Primers (IDT, 10009794). Cells were thawed at 37°C for 2 minutes in a water bath, resuspended in 250µL cold PBS, and then resuspend in 250µL of nuclei permeabilization buffer and rotated for 5 min at 4°C. The nuclei were then resuspended in cold 25µL tagmentation buffer and placed on ice. Nuclei concentration was calculated using a hemocytometer, and then adjusted to to 2,500 nuclei/µL. 20µL of this diluted solution was separated for tagmentation. 1µL of Tn5 enzyme (J. D. Buenrostro et al. 2013) was added to each nuclei sample, and samples were incubated for 60 minutes at 500 rpm at 37°C. After tagmentation, DNA was purified using the MinElute PCR purification kit (28004, Qiagen) following manufacturer's instructions and eluted in 10µL EB. DNA fragments were then amplified by PCR, using 10uL of Tagmented DNA, 25uL of NEBNext 2x PCR MasterMix, 2uL of 25uM i5-primer, 2uL pf 25uM i7-primer, and 11uL of molecular biology water per reaction. The temperature profile used was 72°C incubation for 5 minutes, then 98°C for 30 seconds, followed by eight cycles of 98°C for 10 seconds, 63°C for 30 seconds, and 72°C for 1 minute. PCR reactions were then purified using the MinElute PCR

Purification Kit (Qiagen, 28004) and eluted in 20 µl EB. Size selection was then performed, removing larger fragments with a (0.55x sample volume concentration of SPRIbeads, and then removing smaller fragments using a 1.5 x sample volume of beads. Final ATAC-seq libraries were quantified using Qubit dsDNA HS Assay Kit (ThermoFisher Scientific, Q32851) and checked for library size distribution using 4200 TapeStation and High Sensitivity D1000 ScreenTape and reagents (5067-5583, Agilent Technologies). ATAC-seq library quality was validated using the iSeq 100 Sequencing System, then libraries were sequenced by Novogene and sequence data was processed using the ATAC-seq pipeline for ENCODE data (Kundaje 2017).

2.5: Buffers and Solutions

Complete Cell Culture Media

450mL EMEM

50mL Fetal Bovine Serum

5mL Pen-Strep

Transfection Cell Culture Media

450mL EMEM

50mL Fetal Bovine Serum

Nuclei Permeabilization Buffer

50mg BSA

20uL 10% m/v IGEPAL Ca-630

5uL 200nM DTT

40uL 25x complete EDTA-free protease inhibitor

935uL 1X PBS

Tagmentation Buffer

33uL 1M Tris acetate

22uL 3M K-acetate

36.7uL Mg-acetate

160uL DMF

748.3uL Molecular Biology Grade Water

Chapter 3: Results

To test my hypothesis that BAF complexes containing different ARID1 subunits regulate different sets of cREs and target genes, I first set out to define the chromatin regions that depend on BAF complexes for chromatin accessibility. To accomplish this, I treated cells with an allosteric inhibitor of BAF activity, and measured changes in chromatin accessibility across the genome (Section 3.2). I then set out to identify the potential subsets of chromatin regions that rely specifically on ARID1A- and/or ARID1B-containing BAF complexes for their chromatin accessibility. I used siRNAs to knockdown the expression of *Arid1a* or *Arid1b* and measured the effects on chromatin accessibility (Section 3.3). All experiments were performed in the mouse neuroblast Neuro-2A (N2A) cell line. These cells are fit-for-purpose as they enable studying BAF complexes, and the ARID1 subunits, within both developmental and neuron characteristics.

3.1: Identifying the optimal seeding density of N2A cells for drug treatment and siRNA knockdown experiments.

No previous data on seeding N2A cells for a 48 to 72-hour experiment in 6-well plates was available so seeding density was optimized for the transfection protocol. 6-well plates were seeded with $0.5x10^6$, $1\ x10^6$, $2\ x10^6$, $3\ x10^6$, $4\ x10^6$, and $5\ x10^6$ cells. Plates were collected and cells were counted 24, 48, and 72 hours after seeding to assess growth rate and health of N2A cells. It was determined that a seeding density of $2\ x10^6$ cells was best for a greater than 48-hour experiment as the final concentration of cells after 72 hours was $1.25\ x10^6$ cells at about 80% confluency. These cells looked the healthiest, were spread evenly across the plates with normal clumping, and were at the correct confluency for harvest (Figure 4). All following transfection

experiments began with cells seeded at a density of 2 x10^6 cells per well in a 6-well plate and treated 24 hours after seeding.

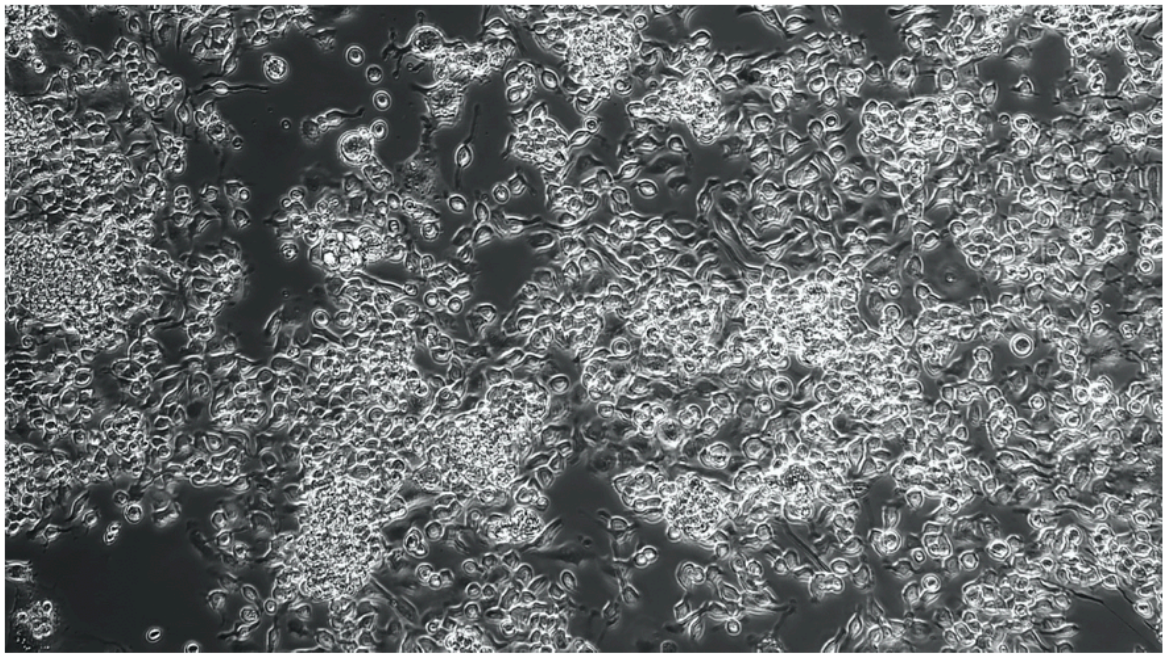

Figure 2: Well seeded with 2x10^6 N2A cells after 72 hours. Approximately 1.25x10^6 cells total in well. Cells are at maximal confluency while remaining healthy with normal clumping and low levels of cell death.

3.2: Characterizing regions of the genome where BAF is required to maintain chromatin accessibility.

To identify the full set of regions of the N2A genome that require BAF to maintain chromatin accessibility, I treated N2A cells with BRM014, a well characterized allosteric ATPase inhibitor of the BAF ATPase subunits SMARCA2 and SMARCA4 (Iurlaro et al. 2021). If ARID1A and ARID1B-containing BAF complexes regulate different cREs and target genes, they would each contribute to chromatin accessibility at a subset of these BAF-dependent loci. Cells were treated with 10uM or 1.0uM of BRM014. As controls, cells were either untreated, or treated with only the drug delivery vehicle: DMSO. I then performed Assay for Transposase Accessible Chromatin (ATAC-seq) on all samples to map regions of accessible chromatin under

each condition, and libraries were sequenced to identify changes in chromatin accessibility when BAF complex activity was inhibited.

After aligning the ATAC-seq sequencing reads to the reference genome (mm10, ENCODE), three of four samples were determined to be of good quality as evidenced by strong enrichment for accessible chromatin at annotated transcription start sites (Figure 3, Table 2), as well as high percentage of unique sequencing reads, and a low fraction of mitochondrial reads (Table 2). The ATAC-seq library of 10uM BRM014 treated cells show low TSS enrichment (Figure 3, Table 2), which could be due to toxicity of the high dose of BRM014.

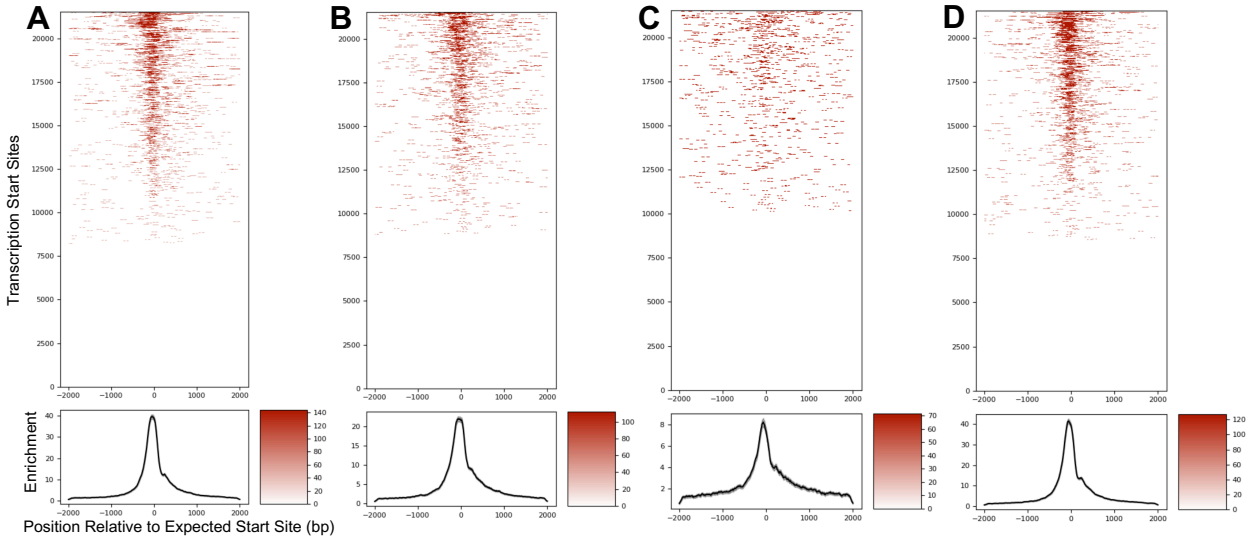

Figure 3: Transcription start site (TSS) enrichment for A) Untreated sample, B) DMSO Treated sample, C) 10uM BRM014 sample, D) 1.0uM BRM014 sample.

Sample Treatment	Percent of Sequence Count		TSS Enrichment	Fraction of Mitochondrial Reads
	Unique Reads	Duplicate Reads		
Untreated	89.0	11.0	39.6048	0.1349
DMSO	85.9	14.1	22.0379	0.2129
10uM BRM014	91.4	8.6	8.21112	0.1238
1.0uM BRM014	80.0	20.0	41.4547	0.2054

Table 2: Percent of sequence counts of unique and duplicate reads, transcription start site (TSS) enrichment, and fraction of mitochondrial reads for all samples. All samples show high percentage of unique reads, and low percentage of duplicate reads. All samples except the 10uM BRM014 treated sample show high TSS enrichment. The low TSS enrichment for the 10uM sample is likely due to the drug treatment's disruption of BAF epigenetic regulation and toxicity. All samples show low fraction of mitochondrial reads, showing quality of ATAC-seq libraries.

The ATAC-seq data show changes in genome accessibility in response to the ATPase inhibitor drug treatment at both the 1uM and 10uM concentrations, with the 10uM sample showing the most dramatic changes (Figure 4). Changes in accessibility are at both intragenic and intergenic regions of the genome (Figure 4), which is in line with cRE characteristics.

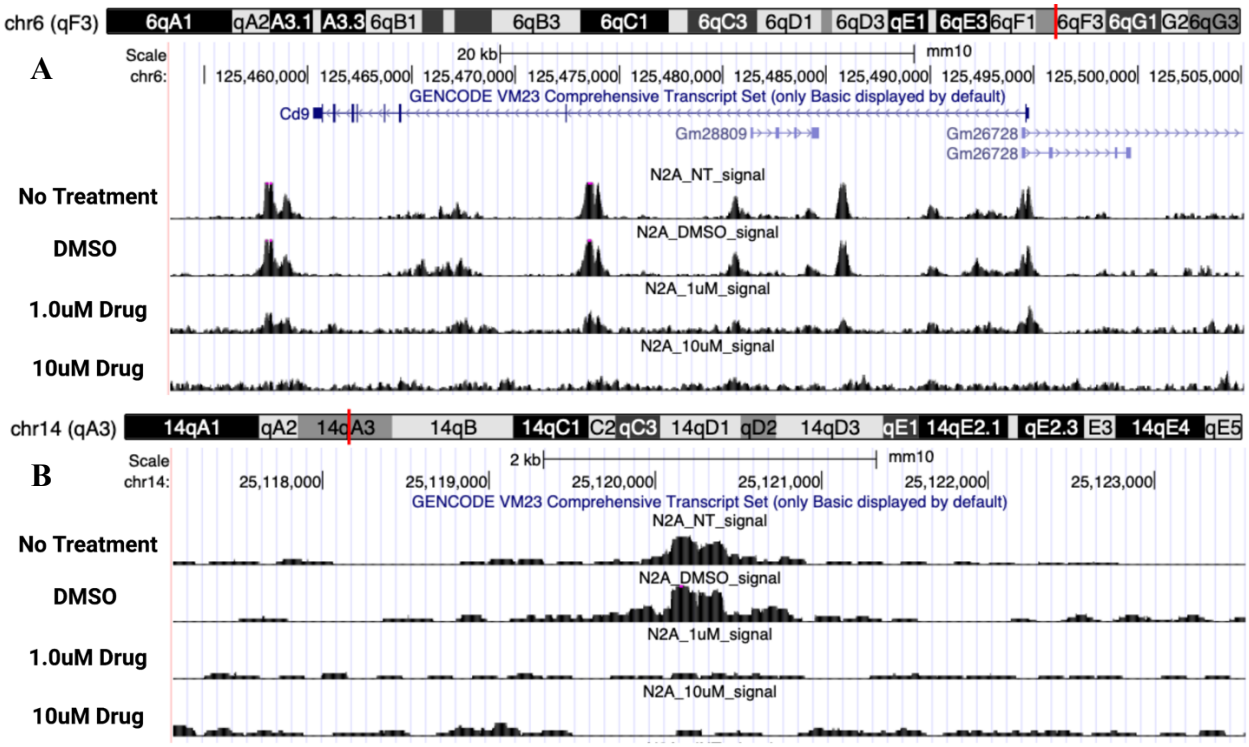

Figure 4: Genome browser images of locations of the genome where peaks are lost with drug treatment. A) CD9 is a protein coding gene for a family of intermembrane proteins that play a role in many cellular processes including differentiation, adhesion, signal transduction, and suppression of tumor metastasis (Ikeyama, et al. 1993). Both drug treatment samples exhibit reduced accessibility across the gene, with effects more pronounced in higher concentration (10uM) sample. B) This area of chromosome 14 is non-protein coding. The no treatment and DMSO treatment samples both show high accessibility, while both drug samples show no accessibility.

3.2: Characterizing regions of the genome where ARID1A and/or ARID1B are required to maintain chromatin accessibility.

3.2A: Optimizing siRNA transfections

To identify regions of chromatin that depend on either ARID1A or ARID1B for their accessibility, I worked to knockdown the expression of *Arid1a* or *Arid1b* and then assess changes in chromatin accessibility using ATAC-seq. However, I first had to optimize protocols for siRNA transfections, which had not been previously performed in the lab. To determine the optimal volume of LipoFECTamine RNAiMAX to maximize transfection efficiency and minimize cytotoxicity, I transfected cells with 10uM TYE-563 labeled non-targeting DsiRNAs, using either 7.5uL, 5.0uL, or 2.5uL of transfection reagent per well. I assessed transfection efficiency at 5, 24, and 48 hours after transfection via fluorescent imaging, and measured cell viability by trypan blue staining and counting on a hemocytometer. As shown in Figure 5, using 7.5uL of Lipofectamine per well showed the highest transfection efficiency by far, and did not lead to a notable difference in cytotoxicity to the lower amounts of transfection reagent. Therefore, 7.5uL of Lipofectamine RNAiMAX was used per well for siRNA transfections. In addition, I noted in these experiments that the TYE-563 labeled non-targeting DsiRNAs appeared to degrade after 48 hours (Figure 5A), so the transfection period was set to 30 hours to achieve maximum transfection and give time for the RNAi mechanism to function.

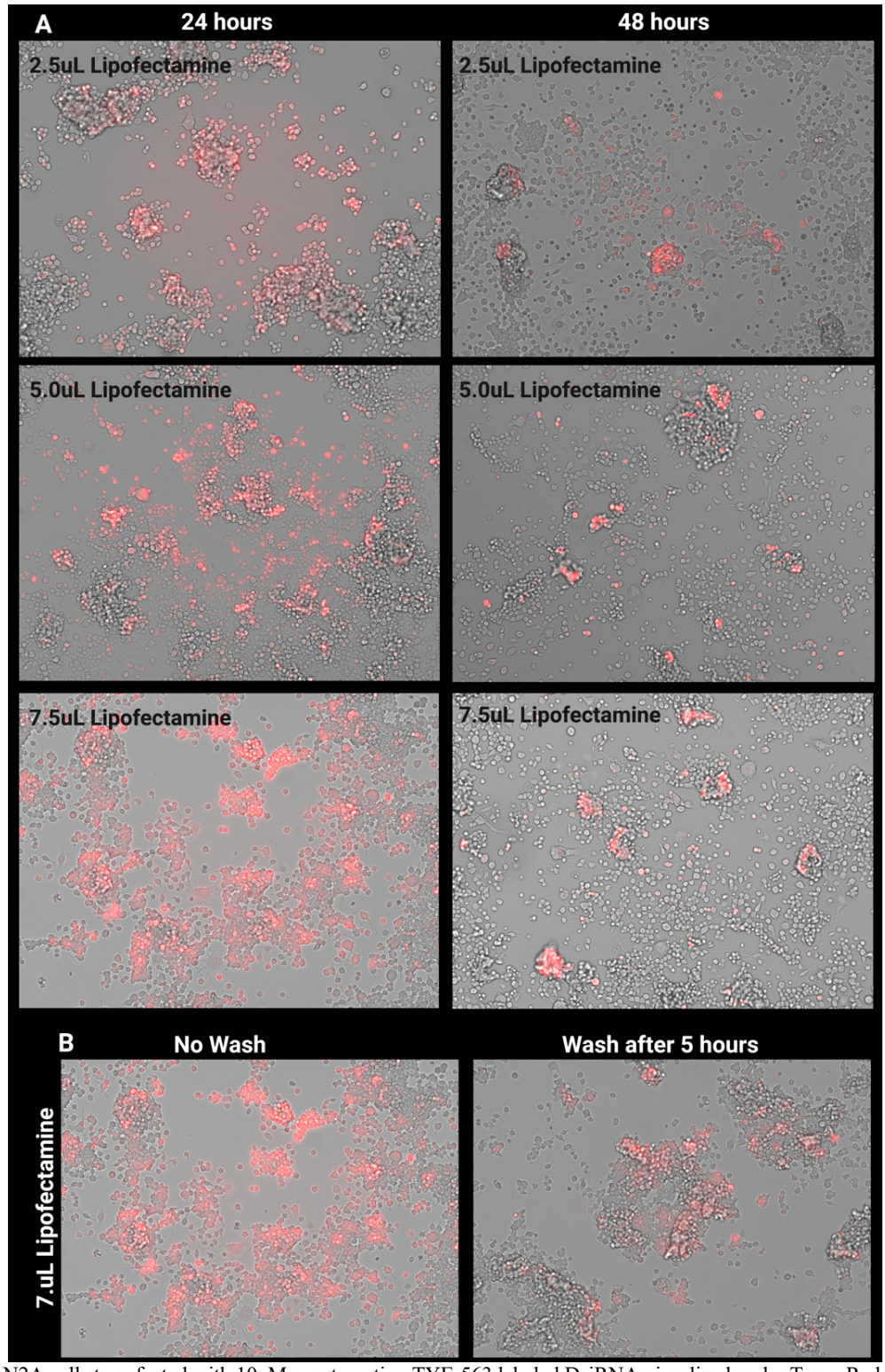

Figure 5: A) N2A cells transfected with 10uM non-targeting TYE-563 labeled DsiRNA visualized under Texas Red fluorescence 24 and 48 hours after transfection. Different volumes of transfection reagent used are 2.5uL, 5.0uL, or 7.5uL per well. The 7.5uL Lipofectamine well has the greatest level of transfection (> 90% of cells after 24 hours). After 48 hours DsiRNAs are notably degraded, suggesting this is too long of a transfection period. B) Wells are transfected with 7.5uL Lipofectamine and imaged 24 hours after transfection. Wash well is washed 5 hours after transfection. Wash wells showed notably less transfection efficiency, and No Wash wells did not show more cell death or unhealthiness, suggesting a wash is not needed or recommended for the transfection.

I also assessed the effects of performing a wash step 5 hours after transfection and replacing with fresh transfection medium, as suggested by some protocols. However, I saw including the wash step reduced the transfection efficiency when compared to plates with no wash (Figure 5) and did not noticeably affect cell viability. Thus, a 5-hour wash was not included in subsequent transfections.

3.2B: Establishing TaqMan Assays to measure *Arid1a* and *Arid1b* expression.

To ensure the TaqMan probe and primer sets were able to quantitatively assess target gene expression, I performed a standard cure qPCR experiment in which I assessed the standard Ct values for different cDNA concentrations to determine the efficiency of the primer assays (Figure 7). All assays showed high levels of efficiency (Table 3), and high levels of specificity and correct length of replicated sequence when run on a gel.

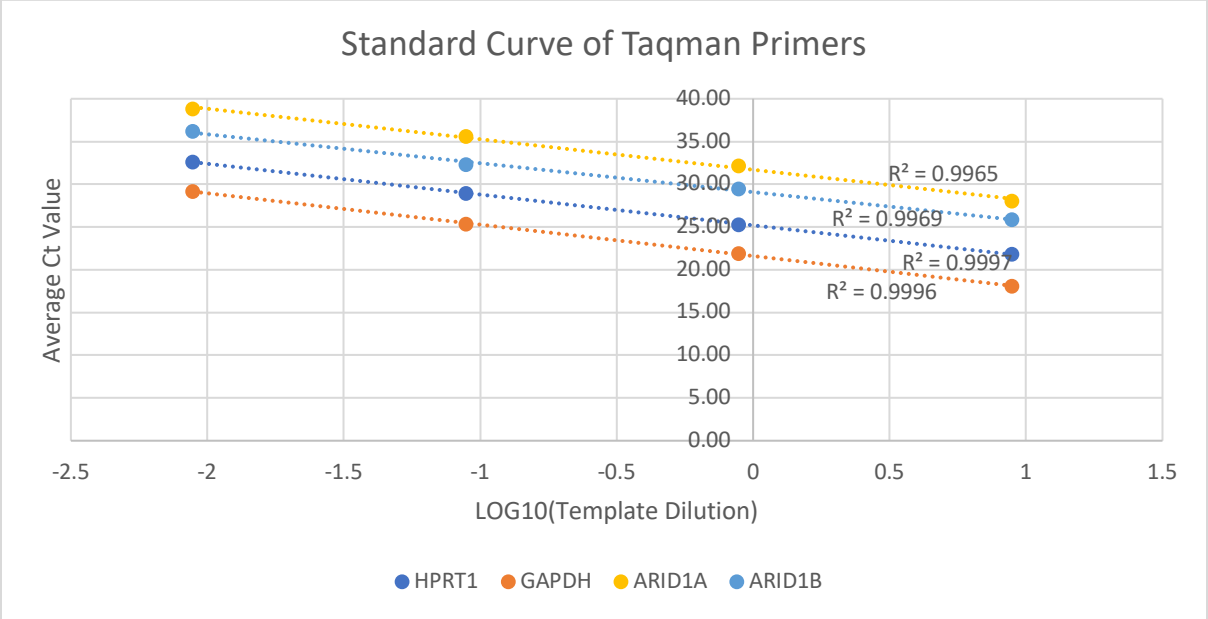

Figure 6: Standard curve created by performing qPCR on a serial dilution of template cDNA (cDNA concentrations of 8.87 ng/uL, 0.887 ng/uL, 88.7 pg/uL, and 8.87pg/uL). R^2 values all >0.99 show consistency of primers. Average Ct values for each template cDNA concentration represent expected Ct values to be used as reference. Slope of trendline used to calculate primer efficiency.

Primer	4pt Efficiency	Chromosome Location
Hprt1	89.5%	Chr.X: 52988078 - 53021660
Gapdh	87.2%	Chr.6: 125161338 - 125166511
Arid1a	90.5%	Chr.4: 133679008 - 133753981
Arid1b	97.5%	Chr.17: 4992755 - 5347656

Table 3: Efficiency of each primer calculated from the slope of the standard curve (E = 100*(-1+10-1/SLOPE)). Efficiency is high for all primers (>85% efficiency). Chromosome location of replicated sequence is noted.

3.2C: Confirming knockdown of positive control gene: *Hprt1*.

To ensure that siRNA knockdown was working, I tested the efficiency of the positive control DsiRNA, *Hprt1*-targeting, provided by IDT (TriFECTa Kit, IDT). Cells were treated with *Hprt1*-targeting siRNAs for 30 hours, collected using trypsinization, and then their RNA was isolated, and cDNA was synthesized. TaqMan qPCR was performed on cDNA of control and experimental samples, and qPCR data was used to calculate knockdown of *Hprt1* and assess the efficiency of the *Hprt1* -targeting siRNAs. The siRNAs showed >96% knockdown of *Hprt1*, and it was determined that the non-targeting siRNA controls could be used to normalize knockdown results (Figure 8, Table 4).

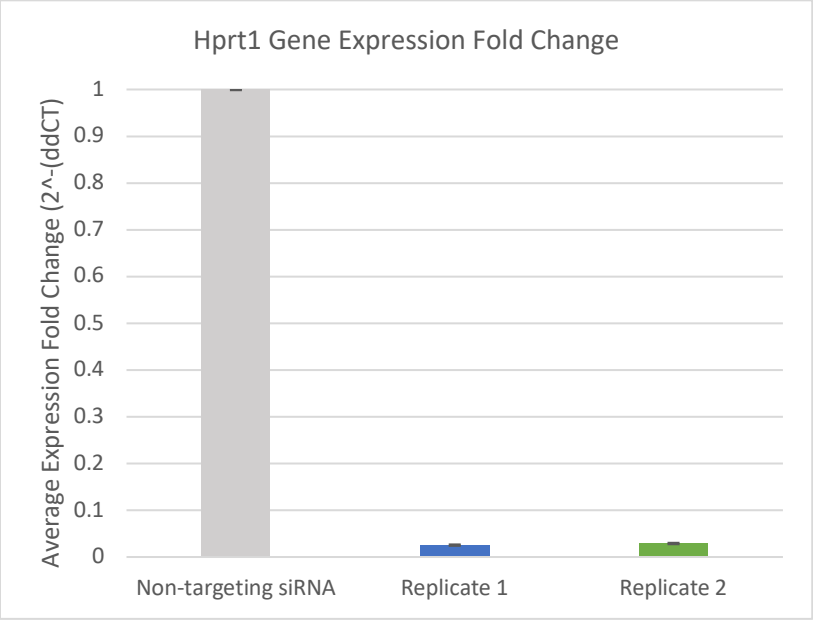

Figure 7: *Hprt1* gene expression fold change after knockdown. Both replicates treated with *Hprt1*-targeting siRNAs show almost no gene expression compared to sample treated with non-targeting siRNAs.

Sample	Percent Knockdown	
Normalized by:	Non-targeting siRNA Sample	No Treatment Sample
Replicate 1	97.46%	97.26%
Replicate 2	97.11%	96.71%

Table 4: Percent knockdown of *Hprt1* mRNA expression in samples treated with *Hprt1*-targeting siRNAs for 30 hours. All samples show >96% knockdown. Samples show same level of knockdown when normalized by non-targeting siRNA sample and untreated sample.

3.2D: Assessing siRNA knockdown of *Arid1b* expression.

To knockdown the expression of *Arid1b*, I tested three *Arid1b*-targeting DsiRNAs separately and in combination. I also assessed their effects on *Arid1a* expression, as sequence similarities between *Arid1a* and *Arid1b* may lead to off-target effects of the DsiRNAs. DiRNAs #2 and #3, as well as a combination of all three DsiRNAs, showed higher efficiency of *Arid1b* knockdown, with both #3 and the combination transfection samples showing >80% knockdown (Table 5). However, all samples also showed high knockdown and change in gene expression of *Arid1a* (Figure 9, Table 5), which made these DsiRNAs ineffective for further use. Unfortunately, this did not allow me to move forward with the *Arid1b* knockdowns as planned. To overcome this, I would design other siRNAs specifically created to avoid regions of similarity between *Arid1a* and *Arid1b* mRNA.

Arid1b-targeting DsiRNA Treatment	Percent Knockdown	
	Arid1b	*Arid1a*
#1	18.96	15.13
#2	77.52	33.10
#3	84.25	49.18
All	82.24	66.32

Table 5: Percent knockdown for *Arid1b* DsiRNAs. DsiRNA #1 shows poor knockdown of *Arid1a* and *Arid1b*. DsiRNA #2 shows some knockdown of *Arid1b*, but also knockdown of *Arid1a*. Both DsiRNA #3 and the combination of all DsiRNAs show high knockdown of *Arid1b*, but notable knockdown of *Arid1a*.

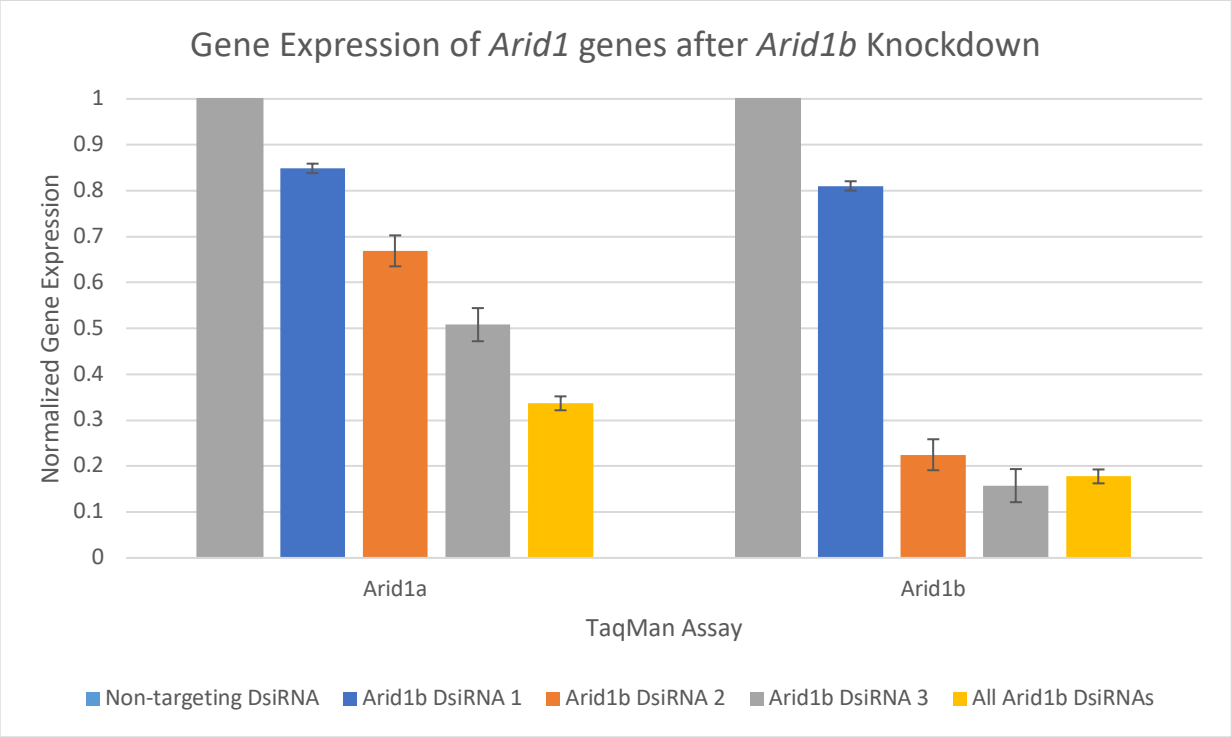

Figure 8: Gene expression of *Arid1b* DsiRNA transfection samples, with error bar values of standard error of the mean, normalized to expression of non-targeting DsiRNA negative control sample. DsiRNA #1 shows high gene expression for both *Arid1a* and *Arid1b*. DsiRNAs #2 and #3, and all DsiRNAs show low expression of *Arid1b*, however all also show a significant decreased expression of *Arid1a*.

3.2E: Assessing siRNA knockdown of *Arid1a* expression.

To knockdown the expression of *Arid1a*, I tested three *Arid1a*-targeting DsiRNAs separately and in combination, and assessed for potential effects on *Arid1b* expression. DsiRNAs #1 and #3 showed notable change in expression of *Arid1b* (Figure 10), with knockdown at approximately 40% for both DsiRNAs (Table 6), and thus were not used for further experiments. DsiRNA #2 and a combination of all three DsiRNAs were determined fit for further experimentation due to a high knockdown of *Arid1a* (>80%), and low knockdown of *Arid1b* (Figure 10, Table 6).

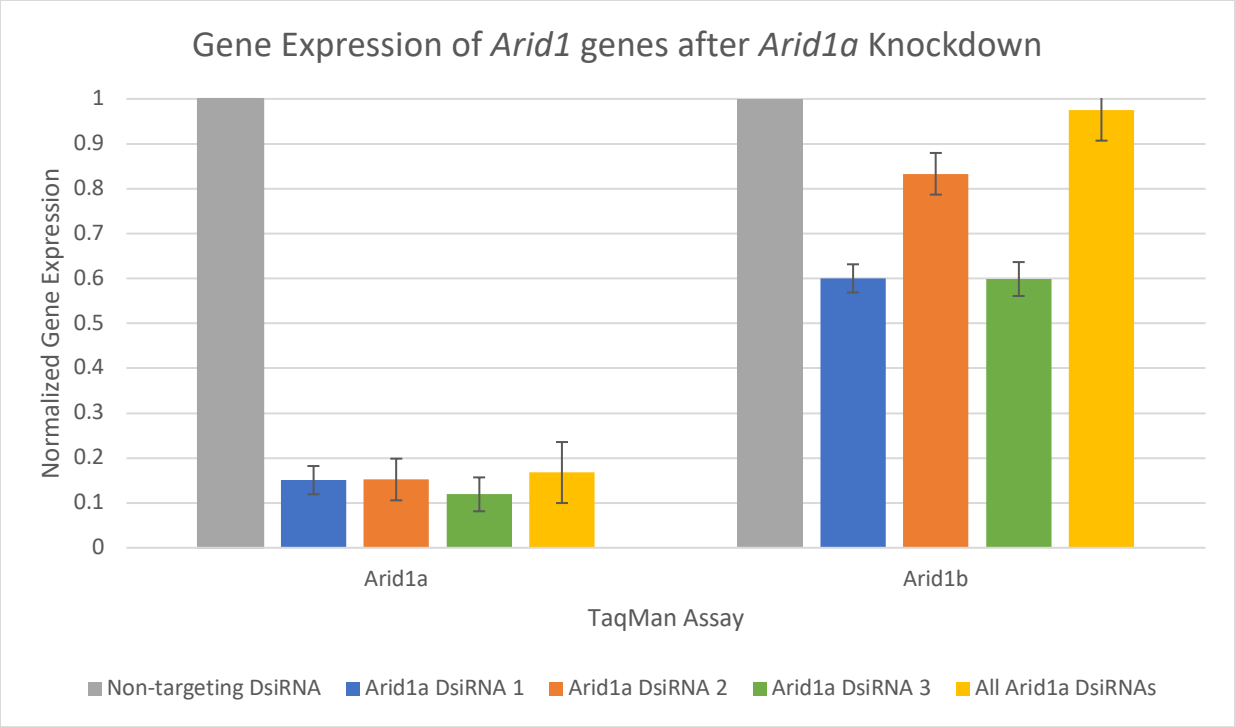

Figure 9: Gene expression of *Arid1a* DsiRNA transfection samples, with error bar values of standard error of the mean, normalized to expression of non-targeting DsiRNA negative control sample. All samples show low gene expression, <0.2, of *Arid1a*. DsiRNAs #1 and #3 show significant change in expression of *Arid1b*. DsiRNA #2 and all DsiRNAs show little change in gene expression of *Arid1b*.

Arid1a-targeting DsiRNA Treatment	Percent Knockdown	
	Arid1a	*Arid1b*
#1	84.93	39.99
#2	84.79	16.68
#3	88.09	40.13
All	83.24	2.51

Table 6: Percent knockdown for *Arid1a* DsiRNAs. All samples show high knockdown, >80%, of *Arid1a*. DsiRNAs #1 and #3 show notable knockdown of *Arid1a*, however DsiRNA #2 and all DsiRNAs show low knockdown of *Arid1b*.

To knockdown *Arid1a* for further analysis, N2A cells were treated with *Arid1a*-targeting DsiRNA #2 and the combination of all *Arid1a*-targeting DsiRNAs. All samples showed high knockdown of *Arid1a* (>80%) (Figure 11, Table 7). All DsiRNAs sample 2 shows very low change in expression (Figure 11) and very low knockdown of *Arid1b* (Table 7), while sample 1 shows some but still low amounts. DsiRNA #2 sample shows upregulation of *Arid1b*, this could

be due to cells reacting to knockdown of *Arid1a* with increased expression of *Arid1b* and incorporation of ARID1B into BAF complexes to compensate for the knockdown.

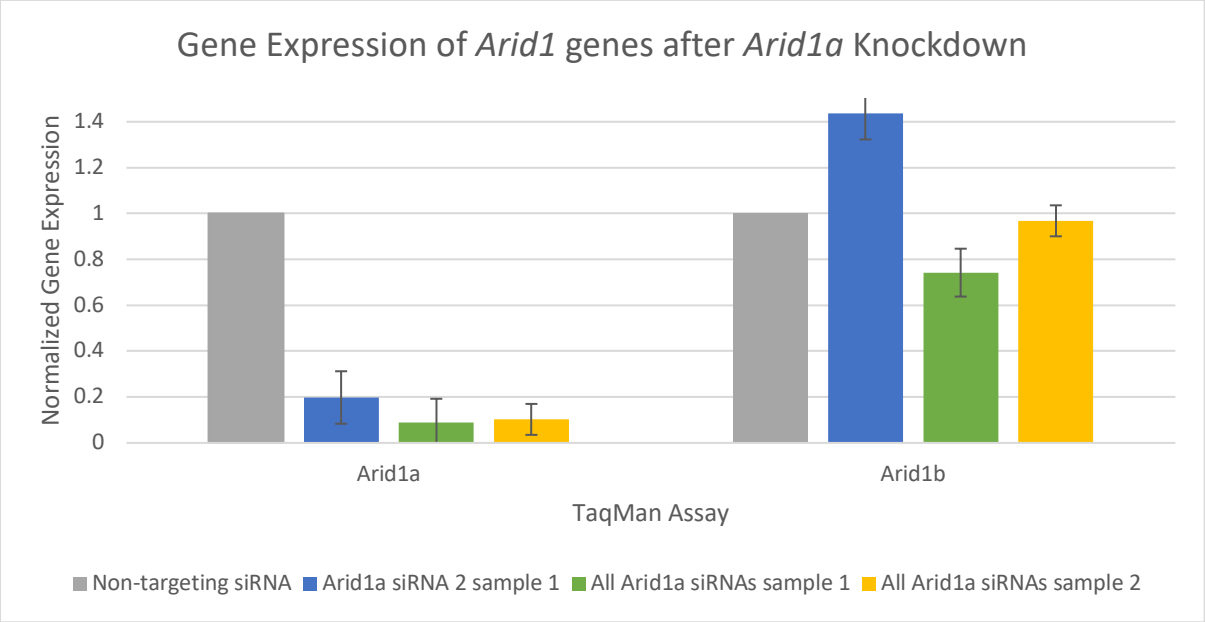

Figure 10: Gene expression of *Arid1* genes after *Arid1a* knockdown. All samples show low expression of *Arid1a* (>20%). All DsiRNAs sample 2 shows very low change in expression of *Arid1b*, while sample 1 shows some but still low amounts. Sample with DsiRNA #2 shows upregulation of *Arid1b*, with 143% expression compared to *Arid1b* expression of the non-targeting DsiRNA negative control.

Arid1a DsiRNA Treatment	Percent Knockdown	
	Arid1a	*Arid1b*
#2	80.20	-43.72
All Sample 1	91.20	25.77
All Sample 2	89.75	3.19

Table 7: All samples show high knockdown of *Arid1a* (>80%). Both samples of all DsiRNAs show low knockdown of *Arid1b*. Sample with DsiRNA #2 shows upregulation of *Arid1b*, with 143% expression compared to Arid1b expression of the non-targeting DsiRNA negative control.

3.2F: Characterizing changes in chromatin accessibility after *Arid1a* Knockdown.

To characterize where in the N2A cell genome the BAF complexes containing ARID1A work to regulate gene expression, ATAC-seq was performed on the *Arid1a* knockdown samples and libraries were sequenced to identify changes in chromatin accessibility when ARID1A-containing BAF complexes were knocked down. After aligning the ATAC-seq sequencing reads to the reference genome (mm10, ENCODE), all samples were determined to be of good quality

as they all show strong enrichment for accessible chromatin at annotated transcription start sites (Figure 12), high percentage of unique reads, and low fraction of mitochondrial reads (Table 8).

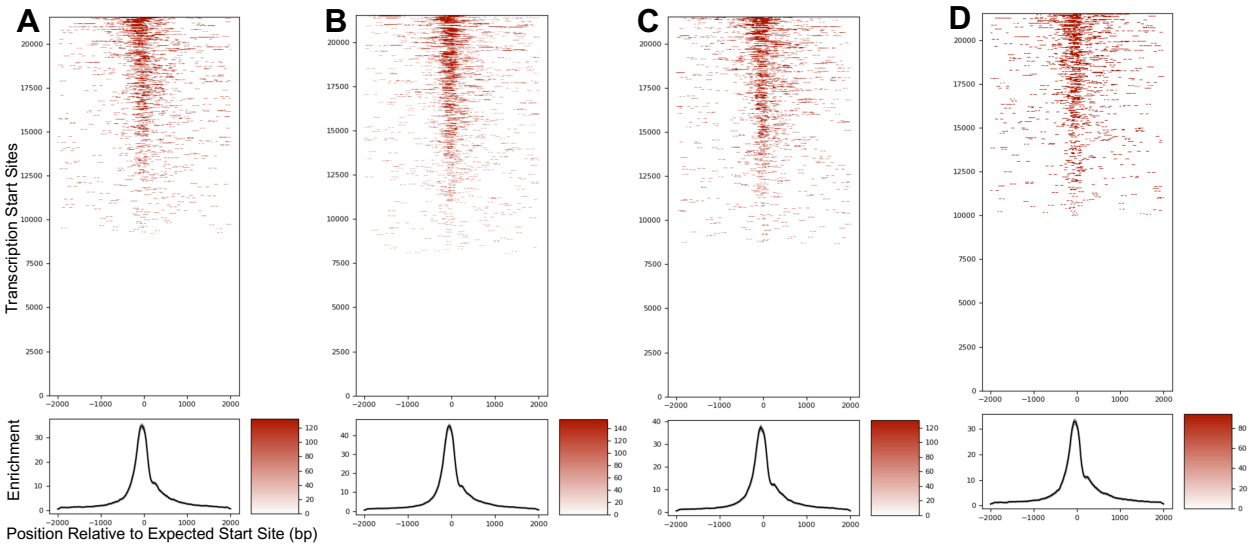

Figure 11: Transcription start site (TSS) enrichment for A) Non-targeting DsiRNA sample, B) Arid1a DsiRNA #2 sample, C) All Arid1a DsiRNAs sample 1, D) All Arid1a DsiRNAs sample 2.

DsiRNA Treatment	Percent of Sequence Count		TSS Enrichment	Fraction of Mitochondrial Reads
	Unique Reads	Duplicate Reads		
Non-targeting	80.7	19.3	34.8191	0.2507
Arid1a #2	80.0	20.0	44.8035	0.2442
All Arid1a Sample 1	83.1	16.9	37.5948	0.2219
All Arid1a Sample 2	76.5	23.5	33.0131	0.3028

Table 8: Percent of sequence counts of unique and duplicate reads, transcription start site (TSS) enrichment, and fraction of mitochondrial reads for all samples. All samples show high percentage of unique reads, and low percentage of duplicate reads. All samples show high TSS enrichment and low fraction of mitochondrial reads, showing quality of ATAC-seq libraries.

Preliminary analysis of the ATAC-seq data shows changes in genome accessibility in response to *Arid1a*-targeting DsiRNA transfection, with changes in accessibility seen at both intergenic and intragenic regions (Figure 12). All regions where changes in chromatin accessibility are seen in the *Arid1a* knockdown samples also exhibit the same changes in accessibility in the BRM014 treatment samples (Figure 12), suggesting these areas are where ARID1A-containing BAF complexes are regulating chromatin accessibility. However, *Arid1a*

knockdown samples do not exhibit changes in chromatin accessibility at all locations where changes are seen in the BRM014 samples. These regions could be areas where ARID1B-containing BAF complexes always regulate gene expression or where they are able to compensate for the *Arid1a* knockdown.

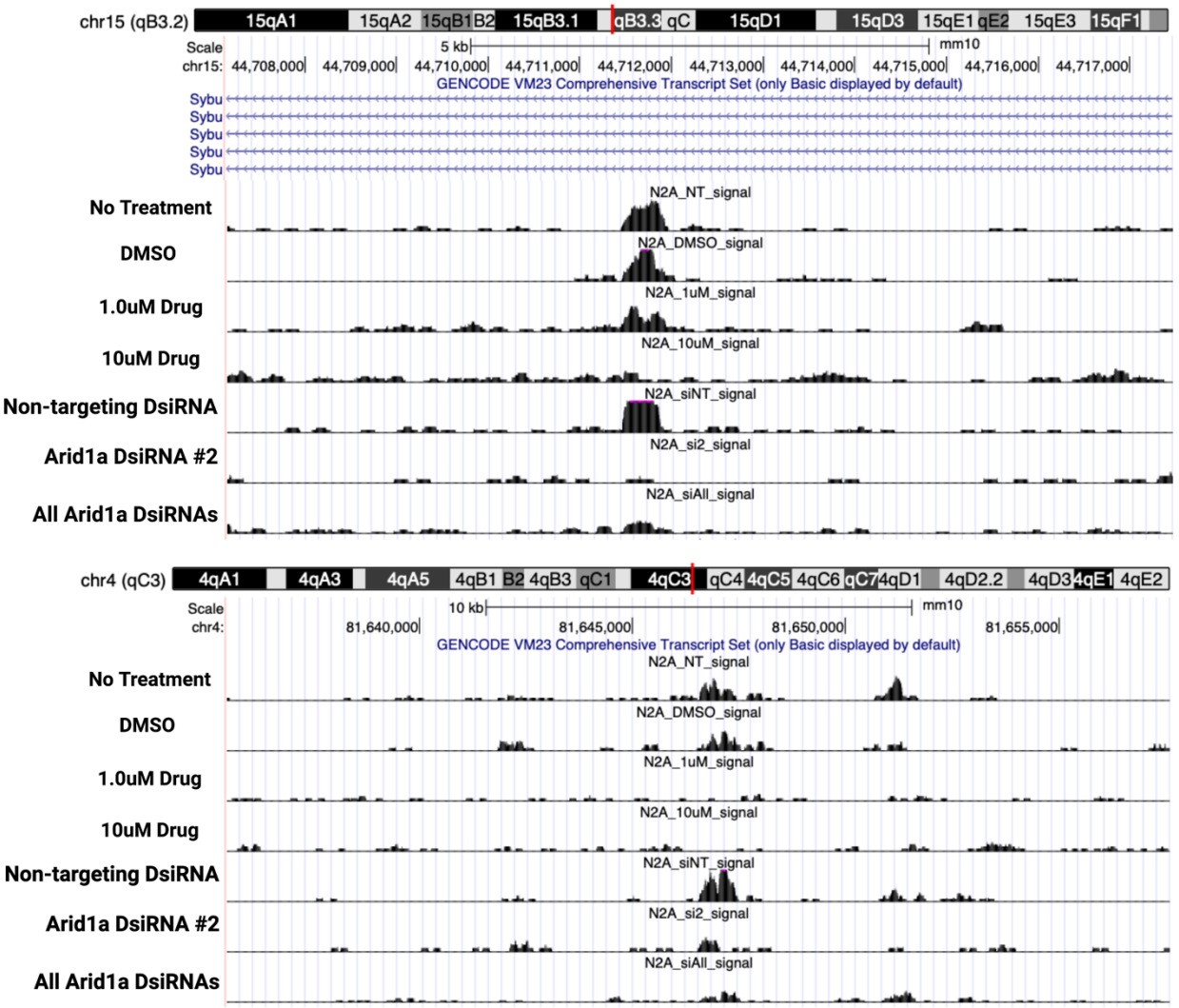

Figure 12: Genome browser images of locations of the genome where peaks are lost with DsiRNA transfection. A) Sybu is a protein coding gene for Syntabulin/GOLSYN, part of a kinesin motor-adaptor complex that is essential for anterograde axonal transport and presynaptic assembly during neuronal development (Cai et al., 2007). Both Arid1a DsiRNA transfection samples exhibit reduced accessibility at this intronic region of the gene. B) This area of chromosome 4 is non-protein coding. Both Arid1a DsiRNA transfection samples both show decreased accessibility.

Further analysis of this data will include performing statistical analyses to identify regions with significant changes in chromatin accessibility upon *Arid1a* knockdown and investigating if they are distinguished by other features such as enrichment for certain TF-

binding motifs or histone modifications. This set of ATAC-seq libraries will also be further compared to the libraries of BRM014-treated samples, to determine at which chromatin regions ARID1A-BAF complexes regulate accessibility.

Chapter 4: Conclusions

In this work, I began to characterize where in the N2A cell genome BAF complexes, specifically those containing either ARID1A or ARID1B, regulate gene expression. First, I treated cells with an allosteric inhibitor of the BAF complex ATPase subunits, and measured changes in chromatin accessibility across the genome. I then optimized siRNA transfection protocols, and validated siRNA knockdown of *Arid1a* and *Arid1b*. I measured changes in chromatin accessibility in *Arid1a* knockdown samples and began preliminary analysis of chromatin accessibility data of both *Arid1a* knockdown and BRM014-treated samples. I found that both datasets exhibit changes, mostly loss, of accessibility across different areas of the genome. These novel datasets present the opportunity to identify more information about the role of BAF complexes, and *Arid1a*, in regulating chromatin accessibility.

There are some limitations to these studies, caused in part by the time-constraints of this project. Conclusions made from the dataset thus far are limited by a lack of replicates, so further experiments are necessary to produce more evidence to make conclusions to identify the different roles of ARID1A- or ARID1B-containg BAF complexes. A successful knockdown of *Arid1b* is also necessary to complete the sequencing dataset for characterization of where in the genome ARID1A- or ARID1B-containing BAF complexes regulate gene expression. To accomplish this, I would design DsiRNAs that target areas of *Arid1b* mRNA that are less conserved across both *Arid1* variants. Using ATAC-seq data from these samples, I would then identify the cREs regulated by either ARID1A- or ARID1B-containing BAF complexes and assess the type of cREs regulated (e.g. promoters, enhancers). I would then assess if the cRES that gain or lose accessibility are distinguished by other features, such as enriched for certain TF-binding motifs, histone modifications, or the identity of cREs and target genes. This may provide

insight into the role of the different ARID1 subunits, how they are recruited, and reveal causality or provide clarity to the mechanisms of disease produced by mutations in *Arid1a* or *Arid1b*. In addition, further experimentation combining a low dose of ATPase inhibitor treatment with *Arid1* subunit-targeting DsiRNA knockdown could be used to sensitize the cells to potentially block compensation and further isolate the individual roles of ARID1A and ARID1B.

Altogether, the assays that I have established, and the datasets and results presented here, provide foundational data that pave the way forward for further and more detailed characterization of the cREs and target genes that may be differentially regulated by ARID1A- and ARID1B-containing BAF complexes. These investigations will likely contribute to a better understanding of the associated SSRIDDs, and to exploring potential therapeutic targets for associated cancers.

Works Cited

Allis, C. D. and T. Jenuwein (2016). "The molecular hallmarks of epigenetic control." Nat Rev Genet 17(8): 487-500.

Bogershausen, N. and B. Wollnik (2018). "Mutational Landscapes and Phenotypic Spectrum of SWI/SNF-Related Intellectual Disability Disorders." Front Mol Neurosci 11: 252.

Buenrostro, J. D., et al. (2013). "Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position." Nat Methods 10(12): 1213-1218.

Cai, Q., et al. (2007). "Syntabulin-kinesin-1 family member 5B-mediated axonal transport contributes to activity-dependent presynaptic assembly." J Neurosci 27(27): 7284-7296.

Celen, C., et al. (2017). "Arid1b haploinsufficient mice reveal neuropsychiatric phenotypes and reversible causes of growth impairment." Elife 6.

Ellegood, J., et al. (2021). "Neuroanatomy and behavior in mice with a haploinsufficiency of AT-rich interactive domain 1B (ARID1B) throughout development." Mol Autism 12(1): 25.

Gasperini, M., et al. (2020). "Towards a comprehensive catalogue of validated and target-linked human enhancers." Nat Rev Genet 21(5): 292-310.

Ikeyama, S., et al. (1993). "Suppression of cell motility and metastasis by transfection with human motility-related protein (MRP-1/CD9) DNA." J Exp Med 177(5): 1231-1237.

Iurlaro, M., Stadler, M.B., Masoni, F. et al. Mammalian SWI/SNF continuously restores local accessibility to chromatin. Nat Genet 53, 279–287 (2021). https://doi.org/10.1038/s41588-020-00768-w

Jimenez, C., et al. (2022). "Structural disruption of BAF chromatin remodeller impairs

neuroblastoma metastasis by reverting an invasiveness epigenomic program." Mol Cancer 21(1): 175.

Jung, E. M., et al. (2017). "Arid1b haploinsufficiency disrupts cortical interneuron development and mouse behavior." Nat Neurosci 20(12): 1694-1707.

Kadoch, C. and G. R. Crabtree (2015). "Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics." Sci Adv 1(5): e1500447.

Mashtalir, N., et al. (2018). "Modular Organization and Assembly of SWI/SNF Family Chromatin Remodeling Complexes." Cell 175(5): 1272-1288 e1220.

Mathur, R., et al. (2017). "ARID1A loss impairs enhancer-mediated gene regulation and drives colon cancer in mice." Nat Genet 49(2): 296-302.

Moffat, J. J., et al. (2021). "Differential roles of ARID1B in excitatory and inhibitory neural progenitors in the developing cortex." Sci Rep 11(1): 3856.

Moffat, J. J., et al. (2019). "The role of ARID1B, a BAF chromatin remodeling complex subunit, in neural development and behavior." Prog Neuropsychopharmacol Biol Psychiatry 89: 30-38.

Moffat, J. J., et al. (2022). "Neurobiology of ARID1B haploinsufficiency related to neurodevelopmental and psychiatric disorders." Mol Psychiatry 27(1): 476-489.

Pagliaroli, L., et al. (2021). "Inability to switch from ARID1A-BAF to ARID1B-BAF impairs exit from pluripotency and commitment towards neural crest formation in ARID1B-related neurodevelopmental disorders." Nat Commun 12(1): 6469.

Raab, J. R., et al. (2015). "Genome-Wide Transcriptional Regulation Mediated by Biochemically Distinct SWI/SNF Complexes." PLoS Genet 11(12): e1005748.

Ronan, J. L., et al. (2013). "From neural development to cognition: unexpected roles for chromatin." Nat Rev Genet 14(5): 347-359.

Schick, S., et al. (2019). "Systematic characterization of BAF mutations provides insights into intracomplex synthetic lethalities in human cancers." Nat Genet 51(9): 1399-1410.

Sokpor, G., et al. (2017). "Chromatin Remodeling BAF (SWI/SNF) Complexes in Neural Development and Disorders." Front Mol Neurosci 10: 243.