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.

Sona Davis April 10, 2024

Investigating the role of the BAF Complex in Human Disease and Evolution

by

Sona Davis

Dr. David U Gorkin
Adviser

Anthropology

Dr. David U Gorkin
Adviser

Dr. John Lindo
Committee Member

Dr. Michal Arbilly
Committee Member

2024

Investigating the role of the BAF Complex in Human Disease and Evolution

By

Sona Davis

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

Anthropology

2024

Abstract

Investigating the role of the BAF Complex in Human Disease and Evolution
By Sona Davis

This Honors Thesis contains two projects each addressing gaps in our current knowledge on the BAF complex. The first project (described in chapter 2) sought to identify areas of the genome where the BAF Complex works to remodel chromatin. The second project (described in chapter 3) sought to determine whether there are genetic variations in BAF subunit genes between modern humans, ancient humans, and non-human primates that could contribute to differences in brain development in these populations. Together, these projects deepen our understanding of the role of the BAF Complex in human evolution and disease. The results provide strong baselines for future research on BAF function, as well as intriguing insights into its implications. Moreover, they pave the way for potential therapeutic interventions and open up new avenues for exploring the intricate interplay between the BAF Complex and various biological processes.

Investigating the role of the BAF Complex in Human Disease and Evolution

By

Sona Davis

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

Anthropology

2024

Acknowledgements

I would like to extend my heartfelt appreciation to my advisor, Dr. Gorkin, whose guidance and mentorship have been invaluable throughout my academic journey. I am so lucky that I stumbled upon your name in a Biology class three years ago, and reached out via email to became one of the first undergrads to join your lab. Over this period, you have not only provided invaluable advice but has also served as an inspiring role model, continuously encouraging me to push boundaries and strive for excellence.

I would also like to express my gratitude to the members of the Gorkin Lab: Alex Gulka, Susan Plott, Sara Flano, Devin Gee, Ziben Zhou, Yonina Loskove, Alan Kang, and Wendy Wang. Special thanks go to Alex Gulka and Sara Flano, who served as mentors for multiple processes in the lab. Their guidance has been invaluable, and I could always turn to them when I was in need of help. The unwavering support, camaraderie, and friendship of each lab member have transformed my lab experience into an absolute pleasure. Their collective contributions have enriched my research endeavors and significantly contributed to my growth as a scholar.

Special thanks are due to Dr. Lindo for graciously agreeing to serve on my committee, offering invaluable guidance, and providing essential resources for the success of this project. I am also grateful to Dr. Arbilly for her contributions as a committee member.

Lastly, I express my heartfelt gratitude to my friends and family for their unwavering support throughout this project. To Mom and Dad, thank you for always encouraging me to explore new avenues, and take on projects beyond the classroom (even when I may already have a lot on my plate). I owe much of who I am today to both of you and everything you have done to make this education, and therefore this project, possible. Thanks to Neel and Elric for being amazing brothers, and to our dog Lucy for being there to offer comfort when needed.

To my best friend and roommate, Emily Silver, thank you for always inspiring me to be better. Experiencing college and the thesis writing process alongside you has been a privilege, and I cannot wait to continue life's journey by your side. To Mo, thank you for being my go-to when I need a break, a pep talk, or anything in between. Your love and encouragement mean the world to me, and no amount of thanks could ever express my appreciation for your unwavering support.

List of Figures

Table of Contents

Overview of Project 1:

Although mutations in BAF complex subunits are common in cancer and NDD, it remains unknown which specific cREs in the genome rely on the BAF complex for their accessibility and function. In my first project, I worked to identify the cREs that rely on BAF complexes for their accessibility and regulatory function in the human cancer cell line K-562. I reasoned that cREs which depend on BAF for their accessibility would become inaccessible if BAF activity were inhibited, as had been shown previously on other cell types (REF). Thus, to experimentally manipulate BAF's function, I used a small molecule allosteric inhibitor of BAF's ATPase subunits SMARCA2 and SMARCA4. Without the ability to hydrolyze ATP, BAF no longer functions properly and cREs that are reliant on BAF lose their chromatin accessibility. Then, I performed ATAC-seq on these cells to map chromatin accessibility genome-wide in the presence and absence of BAF inhibition. I aimed to identify the cREs that lose chromatin accessibility when the BAF complex activity is inhibited. Based on similar work in other cell lines, I expected to see widespread changes in chromatin accessibility in the absence of BAF activity. However, surprisingly, the resulting K-562 data showed few differences in chromatin accessibility between treatment groups. These results show that the cancer cell line K-562 is less sensitive than other cell types tested with the ATP-ase inhibitor used in my study. One possible explanation, to be explored in the future, is that K562 cells may contain a mutant form of the BAF complex that makes them resistant to the ATP-ase inhibitor used in my study.

Overview of Project 2:

In humans, BAF complex proteins are highly conserved, with mutations having profound effects on survival, neuronal development, and cognitive outcomes. The importance of BAF in human brain development led me to question whether genetic variants in BAF subunit genes might play a role in neuronal differences between modern humans and Neanderthals or non-human primates.. Thus, in my second project, I explored the differences in BAF protein coding sequences among archaic human Neanderthals, non-human primates, and modern humans. To achieve this, I utilized lists of genetic variants between modern humans, Neanderthals, and non-human primates that were kindly provided by Dr. John Lindo. I then filtered these lists to identify genetic variants located within genes that encode BAF protein subunits. To identify genetic variants that could impact BAF protein structure and function, I used several computational tools that predict the impact of genetic variants on protein function. I identified several genetic variants specific to Neanderthals or non-human primates (and not found in modern humans) that are predicted to impact the function of BAF complex proteins. This research establishes a list of variants that can be tested in future experiments to determine whether they alter BAF function and/or brain development in ways that may explain phenotypic differences between Neanderthals, primates, and present-day humans. By studying the similarities and differences in these systems, we can gain a deeper understanding of the evolutionary history of the BAF complex's work in the brain.

Chapter 1: Introduction

1.1: DNA Structure and Packaging

DNA, or deoxyribonucleic acid, serves as the genetic blueprint for the development and functioning of all living organisms (Minchin & Lodge, 2019). DNA is a class of the macromolecule known as nucleic acid, characterized by repeating units of nucleotides. Each nucleotide comprises a five-carbon sugar, deoxyribose, a nitrogenous base, and one phosphate group. DNA is composed of four nucleotides, distinguished by the nitrogenous base they contain (Thymine, Cytosine, Adenine, or Guanine). To form one strand, each monomer links to the next through its 3' carbon's hydroxyl group and the following nucleotide's phosphate group. However, to create the well-known coil-like structure, two strands running in the opposite direction connect through the hydrogen bonds of their corresponding nitrogen bases. With their kinked sugar-phosphate backbones on the outside of the strand, the double helix is formed (Minchin & Lodge, 2019).

During the process of transcription, RNA polymerase utilizes the DNA sequence as a template to read and then form a complementary RNA strand (Minchin & Lodge, 2019). Later, ribosomes translate the RNA genetic code into proteins for various processes throughout the cell. This is referred to as "gene expression", or the process by which DNA's genetic information leads to the creation of a functional product, in many cases a protein (Minchin & Lodge, 2019). While DNA is often discussed in this structure, for it to fit and be stored within a cell's nucleus, it must be tightly packaged. DNA is packaged around histone proteins to create a substance known as **chromatin,** which consists of DNA and its associated packaging proteins. The basic units of chromatin are nucleosome particles, which each contain ~150 bp of DNA wrapped about 8

histone proteins. Histone proteins carry a large positive charge that attracts to the negatively charged DNA and stabilizes the nucleosome structure (Simpson et al., 2023).

1.2: Chromatin-Based Regulation of Gene Expression

The structure of chromatin contributes to gene regulation by influencing the accessibility of DNA sequences to proteins that influence transcription (Andersson & Sandelin, 2020). Mammalian genomes contain thousands of non-protein-coding regulatory sequences, known as cis–regulatory elements or **cREs** which help regulate the transcription of genes in response to specific developmental and environmental signals. These cREs are segments of DNA that contain binding sequences recognized by Transcription Factor (**TF**) proteins that activate or repress the transcription of genes located on the same DNA molecule (Andersson & Sandelin, 2020).

However, for proteins such as Transcription Factors to bind cREs and influence transcription, the DNA sequences they bind to must be made accessible through the removal or displacement of the histone proteins. Thus, prior to influencing transcription, cREs undergo a process called “chromatin remodeling”, where the nucleosome particles that package DNA are moved aside to enable TF binding. This process of chromatin remodeling creates short stretches of DNA that are devoid of packaging histone proteins, which are referred to as regions of “accessible chromatin.”

My research has focused on a class of proteins known as **chromatin remodelers**, which are responsible for creating and maintaining accessible chromatin at cREs (Klemm et al., 2019). Chromatin remodelers are enzymes that utilize ATP to slide or eject nucleosomes and alter their positioning on DNA (Kadoch & Crabtree, 2015). With packaging proteins moved, chromatin is left exposed for TFs to bind and influence gene expression. The core of my thesis is dedicated to

a particular family of chromatin remodelers known as BRG1/BRM-Associated Factor (**BAF**), which I will describe in more detail below (Kadoch & Crabtree, 2015).

1.3: BRG1/BRM Associated Factor Family of Chromatin Remodeling

The BRG1/BRM Associated Factor (BAF, also known as SWI/SNF) family of chromatin remodelers are large protein complexes, each consisting of multiple subunits. Each BAF complex contains one ATPase subunit, either SMARCA2 (*BRG1*) or SMARCA4 (*BRM*), in conjunction with multiple other protein subunits (Kadoch & Crabtree, 2015). BAF family proteins exhibit high evolutionary conservation, and orthologs are found in all eukaryotic organisms (Kadoch & Crabtree, 2015). Humans have 31 different genes that encode BAF subunits, with a single BAF complex containing between nine to thirteen of these subunits. Previous studies have shown that BAF complexes are required to maintain accessible chromatin at cREs in human and mouse cells (Shick et Al, 2021; Iurlaro et al., 2021). This activity of BAF complexes is thought to be essential for survival at both the cellular and whole organism levels. Deletion of a single BAF subunit SMARCA4 leads to embryonic lethality in mice prior to implantation (Bultman et al., 2000).

1.4: BAF Mutations in Human Disease

Genes that code for BAF subunits are among the most frequent targets of genetic mutations in cancer and in neurodevelopmental disease. While a normally functioning BAF serves as a tumor suppressor in our body, dysfunctional BAF complexes can cause malignancies (Hodges et al., 2016). It estimated that approximately 20% of all human cancers have a somatic mutation in at least one BAF subunit (Kadoch & Crabtree, 2015). This makes the BAF complex one of the most commonly mutated functional units in cancer.

BAF subunit genes are also among the common targets of mutations in patients with Neurodevelopmental disorders (**NDDs**) (Bögershausen & Wollnik, 2018). Most of these mutations are *de novo*, appearing for the first time in the patient. The phenotypes of these patients vary, but can lead to a variety of diagnoses including Intellectual Disability (**ID**), Autism Spectrum Disorder (**ASD**), and or Developmental Delay (**DD**). When BAF mutations are found in these patients through genetic testing, they are usually given a more specific genetic diagnosis of Coffin Siris Syndrome or SWI/SNF-related intellectual disability disorder (Bögershausen & Wollnik, 2018).

The fact that BAF mutations can cause NDDs reflects the important roles of BAF in gene regulation during brain development. During the development of the nervous system, tissue-specific transcription factors are known to interact with BAF (Ronan et al., 2013). For a cell to specify into a unique identity required for the tissue it will be a part of, tissue-type specific genes must be repressed and expressed. By changing chromatin organization, BAF allows for the TFs to bind and facilitate this process of differentiation via controlled gene expression. When BAF cannot properly remodel chromatin structure, cREs are not accessible for their respective transcription factors to bind, and it leads to altered gene expression. This, in turn, affects cell fate determination.

Chapter 2: Identifying DNA sequence elements that rely on the BAF Complex

2.1: Background

2.1.2: BAF complexes are bound at cREs and required to maintain chromatin accessibility.

BAF operates at thousands of cREs, and demonstrates significant affinity to cREs such as enhancer sites (Kadoch & Crabtree, 2015). Enhancers are genomic regions where TF binding stimulates the transcription of genes. However, further investigation is necessary to understand which specific cREs' localize BAF (Kadoch & Crabtree, 2015).

The BAF Complex has been established as a requirement for proper chromatin accessibility at cREs (Shick et al., 2021; Iurlaro et al., 2021). In order to confirm BAF's role in remodeling chromatin, previous studies arrested its function with a small molecule inhibitor of the BAF ATPase subunits SMARCA2 and SMARCA4. The drug, BRM014, cuts off the complex's energy supply by inhibiting the subunit responsible for hydrolyzing Adenosine-Triphosphate (ATP) (Papillon et al., 2018). Without ATPase, BAF no longer operates and regulatory sequences that rely on BAF have instances of altered chromatin accessibility. I will be using this same small molecule inhibitor in my research as it has shown to effectively cut off ATPase activity (Shick et al., 2021; Iurlaro et al., 2021; Papillon et al., 2018).

In addition to utilizing this drug, previous studies have employed a method of protein degradation to precisely knock-out the ATPase subunit SMARCA2 (Shick et al., 2021). With two sample groups containing methods to ensure disruption of BAF activity, each underwent an

ATAC-seq experiment to measure chromatin accessibility across the genome. The end product of the procedure confirmed that, in both groups, loss of SMARCA4 and BAF function results in very rapidly lost regions of the DNA accessibility. Of the DNA sequence elements that were influenced, enhancer regions vital for cell specialization were the most prominent. The results conclude that BAF complexes are required to maintain chromatin accessibility however calls for future studies on tumor models to generalize the findings (Shick et al., 2021; Iurlaro et al., 2021).

2.1.3: Key gap(s) in knowledge

Despite the recognized significance of BAF in neurodevelopmental diseases and cancer, the individual cREs dependent on BAF for accessibility and function remain unidentified. Moreover, the factors responsible for recruiting BAF to these regions are not clearly understood. A deeper understanding of the specific chromatin regions influenced by the BAF complex will serve as a foundational step for research. To identify the origins of cancer and diseases associated with BAF, it is crucial to determine where to focus attention for developing therapeutic treatments.

2.1.4: Our approach

To address the current gap in knowledge, my project aimed to identify cREs in the human cancer cell line K-562 that rely on BAF for accessibility. The K-562 line originates from the cancerous white blood cells of a leukemia patient. The K-562 cell line was chosen because it is extensively studied, and there is a well-characterized map of TF binding locations in K-562 cell genomes.. This preexisting data allows me to to categorize and contrast which subsets of regulatory elements (promoters, enhancers, silencers) may or may not be dependent on BAF for accessibility. Identifying such patterns within my resulting data could confirm our hypothesis that specific categories of regulatory elements are more likely to be linked to the BAF complex.

It is also important to consider that K-562, as a cancer-derived cell line, may have mutations in BAF subunits that alter normal BAF function. As mentioned earlier, mutations in BAF genes are believed to contribute to more cases of human cancer than those in any of the other three chromatin remodeling families (Kadoch & Crabtree, 2015). This raises the possibility that observations made in K-562 might not generalize to other cell lines. However, despite this potential limitation, we decided to move forward with K-562 as the most suitable cell line for this project due to the abundant availability of data on it.

To establish an experimental group for testing the effects of inactive BAF, I administered the drug BRM014. This small molecular inhibitor effectively halts the ATP-ase necessary for the complex's ability to slide and eject nucleosomes (Papillon et al., 2018). To measure the effects of BAF inhibition on chromatin accessibility, I conducted ATAC-seq on samples with and without BRM014. The ATAC-seq technique cleaves regions of accessible DNA and tags them with short DNA adapters amenable for Illumina sequencing. When ATAC-seq libraries are sequenced and mapped to the reference genome, the result is that sequencing reads pile up to form “peaks” at regions of the genome that were accessible in the assayed sample (e.g. K-562 cells). This data can be analyzed to reveal differences in chromatin accessibility in the presence and absence of BAF. After identifying regions of the genome where chromatin accessibility is lost after BAF inhibition, I could then identify properties of these regions (for example, specific histone modifications and TFs bound there) that distinguish them from regions that do not lose chromatin accessible after BAF inhibition.

2.2: Results

2.2.1: Primer pairs and qualitative PCR assay successfully measure signal to noise ratio of ATAC-Seq Libraries.

I began this project by developing primer pairs that would amplify regions of known accessibility and inaccessibility and could later be used to test the quality of ATAC-seq libraries. Primer pairs consist of two short sequences of nucleotides at the beginning (forward primer) and end (reverse primer) of the fragment of DNA intended for replication. To determine the amplicons to develop my primers around, I identified the housekeeping genes, RPS20, B2M, and GAPDH. These housekeeping genes are found commonly across various tissue types and therefore, would be present in any cell-line of my choice for future experiments.

Next, I wanted to identify regions around these genes that we know to be nucleosome-bound and nucleosome-free. For each of these house-keeping genes, I developed a positive and negative control in areas of the promoter's accessible and inaccessible chromatin nearby. I utilized the UCSC Genome Browser database of "open" and "closed" chromatin in the human genome to determine the primer pairs's 100-200 base pairs amplicon (Kent et al., 2002). The positive control regions were chosen in the genes' promoter peak area of accessibility. The promoter region is where TFs bind in order to initiate transcription and therefore, is a strong basis for chromatin clear of nucleosomes. Whereas, two negative control regions were selected in areas of inaccessibility ~15,000 base pairs upstream and downstream from the promoter peak.

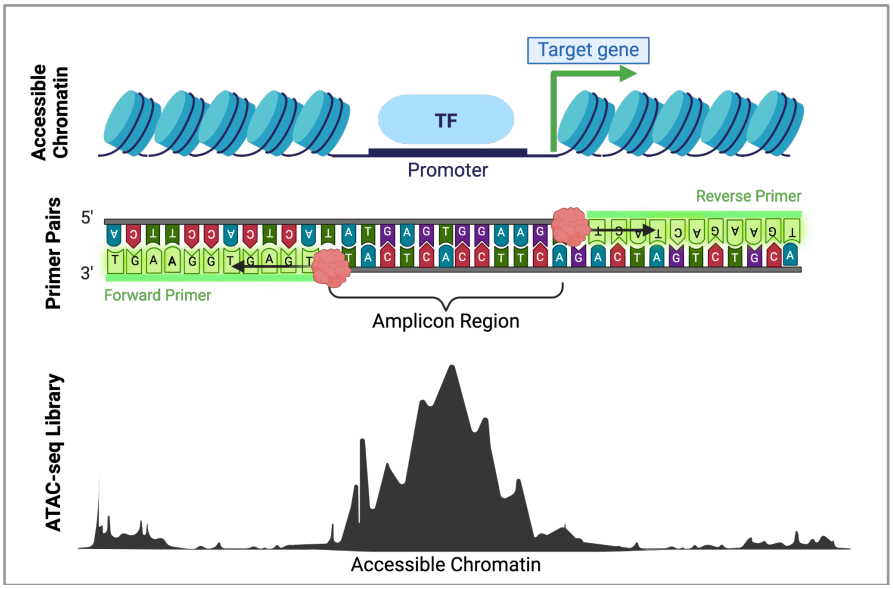

Figure 1. A positive control primer pair, amplifies the known region of accessibility at the target gene's promoter. The designed primers could then undergo qPCR with ATAC-seq libraries to amplify regions of known accessible chromatin. The signal from these regions of accessibility are displayed as a peak.

This resulted in six primer pairs: one positive and one negative control for each of the three genes. To determine their ability to amplify isolated regions, I performed a pCR. The gel electrophoresis bands for RPS20, GAPDH, and B2M were strong and did not contain smears/traces of other DNA bands. This allowed me to conclude that the primer pairs successfully amplify the target region of the same size and in isolation.

Next, I wanted to determine if the positive and negative primers would allow me to determine fold change enrichment or "signal to noise ratio" of a sample. Signal to noise ratio refers to the ability of a library to distinguish the specific DNA sequence that you want to amplify, from the "noise" of any nonspecific amplification or background signal. Following qPCR with these primers, the signal to noise ratio of a library is calculated by subtracting the CT of the positive primer by the CT of the negative primer and squaring this value. A high signal-to-noise ratio indicates that the desired DNA sequence is being amplified efficiently with no trace amount of nonspecific amplicons.

To determine the primer's effectiveness in measuring signal to noise ratio of ATAC-seq libraries, the samples underwent qPCR. The ATAC-seq libraries utilized as the template DNA were previously made using the GM12878 cell line. Our expected results was that they would accurately detect the target accessible DNA sequences amidst background noise (i.e. have a high signal to noise ratio). The larger the signal to noise ratio, the more effective the library is at predicting accessible chromatin areas.

Following qPCR with ATAC-seq libraries, the primer pairs demonstrated high signal to noise ratios. This indicates that the primer pairs can help strongly differentiate between open, nucleosome-free regions versus closed, nucleosome bound regions. The successful creation of positive and negative primer control groups could then be used to verify future ATAC-seq libraries' ability to amplify accessible DNA. This helps determine the libraries' quality before sending them off for high-cost protocols such as sequencing.

2.2.2: ATAC-seq Libraries from K-562 cell-line have high signal to noise ratio.

To quantify and analyze my data, I used the process of Assay for Transposase Accessible Chromatin using Sequencing (ATAC-seq) to identify the cREs that lose chromatin accessibility when the BAF complex is inhibited. ATAC-seq utilizes a protein known as Tn5 Transposase to cleave onto accessible chromatin and insert adapters that tag these fragments (Klemm et al., 2019). These tags help generate a library with peaks that identify regions of accessible chromatin devoid of packaging proteins. A given cell type can have 100,000 or more regions of accessible chromatin, each reflecting a potential cRE. Performing ATAC-seq in the presence and absence of BAF inhibition will allow us to map out which cREs depend on BAF (Klemm et al., 2019).

I first generated three ATAC-seq libraries from different aliquots of K-562 cells. However, prior to performing Illumina sequencing on ATAC-seq libraries, which can be very expensive, I used my primers to ensure that my ATAC-seq libraries contained the expected enrichment of fragments at accessible chromatin regions relative to non-accessible regions of the genome. I refer to this qPCR signal at gene promoters relative to flanking regions at non-accessible regions as "signal-to-noise" ratio.

Then, I ran a qPCR using the designed housekeeping gene primers that were found to successfully measure signal to noise ratio of ATAC-Seq Libraries. As a negative control sample, a genomic DNA library that had not undergone ATAC-seq was also included in the qPCR and data analysis. This genomic DNA library is expected to have roughly equal representation of fragments across the gnome with little-to-no enrichment of fragments at accessible chromatin regions related to non-accessible chromatin regions. As expected, the results of the K-562 libraries showed a larger signal-to-noise ratio in areas of accessible chromatin as opposed to the genomic DNA control.

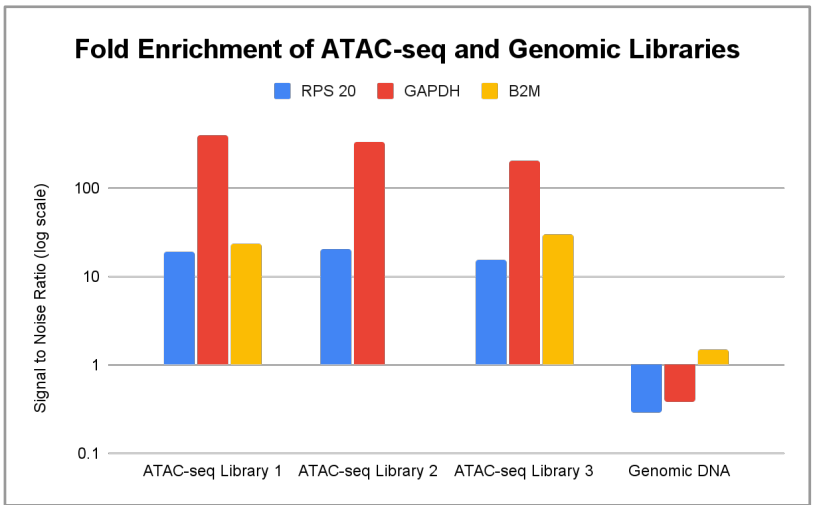

Figure 2. ATAC-seq libraries show expected enrichment of fragments at promoters of housekeeping genes by qPCR. The three K562 ATAC-seq libraries are properly enriched for fragments at gene promoters expected to be accessible, and not at surrounding regions expected to be inaccessible. Also as expected, Genomic DNA does not produce a high signal to noise ratio to identify accessible versus inaccessible DNA.

2.2.3: Identifying optimal concentration to treat K-562 with BAF inhibiting drug.

Once the ATAC-seq protocol was mastered as a means to measure chromatin accessibility across a genome, I wanted to understand how these accessibility patterns changed in response to BAF inhibition. To accomplish this, I aimed to observe the alterations in chromatin organization before and after inhibiting BAF function with the drug BRM014. Upon applying BRM014, I could interfere with BAF function and investigate its impact on chromatin accessibility. However, before beginning drug treatment, I needed to determine the maximum dose of the inhibitor we could use without leading to loss of viability. To establish the most effective dose for experimental treatment, I underwent a dose escalation.

Twenty-four hours after treating K-562 cells with varying concentrations of BRM014 drug dosages, the 100 nM and 10 nM wells were the highest treatment groups that did not exhibit cell death. After an additional forty-eight hours, the 100 nM and 10 nM wells continued to be the highest drug dosages that did not cause significant cell death. I discerned these concentrations, 100 nM and 10 nM, as the maximum doses that maintained cell viability without compromising viability.

By selecting these concentrations, I could confidently assert that the cells were being treated with the most potent levels of the BAF inhibiting drug, and therefore would likely possess inhibited BAF activity. However, the results of the drug dosage also ensured that the treated cells remained viable after treatment. These concentrations are optimal for future experiments because they ensure the highest levels of BAF inhibition while preserving cell health for experimental procedures following drug treatment.

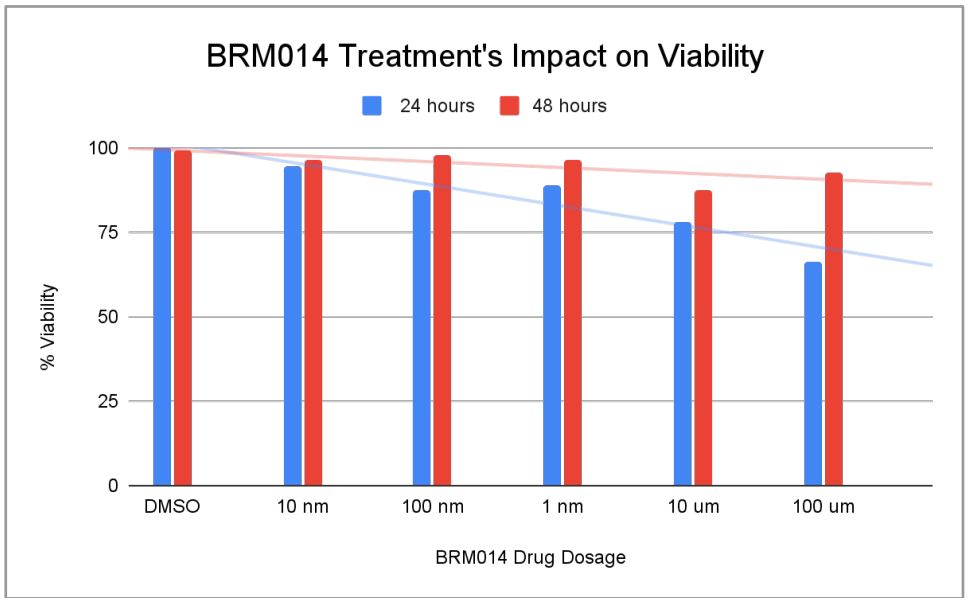

Figure 3. The 10 nM and 100 nM BRM014 dosage groups are optimal for future experimentation. K-562 cells were seeded in a 12-well plate, treated with varying concentrations of the ATPase inhibitor, and monitored over a 48-hour period. The groups treated with a concentration of 10 nM and 100 nM demonstrated the highest preservation of cell viability while still ensuring BAF inhibition drug effects.

2.2.5: Surprisingly, BAF Inhibition does not lead to loss of chromatin accessibility.

After demonstrating that I could successfully perform ATAC-seq on K-562 cells (as described in section 2.2.2), and identifying the maximum sub-lethal dosage of BRM014 to treat my cells with (as described in section 2.2.3), I moved forward with the full drug treatment experiment to inhibit BAF followed by ATAC-seq to measure changes in chromatin accessibility (Figure 4).

I performed ATAC-seq on four treatment groups of K-562 cells. The first, was an experimental group of cells treated with 10 nM BRM014, then a second experimental group of cells treated with 100 nM BRM014. I decided to test two doses of drug, as seen in previous studies, to increase the chances that we would include a dose that impaired BAF function but did not make the cells acutely sick. Next, I created two control groups of cells, one of which was treated with the DMSO vehicle that BRM014 was previously diluted in for groups 1 and 2. The other control and final group was not treated at all. Each of the outlined groups were collected

and assayed by ATAC-seq in two separate biological replicates. These groups will be referred to as No Tx (rep1 and rep2), DMSO (rep1 and rep2), 10 nm (rep1 and rep2), and 100 nm (rep1 and rep2).

The ATAC-seq libraries were sequenced by the Novogene commercial service. Then, I was assisted by lab members Alek Gulka and Dr. Gorkin to align the reads in the form of FASTQ files to the hg38 human reference genome. This allowed for the creation of peak calls (bed files) and signal tracks (bigWig files) for all four treatment groups and their replicates. The bigWig files contain a continuous signal track across the genome reflecting the number of reads mapping to each region. The bed files can be uploaded to the UCSC Genome Browser to visualize signals from accessible chromatin accessibility signals as peaks.

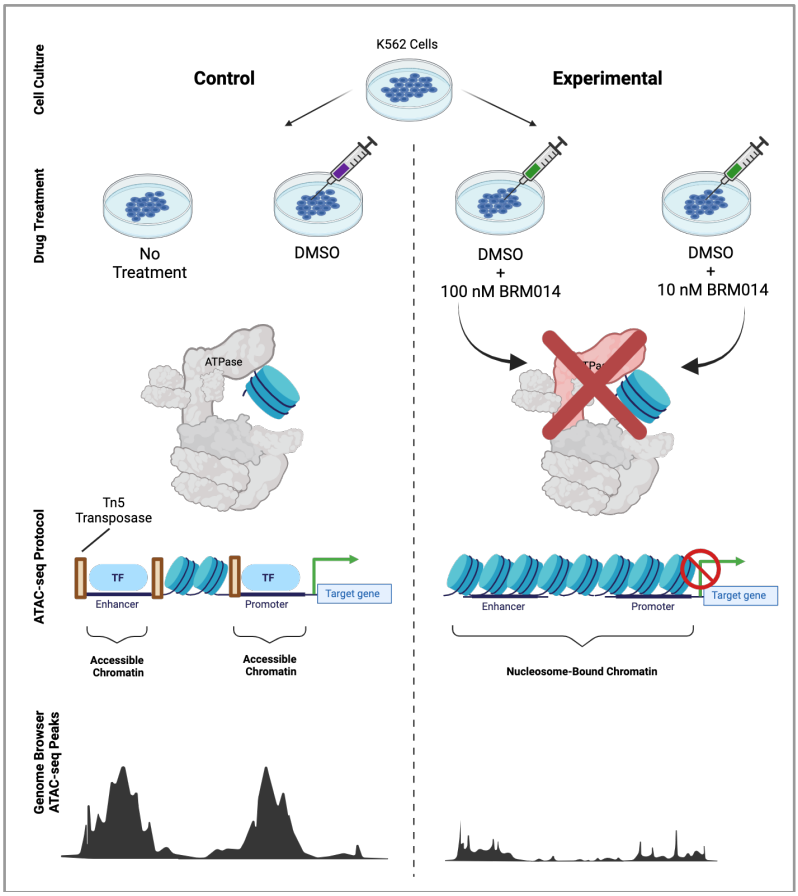

Figure 4. Once cells undergo BRM014 treatment and ATAC-seq protocol, the Genome Browser displays accessibility signals as peaks. We anticipate groups with inhibited BAF to exhibit a decrease in accessibility peaks.

For deeper downstream analysis, the bigWig files were advantageous to look at the effects of BRM014 treatment on chromatin accessibility in specific regions. Each replicate of the treatment groups possesses a corresponding bigWig file that was studied with multiple computational tools. Additionally, all of the peaks across the genome (merged peaks) for each sample were divided into signals far from a TSS (enhancer regions) and signals close to a TSS (promoter regions). The merged peaks within 1,000 bases of a known TSS are also referred to as TSS Proximal, and those that are not within 1,000 bases of a TSS are TSS Distal. I compared these bigWig files containing ATAC-seq signals (a measure of chromatin accessibility) across the genome for each sample. Through the analysis of different types of cRE regions, I aimed to investigate whether BAF is more involved with specific gene categories over others to influence chromatin accessibility.

To visualize and compare the impacts of BAF inhibition based on the bigWig files and on varying types of cRE regions, I plotted the data into Heatmaps. Heatmaps are utilized to illustrate accessible regions within the genome through the depiction of colors. The intensity of these colors on the heatmap correlates with heightened accessibility, facilitating the clear visualization of patterns and trends. When using the plotHeatmap command, I specify the number of "clusters" or groups with similar data that I wish to showcase on the heatmap. This feature highlights patterns within subgroups that may not exhibit the anticipated changes in accessibility.

Upon uploading the initial bed files onto the UCSC Genome Browser, we anticipated observing distinct signal variations across treatment groups, as seen previously in the Gorkin lab with Alex Gulka's work on GM12878 cells. These cells underwent the same methods of BRM014 treatment and ATAC-seq library generation with a DMSO control and treatment group

of a slightly higher 10 uM concentration of BRM014 after 24 hours. The GM12878 cells displayed a reduction in signal upon BRM014 treatment.

The GM12878 cells' ATAC-seq peaks visible on the Genome Browser displayed a reduction in accessibility signal when BAF was inhibited. This aligns with previous work on other cell-lines treated with BRM014. As such, I expected to see similar differences in the accessibility peaks of my four K-562 treatment groups. However, upon initial exploration of the browser interface containing the K-562 experimental results, there was a lack of discernible changes in accessibility. Through further analysis, as shown in the Figure 7 heatmap and Figure 8 USCS Genome Browser clip, I did not observe any severe changes in accessibility.

I surprisingly saw one cluster (Cluster 5) that appears to be increasing in signal. Later on, regions within this cluster underwent closer inspection to determine whether it was a dependable cluster of differentiated signal or if it was an isolated artifact stemming from noisy accessibility signals. These regions showed inconsistency across the 10 nM Treatment Replicate 1, and peaks that were being called as a gain in accessibility were not showing a dramatic increase signal as the heat map led us to believe. Based on this deeper analysis, I confidently categorized the discrepancy in signal as noisy peaks and not a legitimate difference in accessibility.

Similarly, in the TSS distal (enhancer) heatmap (Figure 9), the 10 nM treatment showed a loss in accessibility (Cluster 4) and regain in accessibility (Cluster 5) in both replicates. However, after repeated methods to study the cluster regions, these peaks were too noisy to draw a conclusion of increased or decreased accessibility. Finally, in the heat map showing TSS proximal (promoter) regions (Figure 10), there are no differences in chromatin accessibility across treatment groups. It is particularly surprising that BAF inhibition in the K-562 libraries

did not lead to loss of chromatin accessibility. When ATAC-seq libraries of the GM12878 cell line were treated with the same batch of the BRM014 drug, there was a clear loss in accessibility.

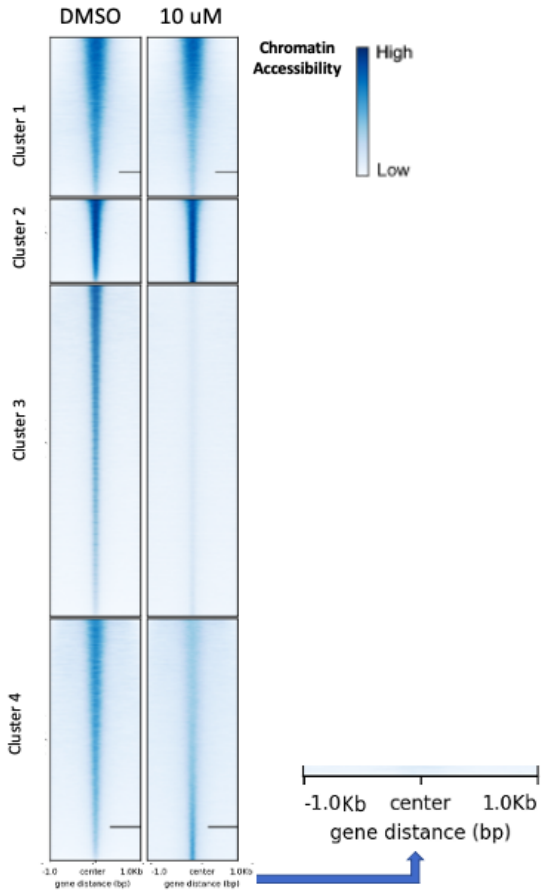

Figure 5. The GM12878 cell line displays a distinguishable loss in chromatin accessibility after treatment with 10 uM of BRM014. This heat map was developed by Alex Gulka of the Gorkin Lab utilizing the GM12878 cell line. Through the same methods of BRM014 treatment (although slightly higher concentration) and ATAC-seq library generation, the control (DMSO) and treatment group (10 um after 24 hours of dosage) could be compared for changes in accessibility. There was a noticeable loss in chromatin accessibility when BAF was inhibited.

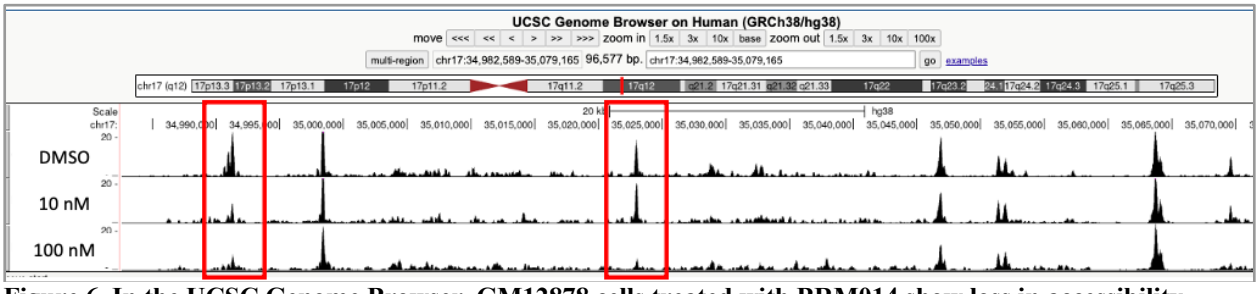

Figure 6. In the UCSC Genome Browser, GM12878 cells treated with BRM014 show loss in accessibility peaks when compared to the DMSO control. The decrease in signal in both doses of BRM014 are a result of BAF being properly inhibited and no longer clearing nucleosomes.

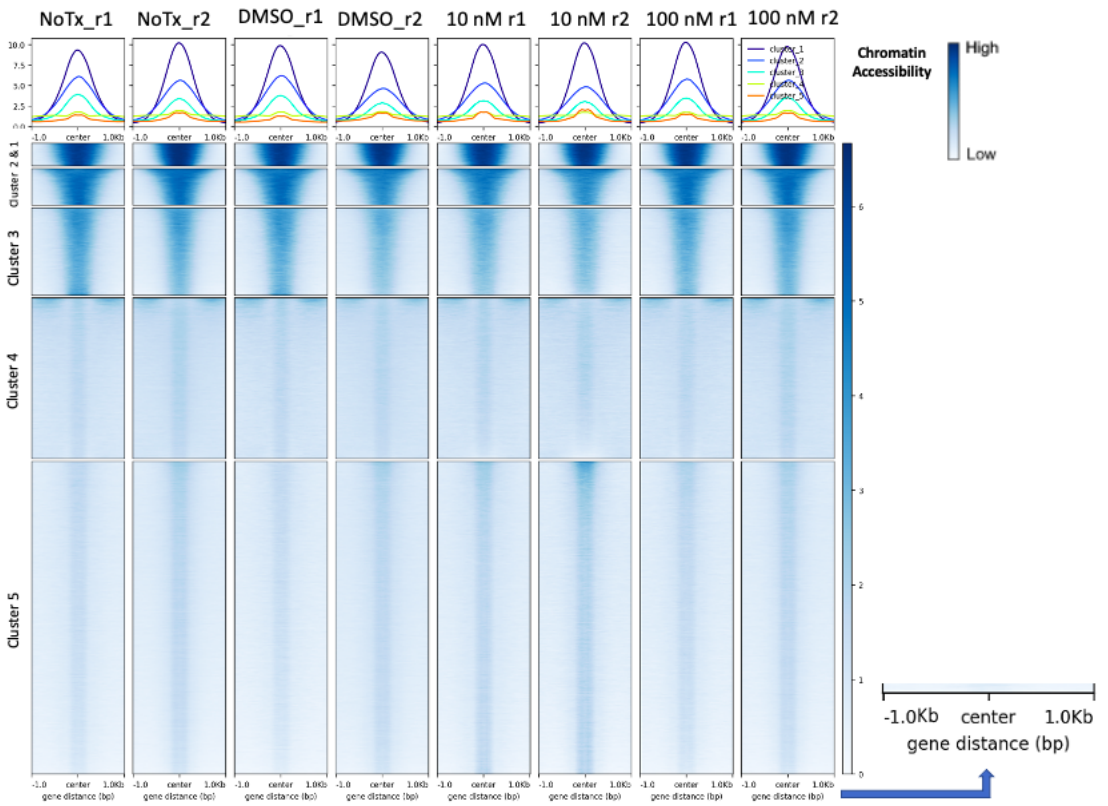

Figure 7. The heatmap of the Merged Peaks bigWig shows no variation in chromatin accessibility between treatment groups. This heat map was developed with the deeptools plotheatmap command. Each treatment group's bigwig files were provided as scores for the merged peaks bed file region. The graph shows no changes in chromatin accessibility except slightly in Cluster 5 that was later identified as noisy peaks.

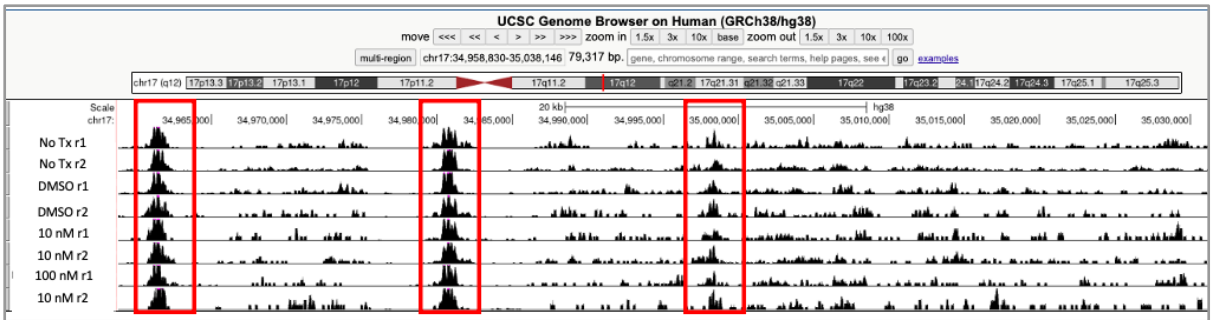

Figure 8. In the UCSC Genome Browser, K-562 cells do not display any significant changes in accessibility peaks when treated with BRM014. This is surprising given the change in accessibility upon BAF inhibition in the GM12878 data and our known understanding of how BAF alters chromatin accessibility.

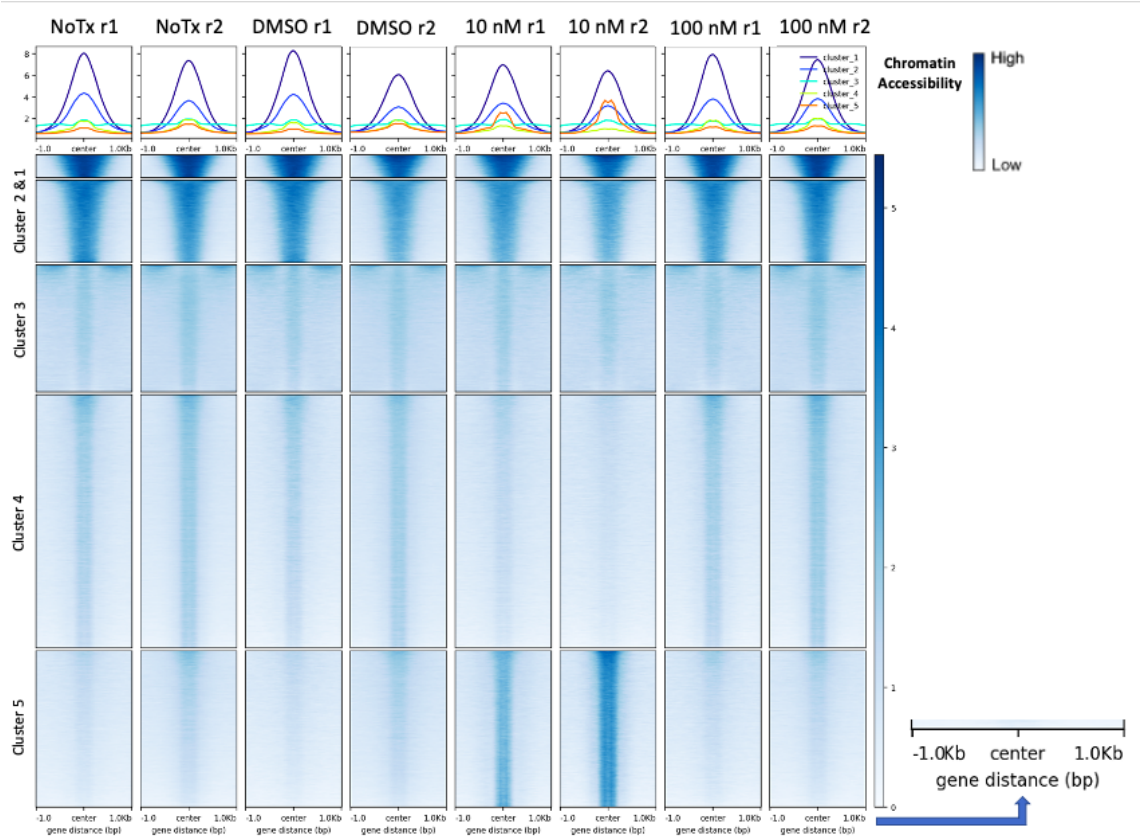

Figure 9. The Heatmap of the TSS Distal Peaks shows no significant variation in chromatin accessibility among treatment groups. This heatmap was developed with the deeptools plotheatmap command. All eight treatment group's bigwig files were provided as scores for the TSS distal file regions. These are peaks that are not within 1,000 bases of a TSS. The graph shows no changes in chromatin accessibility except slightly in Cluster 5 that was later identified as noisy peaks.

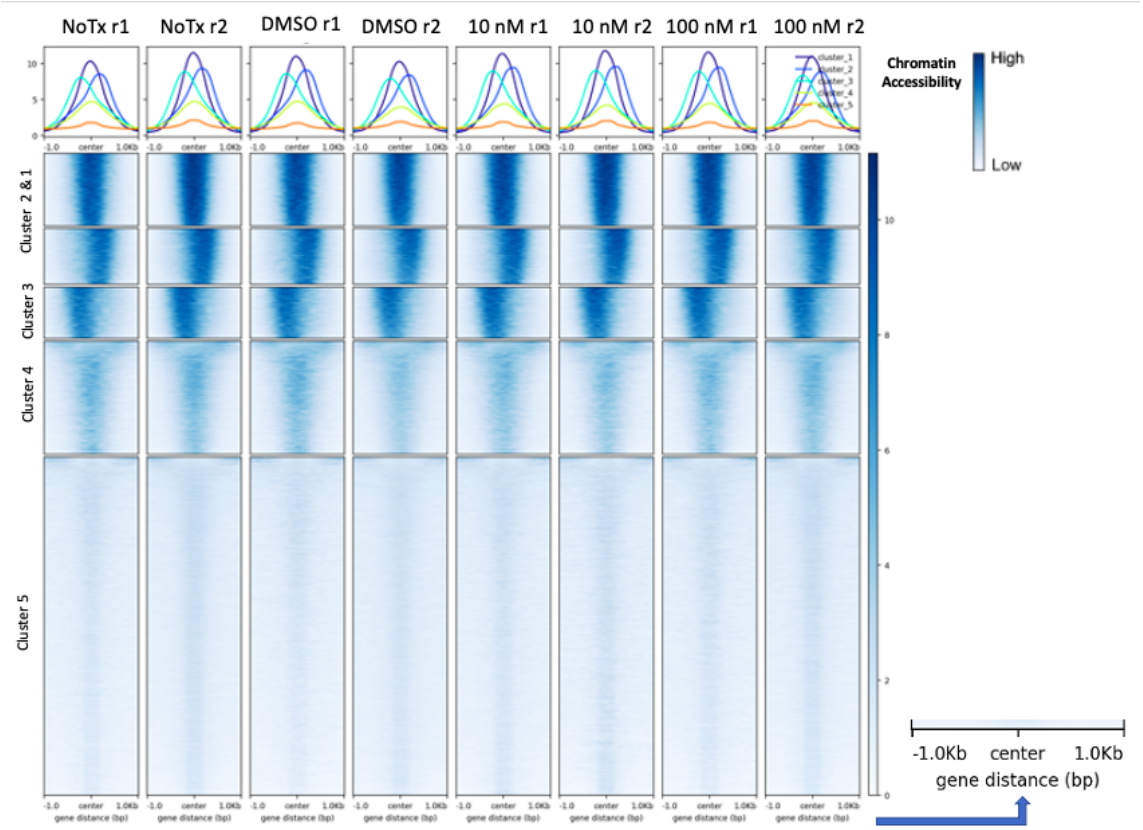

Figure 10. The heatmap of TSS Proximal regions shows no changes in chromatin accessibility. This heat map was developed with the deeptools plotheatmap command. All eight treatment group's bigwig files were provided as scores for the TSS proximal file regions. These are peaks that are within 1,000 bases of a TSS. The graph shows no changes in chromatin accessibility. (The cluster shifts are not relevant or indicative of any differences in accessibility.)

2.3: Methods

2.3.1: Primer Design

I was able to produce primer pairs to replicate these regions by inputting the sequencing of each housekeeping genes' positive and negative control regions I wanted to amplify, into the Primer3 program (Koressaar & Remm, 2007). This process was repeated for each housekeeping gene, to produce three total positive primer pairs and six total negative primer pairs. The Primer3 software produces accurate primer pairs that account for melting temperature to avoid denaturation (Koressaar & Remm, 2007). This is an important consideration for the pairs that will later be placed in a high temperature qPCR environment. I then ran the outputted forward

and reverse primers into the UCSC In-Silico PCR software to confirm they will replicate the desired amplicon and no others.

2.3.2: pCR

The polymerase chain reaction (PCR) method involves reading and amplifying a specific DNA region using forward and reverse primers. Following the Thermo Fisher protocol, a mixture containing primer pairs, template DNA, water, and Thermo Scientific PCR Master Mix was loaded into the Bio-Rad C1000 Touch Thermal Cycler (Thermo Fisher Scientific, 4444556; Bio-Rad, 1851148). In this precisely controlled temperature environment, the mixture underwent 35 cycles of denaturation, annealing, and elongation. In the heat of denaturation, the template DNA fragments and forward and reverse primers attach, annealing to their complementary sequences. Subsequently, components such as DNA polymerase within the PCR Master Mix facilitate the extension of DNA, synthesizing the sequence between the primers. With each cycle, the DNA quantity in the targeted region doubles. Upon completion of the 35 cycles, a highly concentrated sample of the original target sequence is produced.

2.3.3: Gel Electrophoresis

Gel Electrophoresis verifies the specificity of both positive and negative primers in amplifying distinct DNA regions. The gel electrophoresis box generates an electric current that when run through a gel containing naturally negative-DNA, causes it to migrate towards the positively charged electrode (Bio-Rad, 1703930). When current is applied to a gel containing naturally negatively charged DNA, it induces migration towards the positively charged electrode. Larger, heavier DNA segments migrate at a slower rate than smaller, lighter sequences, creating a vertical gradient of longer to shorter DNA segments across the gel. Evaluating the

experimental strands alongside molecular weight markers allows for the determination of region size, quality, and purity.

The agarose gel mixture was prepared by adding agarose powder to a 200 mL solution of water and 1x working solution of Thermo Fisher 10X TBE Buffer (Tris-borate-EDTA) (Research Products International, A20080-100.0; Thermo Fisher Scientific, 75800-950). This mixture was heated and poured into a gel mold equipped with combs to create wells for DNA insertion. SYOR Gel Staining (ApexBio, A8743) was added to enhance the visibility of DNA bands in the solution.

After solidification of the gel, it is placed in the electrophoresis box containing 1x TBE Solution. Before beginning the electric current, the PCR DNA containing the primer pairs' amplicons and a comparative DNA ladder are loaded into the wells. Once a voltage of 150 volts is applied for 35 minutes, the gel is transferred to an imaging system for analysis. This was done to all samples of housekeeping gene primers to evaluate their efficiency at amplifying only the specific region’s length with no impurities.

2.3.4: qPCR

Quantitative PCR, also known as real-time PCR, monitors DNA amplification in real-time by measuring the increase in fluorescence. As the primers’ target sequence is amplified, the amount of starting material with fluorescence doubles, and the intensity increases. The progression of the sample's reaction is evaluated through the Cycle Threshold (CT) value, indicating the cycle number at which the fluorescence signal crosses a predetermined threshold. The output amplification graph of a complete 40-cycle qPCR has a CT value on the X-axis and fluorescence intensity on the Y-axis.

This project will determine the success of ATAC-seq through establishing signal to noise ratio. Following qPCR, the signal to noise ratio of a library is calculated by subtracting the CT of the positive primer by the CT of the negative primer and squaring this value. The larger the number, the more effective the library is at predicting accessible chromatin areas.
Before initiating qPCR, the DNA template underwent serial dilutions of 1:10, 1:100, 1:1000, and 1:10000. I Anticipated that samples with lower dilutions of DNA would require more cycles to reach the CT value. If observed, this would suggest the reliability of the primers for testing the signal-to-noise ratio.

The reverse and forward primers of each gene's positive or negative pairs were loaded into a 96-well Fast Plate based on their diluted amounts, each well containing three replicates (refer to the image for guidance). Similarly to PCR, TaqMan™ Fast Advanced Master Mix, DNA Template, and water were added to facilitate the continual production of the sequence of interest (Thermo Fisher Scientific, 4444556). The No Treatment Control (NTC) wells lack DNA templates and, if the experiment is accurate, expected to show no amplification.

Following quantitative PCR, Qubit dsDNA Assay Kits and Agilent D1000 ScreenTape System Quick Guides can be utilized for additional purification and analysis of the amplified DNA. The Qubit assay yields a calculated concentration of double-stranded DNA in the sample relative to other reagents (Thermo Scientific, Q32851). Whereas, the Agilent D1000 Screen Tape assesses the quality and size distribution of the library by separating fragments based on their size (Agilent Technologies, 5067-5584).

2.3.5: Drug Inhibitor Dose Escalation

K-562 cells were seeded in a 12-well plate, treated with varying concentrations of the ATPase inhibitor, and monitored over a 48-hour period. After a passage, 200,000 K-562 cells were

seeded in a twelve-well plate for incubation. The BRM014 drug, dissolved in a 10 mM stock, was diluted to concentrations of 100 μM, 10 μM, 100 nM, and 10 nM using Dimethyl sulfoxide (DMSO) buffer (MedChem Express, HY-119374; Fisher Scientific, BP231100). Following a 24-hour incubation, 10 μL of the various drug concentrations were added. Each treatment was conducted in triplicate to ensure the reproducibility of effects. This created nine wells of experimental treatments and three wells that served as the No Treatment Control. NTC wells were treated solely with a vehicle composed of the solvent used for delivering the inhibitor (DMSO), and not the actual inhibitor itself.

After an additional 24-hour incubation, the plates were gently mixed and homogenized in preparation for the standard counting procedure. The percent viability was calculated by dividing the number of live cells by the total cell count sum of both live and dead cells.

2.3.6: Cell Culture and Treatment

The K-562 cell line is an immortalized leukemia-derived cell line that is non-adherent. As suspension cells, they grow best free floating in a liquid medium. The K-562 media is composed of 10% Fetal Bovine Serum, 1% Penicillin Streptomycin, and 89% RPMI 1460 (Fisher Scientific 10-082-147; 15-140-122; A1049101). Cells are cultured in 20 mL of this medium and kept in T75 flasks, incubated at 37°C for 48 hours.

The cells thrive at a concentration of 200,000 cells per mL and should be split when they reach 4 million in the flask. This necessitates a “passaging” process approximately every 48 hours. During a passage, the cells are spun down and the original media is removed. During a passage, cells are spun down, the original media is removed, and approximately 200,000 cells are transferred into a new flask with 20 mL of fresh media. They are then incubated until the next passage.

To administer the BRM014 drug and DMSO vehicle, 200,000 K-562 cells were seeded in a six-well plate for incubation. After 24 hours, two wells received the DMSO control, two wells with a 100 nM concentration of the inhibitor, and the remaining two wells were treated with a 100 nM concentration of the drug. After an additional 24 of incubation, all samples were snap-freezed through immersion in liquid nitrogen (Thermo Scientific, 1222Y07). This step is essential to halt cellular processes, enabling later experiments to capture the immediate effects of the BAF inhibitor on the K-562 cells.

2.3.7: ATAC-Seq

The ATAC-seq protocol for K-562 cells was adapted from Buenrostro et al. (2015) to efficiently permeabilize nuclei and tag accessible chromatin. The process begins with nuclei preparation using a permeabilization buffer containing BSA (Sigma, A7906), IGEPAL (Sigma, 18896), DTT (Sigma, D9779), PBS (Thermo Fisher, 10010-23), and complete EDTA-free protease inhibitor (Roche, 05056489001). After isolating the nuclei, DNA is prepared for Tn5 tagmentation using a buffer comprising Tris acetate (Thermo Fisher Scientific, BP-152), K-acetate (Sigma, P5708), Mg-acetate (Sigma, M2545), DMF (EMD Millipore, DX1730), and water (Corning, 46000-CM).

Finally, the samples undergo PCR amplification of tagmented DNA fragments and MiniElute PCR purification to ensure that the end product consists solely of DNA (NEB, M05421; Qiagen, 28004). Quality checks, such as concentration analysis through the Qubit dsDNA HS Assay Kit or library size distribution using 4200 TapeStation, were then performed (Thermo Scientific, Q32851; Agilent Technologies, 5067-5584). At the end of a successful protocol, I have generated six ATAC-seq libraries (two replicates for each treatment group), each ready for the next sequencing step.

2.3.8: Next-Generation Sequencing and Analysis

Before sending the ATAC-seq libraries for genome-wide sequencing, initial iSeq testing was undergone to confirm that the libraries accurately identify regions of accessible chromatin. The sequencer scans at transcription start sites (TSS) where gene expression begins and chromatin are generally clear of nucleosomes. By showing high signal to noise ratio in these areas of known accessibility, the iSeq 100 Sequencing System validated the quality of the libraries (Illumina, 20021532). Then, libraries were pooled and normalized to a concentration of 1 ng/μL to be sequenced by Novogene. To decipher the data provided by Novogen Whole Genome Sequencing, it was put through the ATAC-seq pipeline for ENCODE data (Yan et al., 2020). Each step of this processing pipeline assesses, screens, and filters the raw sequencing reads to generate peak calls (bed files) and signal tracks (bigWig files).

The computational tools applied throughout this project are from the deepTools suite used for exploring sequencing data (Ramirez et al., 2016). The first command, computeMatrix, creates an intermediate file that can be used for the plotHeatmap command. For this command, all treatment group's bigwig files were provided as scores for each bed file region (merged peaks, TSS dist, and TSS prox). This created three separate files that when inserted into the plotHeatmap command, create heatmaps for all treatments in the specified bed region.

I proceeded to analyze clustered regions exhibiting changes in accessibility by filtering for DNA within these regions using command line tools. By generating a bed file of these areas, I could input them into the UCSC Genome Browser to verify if they represented genuine alterations or were merely noisy peak calls. Ultimately, the heatmaps I generated serve as a tool for comparing and evaluating changes in accessibility specifically influenced by the BRM014 treatments.

2.4: Conclusions

2.4.1: The inhibition of the BAF complex in the K-562 cell line does not result in significant changes in accessibility.

The culminating results and data analysis of these aforementioned experiments lead us to conclude that K-562 cells' accessibility is not dramatically altered when treated with the BAF-inhibiting BRM014 drug. This is demonstrated by a lack of significant differences in all treatment groups' heat maps where intense signal signifies areas of open chromatin in the ATAC-seq libraries. This is true for the merged peak regions of DNA as well as in the areas where enhancers and promoters are expected to be located. Regardless of the category of cis-regulatory element, there seems to be little impact of BAF loss of function on the nucleosomal organization within K-562, relative to other cell lines tested.

This is a novel finding as we consider the effects of the same batch of the drug on other cell lines. In addition to the work done within the Gorkin lab on GM12878, there are cell types in the published record that lost accessible chromatin when BRM014 was administered. This includes cell lines HAP1, SKMEL5, SBC5, and H1299 where each study reinforces the effectiveness of BRM014 to inhibit BAF and alter nucleosomal organization (Shick et Al, 2021; Iurlaro et al., 2021; Papillon et al., 2018). This leads us to hypothesize that the K562 cell-line might contain a BAF mutation in which either the BRM014 drug does not affect BAF's ability to function or there is an already existent absence of working BAF.

2.4.2: Limitations

There are several limitations to the study that may have impacted its output. This includes the choice of the cell line K-562. Given its cancer-derived qualities, there is a possibility of pre-

existing BAF mutations within its genome. If this is the case, BAF may already be partially dysfunctional in the cell-line, which could affect the sensitivity of K-562 cells to BRM014. Alternatively, K-562 could have other chromatin remodelers able to compensate for BAF inhibition. However, when beginning the study, we did not know how the BRM014 influence on K-562 would differ. Especially considering that previous research has used many cell-lines of a cancer background and seen the changes in accessibility we had expected. With this pre-existing blind spot, and the widely characterized TF binding regions of K-562, we decided that this would be the best vehicle to test our hypothesis.

Another possible limitation is that the drug treatment on the cells was ineffective. Although I performed a dosage escalation to try to identify the optimal dose, it may be that higher doses are needed to cause BAF chromatin remodeling inhibition in K-562. This is a probable reason, as the 10 nM and 100 nM I used to administer BRM014 is ten times less concentrated than the GM12878 studies conducted within the lab and would explain why there were no significant differences between the control and treatment groups' chromatin accessibility. While we can not confirm this for certain, it is important to note that the same batch of the BRM014 inhibitor was used to treat the GM12878 cells. In this experiment, the drug had a reproducible impact on the GM12878 that aligns with our current knowledge of the BAF complex's role to remodel chromatin.

2.4.3: Potential Next Steps

To either verify or discount the results of my work, there are potential experiments that can be employed. To further investigate the impact of the BRM014 drug on ATAC-seq libraries, we can conduct another dose escalation for a longer time period and at higher concentration. It is possible that the K562 cells are regaining BAF function following long periods after drug

treatment. By examining the libraries after longer time periods, we can rule this out as a potential confounding variable. Additionally, administering BRM014 at higher concentrations will provide a broader range of data on how BRM014 impacts accessibility.

Another next step may be to explore other chromatin remodelers that could be compensating for the loss of BAF function in K-562. In a study done by Martin et. Al in 2023, it was demonstrated that remodeling complex EP400 was capable of reestablishing accessibility at promoter cites following BAF-inhibition (Martin et al., 2023). To mirror their approach, this alternate hypothesis can be tested by running a western blot on untreated cells and cells treated with BRM014. Western blots detect distinct proteins in a sample using an antibody that specifically binds to the target protein of interest. By analyzing the increase or decrease in intensity of other BAF proteins' signals, we can interpret changes in their expression. If proteins correlated to other BAF subunits or chromatin remodeling complexes increase in signal when BAF is inhibited, it is possible that it may be taking over the role normally performed by BAF.

Finally, to confirm that nucleosomal organization within the K562 cell line is unchanged by BAF inhibition, we can utilize a PROTAC to physically degrade BAF's ATP subunits. Proteolysis targeting chimeras, or PROTACs, have been shown to more selectively hinder the function of the target proteins (Kargbo, 2020). We can then run a Western Blot to affirm the successful disintegration of the SMARCA2 and SMARCA4 target proteins. Once we are confident that BAF has been inhibited, ATAC-seq can be performed. We can cross-compare the treatment groups with those of the small-molecular inhibitor to discover if K-562 truly possesses abnormalities.

Chapter 3: Identifying Variation in BAF Complex Genes Among Non-Human Primates, Neanderthals, and Homo Sapiens

3.1: Background

3.1.1: The BAF Complex in Brain Development

During human development, chromatin remodeling complexes such as BAF work alongside transcription factors to precisely conduct chromatin organization and gene expression. This process specializes cells into their tissue-specific form, with BAF having a particular link to neural tissue that develops into the central nervous system (CNS) (Sokpor et al., 2017). The growth of the CNS is referred to as neurodevelopment, a process that determines the morphological development of the organism's spinal cord and brain. Mutations in the BAF Complex are more likely than any of the other chromatin remodeler families to result in neurodevelopmental disorders (Sokpor et al., 2017).

One 2000 study underwent gene targeting in mice embryos to produce heterozygous and homozygous null mutants of BAF's SMARCA4 (*BRG1*) ATPase subunit (Bultman et al., 2000). The developing mice with homozygous BAF knockout resulted in early embryo death. Even within the heterozygous treatment groups who survived, these mice experienced higher rates of tumors and nervous system defects. This work underscores the necessity of a properly functioning BAF in gene regulation and embryonic development.

In humans, there are thirteen BAF subunits that are linked to diagnosable human mental disorders, many of which are noted for Coffin-Siris Syndrome (CSS) risk (Valencia et al., 2023). Coffin-Siris is a genetic diagnosis in which patients display underdevelopment of their fingers and toes as well coarse, dysmorphic facial features. Most often, patients who are displaying these

symptoms as well as have a delayed development of speech and experience symptoms of intellectual disabilities (ID) or autism spectrum disorder (ASD) go to a clinical setting for genetic tests. Upon finding a mutation in one of the genes encoding BAF proteins these patients are then diagnosed with CSS. It is important to recognise that the association of these proteins with NDD diagnoses stems from identifiable phenotypic traits that result from improper development when BAF is not functioning correctly.

Given the significant impact that mutations in BAF complex proteins can have on cognitive outcomes, it prompts consideration of potential genetic differences across human evolution. Variations in BAF proteins between modern humans, Neanderthals, or non-human primates could potentially influence neuronal disparities between these species. This study aims to investigate BAF in an evolutionary perspective to determine its influence on the brain development of different species.

3.1.2: The Evolutionary Conservation of the BAF Complex

Evolutionary conservation refers to DNA sequences that remain highly unchanged across human history due to the essential role it may play in organisms' fitness and reproduction. The BAF Complex proteins exemplify this, having been initially identified in yeast and demonstrating remarkable stability over 500 million years of evolution (Kadoch & Crabtree, 2015). Their conservation is compounded by their presence across all eukaryotic organisms, from fruit flies to humans. As my following work will highlight, a similar pattern emerges in Neanderthal and non-human primate groups, with human BAF coding sequences sharing a 99.992% identity with non-human primates and 99.9978% with Neanderthals.

Moreover, studies have revealed evidence of negative selection acting upon the genes encoding BAF subunits. Negative selection, a hallmark of evolutionary conservation, suggests

that mutations in these genes are less tolerated and if found, could have severe deleterious and pathogenic effects (Hodges et al., 2016). The conservation of BAF Complex genes underscores their fundamental importance in various cellular processes and suggests that any changes to them would likely be detrimental to the organism's survival (Hodges et al., 2016).

3.1.3: Single Nucleotide Polymorphisms and Missense Mutations

In the DNA that makes up the human genome, there are an estimated 3 billion base pairs (Collins & Fink, 1995). However, the well characterized human reference is not entirely identical to every single existing individual. This is largely due to single nucleotide polymorphisms (SNPs) or locations in the genome where some individuals have one nucleotide and others have a different nucleotide. There are an estimated 1.4 million SNPs existent in the human genome and most do not impact transcription, translation, and the activity of our genes. In this project, the term "variant" will be used to describe nucleotide differences within the NHP/Neanderthal genome that deviate from the expected/predominant nucleotide sequence in the human reference genome.

In contrast to SNPs, mutations are changes that occur in DNA sequences that are unpredictable or uncommon. Mutations can be a result of a base pair being inserted, deleted, or changed in any location. Sometimes these mutations have no effect on protein coding or gene expression and other times the substitution of nucleotides can result in the incorrect protein being coded for. When the substitution of one amino acid for another occurs as a result of a single nucleotide change, this is called a non-synonymous or missense mutation. This paper will focus specifically on missense mutations and how their presence as variants in the Non-Human Primate and Neanderthal genomes alter the encoded amino acids. In the context of BAF

protein's highly conserved coding regions, existing variants hold particular significance, as they may offer insights into evolutionary divergence of brain development.

3.1.4: Non-Human Primates

Pan Troglodytes (Chimpanzees) and Pan Paniscus (Bonobos) diverged from the common ancestor of modern humans approximately 6 million years ago, making them our closest living relatives and worthwhile point of study. The genomes of chimpanzees and bonobos exhibit a remarkable 99.6% identicality, with chimpanzees sharing 98.8% of corresponding sequences with the human genome, and bonobos sharing 98.7% (Prüfer et al., 2012; Waterson et al., 2005). Despite this high degree of genetic conservation, the small percentage of divergence manifests as three distinct species, each possessing unique anatomical and physiological traits. As such, we expect any existing genetic variation in protein coding genes between humans and their primate counter-parts to be significant drivers of their differences (Prüfer et al., 2012; Waterson et al., 2005).

Such variants are infrequent and require a deep analysis of their impact on protein function. In Waterson's 2005 study, they discovered 12,164 documented disease variants in 1,384 human genes, with only 16 instances where the altered sequence in a disease allele matched the chimpanzee genome (Waterson et al., 2005). Six of these were identified as de novo human mutations associated with Mendelian disorders, such as an allele linked to causing pancreatitis in humans, which might indicate a digestive adaptation within chimpanzees. This study underscored the importance of investigating genetic variants between these groups as a means to fully understand the variability between humans, chimps, and bonobos (Waterson et al., 2005).

In a 2018 project, a team looked into the neuronal development of human, chimp, and bonobo by programming induced pluripotent stem cells and watching their growth into a neuronal network (Marchetto et al., n.d.). They observed significant differences in neuronal maturation and differential gene expression that could be a result of the BAF complex's work in neurodevelopment. The study calls for further investigation of specific genes that may be causing neural dissimilarities. My work will do just that through directly comparing the BAF genes in these Non-Human Primates (NHP) that regulate neuronal cell differentiation (Marchetto et al., n.d.).

3.1.5: Neanderthals

Neanderthals, known formally as Homo Neanderthalensis, are an archaic relative of humans thought to have inhabited the Earth between 30,000-230,000 years ago (Meyer et al., 2012). Noonan et. al's study found that Neanderthal and human genomes are 99.5% identical and identified 171 sites of base-pair substitutions among the 37,636 aligned nucleotides of humans, Neanderthals, and chimpanzees (Noonan et al., 2006). This indicates that merely 0.5% of the Neanderthal genome differs from the genetic makeup shared by humans and chimpanzees at specific positions. As with the non-human primate variants, it is likely that these variants are responsible for the discrepancies between the three species and can provide insight into the evolutionary relationship between them (Meyer et al., 2012; Noonan et al., 2006).

One of the most prominent disparities between Neanderthals and Homo sapiens is in our brains. Fossil evidence tells us that these individuals had a larger brain than modern day humans likely due to a higher growth rate during infant development (Ponce de León et al., 2008). This difference in development between Neanderthals and modern day Homo Sapiens raises the question of larger evolutionary and cognitive implications. As the closest known ancestors to

humans, studying the Neanderthal genome and how it manifests in brain development and diseases, can provide exciting advancement in understanding our own brains (Meyer et al., 2012; Ponce de León et al., 2008).

3.2: Results

3.2.1: Identifying variants in BAF Protein subunits' genes.

I sought to identify genetic variants that might change the protein structure and function of BAF subunits. To do so, I began with a list of all genetic variants between the human and Neanderthal genome (N=1,419,541) as well as those between the human and NHP genome (N= 6,576,748). I then utilize the variant calling files from the Lindo lab to filter this to only variants within protein-coding sequences of the 31 known BAF subunit genes (N=41 for NHP and N=42 for Neanderthal). I focused on protein-coding variants only for this analysis, because these are the most likely to impact protein function, though I note that non-coding variants in splice site or transcriptional regulatory sequences can also alter protein function. With this list of variants I calculated the percent identity, or relative amount of how many amino acids are identical between the species. The BAF Complex has a 99.992% identity between humans and Non-Human Primates and a 99.9978% identity between humans and Neanderthals.

3.2.2: Predicting the impact of genetic variants on BAF protein structure and function.

By assessing the effect of a variant on a protein, I can reveal how likely it is to induce protein malfunction and lead to adverse downstream effects such as genetic diseases or disorders. A harmful mutation in any of the BAF proteins would likely disrupt the functioning of the entire complex. This could subsequently impede BAF's roles in chromatin organization, tumor suppression, and neurodevelopment. To explore the extent to which these alternate base pairs

impact BAF proteins' role, I inputted them into the CADD, SIFT, and PolyPhen softwares (Kircher et al., 2014; Sim et al., 2012; Adzhubei et al., 2010).

The Combined Annotation-Dependent Depletion (CADD) method, as developed by Kircher et al. in 2014, utilizes multiple gene annotations to generate a single measure of genetic variation. This measure, known as the C-score, works as a ranking system, evaluating a variant's relative potential to affect protein structure/function, or to instigate severe disease. A higher C-score indicates a greater likelihood of a variant being deleterious, potentially disrupting protein coding and other essential downstream processes. Variants with scores above 20 are predicted to represent the top 1.0% of the most detrimental substitutions possible within the human genome (Kircher et al., 2014).

CADD is specifically effective at measuring "deleteriousness" by comparing inputted variants to well-characterized, fixed alleles in the human genome. This large-scale data set was assembled using the ENCODE project and UCSC Genome Browser Tracks. With this information, CADD pinpoints variants that have been removed from the human population over time due to natural selection but exist in the genome of interest. These extracted variants with a high selective constraint are tagged as deleterious and therefore can help predict pathogenicity as an, "organismally relevant estimate of variant impact" (Kircher et al., 2014).

The Sorting Intolerant from Tolerant (SIFT) algorithm accounts for common variants and human divergence data by comparing amino acid conservation across species (Sim et al., 2012).. Highly conserved mutations suggest a change in that position might be deleterious. Upon analyzing sequence alignments and calculating conservation scores, SIFT assigns a score ranging from 0 to 1. Genetic variants with scores between 0 and 0.5 are labeled as having a "Damaging" or "Tolerated" effect on the protein (Sim et al., 2012).

PolyPhen-2 is a software tool designed to identify variations in DNA base pairs (SNPs) that may lead to changes in protein translation (Adzhubei et al., 2010). It utilizes sequence-based and structural predictions to analyze the impact of amino acid substitutions on protein structure and function. The tool assigns qualitative assessments (benign, possibly damaging, or damaging) to each variant and provides a confidence score to indicate the reliability of the prediction (Adzhubei et al., 2010). By inputting my list of variants into these three computational tools, I was able to predict their mutational impact on each of their respective BAF proteins and overall function of the BAF complex. The results of these tools are displayed in the table below.

CADD	Chr3: 47629730 (G/A, SMARCC1)	23
	Chr3: 52678736 (A/T, PRBM1)	22.1
	Chr12: 46205253 (G/A, ARID2)	22
	Chr17:38788469 (A/C, SMARCE1)	24.6
PolyPhen	Chr3:47629730 (G/A, SMARCC1)	Possibly Damaging
	Chr17:38788469 (A/C, SMARCE1)	
SIFT	Chr6:157505578, (C/T	Damaging
	Chr17:38788469 (A/C, SMARCE1)	
	Chr17:38786509 (T/G, SMARCE1)	

Figure 11. The CADD, PolyPhen-2, and SIFT softwares outputted scores to delineate the NHP variants that impact protein structure. Of the 41 Non-Human Primate BAF variants: 4 presented a CADD score above 20, 2 were identified by PolyPhen as possibly damaging, and 3 were labeled by SIFT as damaging.

CADD	Chr1:27102188 (A/G, ARID1A)	21
	Chr6:157471861 (C/T, PHF10)	21.1

	Chr6:170115851 (G/A, PHF10)	31
	Chr6:170115857 (A/C, PHF10)	21.1
	Chr12:50483727 (G/A, SMRCD1)	23
	Chr17:61919841 (G/A, SNARCD2)	22.3
	Chr17:61919842 (C/T, SMARCD2)	20.1
	Chr19: 11095955 (G/T, SMARCA4)	43
	Chr19:11095956 (A/T, SMARCA4)	23.7
PolyPhen	Chr6:170115851 (G/A, PHF10)	Possibly Damaging
SIFT	Chr6:157471861 (C/T, PHF10)	Damaging
	Chr6:157507698 (G/C, ARID1B)	
	Chr6: 170115851 (G/A, PHF10)	
	Chr6:170115857 (A/C, PHF10)	
	Chr19: 11095955 (G/T, SMARCA4) *Stop Codon	
	Chr20:60733994 (C/T, SS18L1)	

Figure 12. The CADD, PolyPhen-2, and SIFT softwares outputted scores to delineate the Neanderthal variants that impact protein structure. Among the 42 Neanderthal BAF variants: 9 presented a CADD score above 20, only 1 was identified by PolyPhen as possibly damaging, and 6 were labeled by SIFT as damaging.

3.2.3: Identifying variants predicted to be “Damaging”by all three computational tools.

To delineate the alleles that were most likely to lead to BAF protein complete loss of function, I picked out NHP and Neanderthal variants that were considered “Damaging” by all three computational tools. This gave me a credible variant of interest for each group. The Non-Human Primate variant I located in Chromosome 17 results in the impaired function of the SMARCE1 subunit. Mutations in SMARCE1 are highly associated with Coffin-Siris Syndrome

(Sokpor et al., 2017). The Neanderthal variant I discovered, impacts the PHF10 subunit, whose transcription levels increase along with melanoma progression (Soshnikova et al., 2021).

	Position	Gene	Ref/Alt	Amino Acid Substitution	CADD Score	SIFT Score/Prediction	Polyphen Score/prediction
Non-Human Primate	Chr17: 38788469	SMARCE1	A/C	Valine/ Glycine	24.6	0 Damaging	0.966 Possibly Damaging
Neanderthal	Chr6: 170115851	PHF10	G/A	Arginine/ Tryptophan	31	0 Damaging	0.999 Possibly Damaging

Figure 13. One variant in the Non-Human Primate genome and one variant in the Neanderthal genome were identified as being "Damaging" by CADD, SIFT, and PolyPhen.

3.3.7: Determining variants' presence in the human population.

Once I had established a list of variants present in the Neanderthal and NHP individuals, I next wanted to understand their prevalence in modern day humans. As a dataset of 76,156 human genomes, the Genome Aggregation Database (genomAD) allowed me to predict the frequency of alleles in the broader human population (Chen et al., 2024). I utilized version 2.1.1, as it is mapped to the hg19 reference used consistently throughout my computational analyses. Alongside known gnomAD variants, this database integrates information on ClinVar variants sourced from clinical contexts that have documented genetic variations observed in patients.

This information provides insight into the link between the presence of mutational base pairs and their likelihood to be found in association with a particular phenotype or disease. By uploading the variants in NHP and Neanderthal which have been verified as "Damaging" by CADD, SIFT, and PolyPhen, we can understand whether the variant is observed in the modern-day human population. These tools were utilized in conjunction with the UCSC Genome Browser to validate whether they are common SNPs and would be present an anomaly if found in any individual's sequenced genome. If present, we can conclude this variant does not have an

impact on protein function that is detrimental to an individual's survival. However, if the damaging variant is absent in the human population, it suggests that the protein it affects is crucial. Those with this variant may not survive long enough for their symptoms or genotype to be documented due to the loss of essential protein function.

When I searched for the Non-Human Primate variant (Chr17: 38788469, A/C) in these databases, I found that it lacks known genomAD or ClinVar variants in the human population. Similarly, the UCSC Genome Browser indicates the absence of common SNPs at this specific location Yet, it predicts a severe loss-of-function effect on the SMARCE1 protein if a missense mutation were to occur in this position.

When I did the same to the Neanderthal variant (Chr6: 170115851, G/A), genomAD observed an identical substitution in three individuals within the dataset. It represents a missense mutation with a predicted loss-of-function effect on the PHF10 protein. However, there is no ClinVar data in which this exact exchange has been found in a clinical patient with documented phenotypic effects. According to the UCSC Genome Browser, this base pair is not a common dbSNP and remains consistent throughout the human population.

Through this analysis, both variants were ruled out to be human polymorphic SNPs, indicating their uncommon occurrence in humans. Consequently, these base pairs are highly conserved in humans, and an alternative allele can be considered significant. For the SMARCE1 variant, there is no documentation of a human possessing the Adenine nucleotide in place of the reference Guanine. It is therefore highly likely that individuals who possess the alternate A allele instead of G, are not surviving to the point of being documented. Its presence in Non-Human Primates underscores the importance of further investigation into this specific base pair.

3.3: Methods

3.3.1: Obtaining variants between the human genome and Neanderthal/Non-Human Primate Genomes

Our genomic data on Neanderthals was generated by Matthias Meyer et a.l from bone remains located in the Altai mountains of Siberia (Prüfer et al., 2014). Of the DNA extracted from this over 50,000 year old bone, their team was able to develop a single-stranded library preparation method to generate a genome sequence with a near 30-fold coverage. In order to compare the Neanderthal genome to that of the modern-day human, it was aligned to the hg19 Human Genome Assembly (Prüfer et al., 2014).

I received this data from Dr. John Lindo and his lab, who conducted the processing of the libraries and performed variant calling. This variant calling applies advanced sequencing technology to analyze the distinctions between the Neanderthal genome and hg19 human reference. By leveraging genetic information such as single nucleotide polymorphisms (SNPs) and known mutation sites, the process generates a Variant Calling File (VCF) highlighting bases that differ from the hg19 genome. The VCF file provided by the Lindo lab contains variants in the Altai Neanderthal DNA.

The Non-Human Primate (NHP) data is a high-coverage genome obtained from 13 bonobos (Pan Paniscus) and 58 chimpanzees (Pan troglodytes) (Prado-Martinez et al., 2013). Of the chimpanzees, 18 come from central central populations, 19 eastern, 11 western, and 10 are from Nigeria–Cameroon chimpanzees. It is important to note that this file serves as a reference and not a variant calling file. Therefore, base pairs that differ from the human reference genome may be polymorphic. This means that for any nucleotide at a position in the DNA sequence, there are multiple variations of that nucleotide present in different individuals of the

bonobo/chimpanzee population. With this NHP and Neanderthal data, we aimed to explore divergent nucleotides within BAF coding regions that might explain our modern day differences from this ancestor (Prado-Martinez et al., 2013).

3.3.2: Filtering for Coding Sequences within Exons

To focus only on genes that code for BAF proteins, the human reference genome was filtered for coding sequences within exon regions. To do this, I first downloaded a GRCh37 (hg19) version file of Gencode's Basic Gene Annotation that specifies gene locations and status as start/stop sites, genes, transcripts, exons, untranslated or coding regions. The hg19 version of the Homo Sapien Reference Genome was used intentionally in place of the updated hg38 version because when the Neanderthal and Non-human Primate genomes were created, they were aligned to hg19. While the older version has less coverage of alternate sequence regions, it allows for the project's overall goal to compare modern day human DNA sequences to Neanderthals and Non-Human Primates.

With this annotated file, I was able to filter for genes who were labeled as "CDS" or coding sequences. These are regions of DNA that are read and determine the amino acids to make proteins. We are only interested in these regions of the DNA because mutations within them may code the incorrect protein or insert a premature stop codon. Both of these events occurring in a BAF gene's protein coding sequencing would influence the overall function of a subunit and potentially inhibit the complex's activity altogether.

Next, I accumulated a list of all BAF subunits and their respective Havana gene ID. Again, I filtered the basic gene annotation file to output only the coding sequences in which the gene_ID column possessed one of the thirty-one BAF subunits. This ensured that my final file, "bedtools.BAF.CDS.bed" was a hg19 compilation only of DNA that influences BAF protein

structure. All computational work was performed on the Linux Command Line in the Gorkin Lab server.

3.3.3: Bedtools Intersect

Once I had established a human reference file, my goal was to cross compare this genome to Neanderthals/Non-Human Primates and look for dissimilarities in nucleotides (variants). This process was facilitated by downloading the BEDTools software onto the server. BEDTools provides multiple tools to test the correlation between genomic datasets, most notably for my work being the *intersect* command. The intersect command screens for overlaps between two genomic files and can output the areas of overlap or non-overlap between them (Quinlan & Hall, 2010).

Utilizing BEDTools, I sought to identify Neanderthal and non-human primate variants within regions coding for BAF proteins. By inputting the Neanderthal/non-human primate Variant Call Format (VCF) files and bedtools.BAF.CDS.bed, I directed the command to extract the original Neanderthal/non-human primate VCF entries where overlaps occurred, saving them to separate files. Once these files were generated for both groups, I could directly examine the locations of nucleotide differences and assess their impact on BAF protein function.

3.3.4: CADD

The Combined Annotation-Dependent Depletion (CADD) was used in this project to evaluate the influence of variants found in the Neanderthal and Non-Human Primates' BAF protein coding regions that are not present in modern-day humans (Kircher et al., 2014). Using the list of overlapping variants in the BAF protein regions, CADD was able to delegate a C-score for each variant. To compile a list of relevant variants, only those with a C-score exceeding 20

were extracted, indicative of being among the top 1% most deleterious. This process was conducted separately for Neanderthal and Non-Human Primate variants (Kircher et al., 2014).

3.3.5: SIFT

The Sorting Intolerant from Tolerant (SIFT) algorithm was developed by Sim et. al in 2012 to determine the influence of a missense variant's change in amino acids on the coded protein's structure or function (Sim et al., 2012). In my project, each Neanderthal and NHP variant was submitted into the SIFT web server (http://sift-dna.org) to receive a score on a scale of 0 to 1. Variants that were identified as "Damaging" were considered as variants of interest for further study (Sim et al., 2012)..

3.3.6: PolyPhen-2

The PolyPhen-2 software tool traces differences in base pairs (SNPs) linked to changes in protein translation (Adzhubei et al., 2010). PolyPhen incorporates both sequence-based and structural feature predictions based on the amino acid substitution. This allows it to analyze how structural and chemical variant properties impact proteins' biochemical interaction. As an output, it provides a qualitative assignment (benign, possibly damaging, or damaging) for each variant, along with a confidence score to reflect the reliability of the prediction (Adzhubei et al., 2010).

The software offers two options HumanDiv and HumanVar. HumanVar datasets compile genetic variations within the human population, providing insights into genetic diversity. Werease HumanDiv datasets focus on quantifying genetic differences between populations. The HumanVar datasets setting was chosen for my work to study traits or diseases that may be influenced by genetic variations across human populations. These datasets helped me understand human evolution, and my extracted variants' associations with traits or diseases.

3.4: Conclusions

3.4.1: Eight individuals of the non-human primates analyzed in my study have a genetic variant in SMARCE1 that is predicted to cause NDD in humans.

The CADD, SIFT, PolyPhen, UCSC Genome Browser, and genomAD tools unanimously find that the identified NHP variant would result in the loss of SMARCE1 protein function. SMARCE1 is a crucial and highly conserved protein within the BAF Complex. Its functional loss would significantly affect the BAF Complex's role in neuronal cell differentiation and the development of neural structures. If a human possessed this variant, they would likely show phenotypic traits such as developmental delays and abnormalities in fingers, toes, and facial features that are typical of Coffin-Siris Syndrome.

This finding sheds light on the potential evolutionary differences in brain development among humans, Neanderthals, and other primates. The impact of the identified variant on SMARCE1 may provide insight into the genetic mechanisms underlying neural development. With this information, we can begin to unravel the genetic reasoning for the differing cognitive abilities and neurobiological structures in humans compared to Neanderthals and other primates. In addition, my work extends beyond just understanding the differences; they also offer clues about the similarities and shared developmental pathways among these species. In this project, there is data on the many shared base pairs between the human reference and Non-Human primates. By identifying DNA sequences that lead to similar phenotypic outcomes across species, we can explore conserved genes in brain development that have persisted throughout evolutionary history. This research establishes a list of variants that can be tested in future experiments to deepen our understanding of what makes the human brain distinct while

highlighting the shared genetic foundations that unite us with our primate and Neanderthal relatives.

3.4.2: Limitations

The primate dataset comprises a collection of polymorphisms, exhibiting variability where some individuals harbor this specific variant base-pair while others do not. For instance, among chimpanzees, approximately 16% possess the SMARCE1 variant and within the bonobo population, none possess the SMARCE1 variant. Additionally, this data lacks phenotypic information on the animals carrying these variants. This adds a layer of uncertainty to the analysis as we are unsure if the variant is cause for intellectual or developmental differences in the primate populations as it would be if found in a human genome.

Although the utilized computational tools provide valuable predictions to identify potentially damaging variants, their assessments are not definitive. As such, it is possible that the two discovered variants do not impact protein function as severely as I have predicted with these tools. To ascertain the actual impact of these variants, experimental validation would be necessary. This could involve utilizing CRISPR genome editing technology to introduce the variant into model organisms or cell lines and observe phenotypic changes.

Other confounding variables such as complexity of genetic interactions are also posed. This project focused solely on coding mutations and disregards non-coding mutations or other contributing genetic variations. Finally, limitations arise from the scarcity of available Neanderthal and non-human primate sequences. The Neanderthal genome that we employed has been derived from one individual and there is little additional information that has arised since its generation. There is an additive barrier to gathering Bonobo data due to their endangered status that hinders continued usage for research.

3.4.3: Potential Next Steps

To validate the findings of my project one potential avenue is to reconsider the analysis using a dataset of NHP known variants or one that includes phenotypic information within primate populations. If the SMARCE1 allele is among this non-polymorphic data, we can more definitively conclude that it is of significance. The additional known neurodevelopmental phenotypes can then give a more nuanced understanding of the functional consequences of genetic variants within primate populations.

Upon confirming the variant as non-polymorphic, a final model to test my findings is to integrate variants into neuronal cells and organoids. Through CRISPR genome editing, we could specifically locate Chr17: 38788469 of SMARCE1and Chr6: 170115851 of PHF10 and change the reference nucleotide to the alternate. Then, by observing differences in cellular behavior, development, and neural network formation we could shed light on the functional impact of these variants. To validate the functional significance, we could then use animal models such as mice. Observing phenotypic changes in behavior, cognitive abilities, and brain morphology could provide direct evidence linking genetic variants to neurobiological traits.

Chapter 4: Discussion

This thesis presents a comprehensive investigation into the BAF complex inhibition in K-562 cells and the impact of SMARCE1 variants in Neanderthals and Non-Human Primates, offering profound insights into chromatin dynamics and neural development. My findings from both projects, unveil functional and evolutionary dimensions of BAF's role in disease and evolution.

My examination of BAF complex inhibition in K-562 cells challenges existing assumptions regarding its effect on chromatin accessibility. Surprisingly, treatment with the BAF-inhibiting BRM014 drug fails to significantly alter accessibility in K-562 cells. However, limitations such as the choice of the K-562 cell line and uncertainty about the drug dosage strength underscore the necessity for further investigation. My exposure of these underlying complex regulatory mechanisms pave the way for alternative projects within the Gorkin lab. This work has the power to uncover mechanisms in BAF that can aid in the development of therapeutic interventions intended to enhance drug specificity in BAF inhibition.

My identification of a SMARCE1 variant in non-human primates sheds light on potential evolutionary differences in brain development among humans, Neanderthals, and other primates. The analysis of shared base pairs between human reference and non-human primates underscores both the uniqueness and shared genetic heritage among these species. By identifying DNA sequences that may alter the genetic and phenotypic outcomes across species, we gain deeper insights into the genetic divergence of brain development throughout evolutionary history. Together, the insights from both thesis projects contribute a diverse array of new knowledge on the BRM/BRG1 Associated Factor Complex. They lay the groundwork for future scientific inquiries aiming to explore the vital role of BAF through its loss of function, whether through

ATP-ase inhibition or SMARCE1 protein loss. By bridging functional and evolutionary perspectives, my research paves the way for a deeper understanding of the intricate mechanisms underlying brain development across diverse species.

Works Cited

Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., Kondrashov, A. S., & Sunyaev, S. R. (2010). A method and server for predicting damaging missense mutations. *Nature Methods*, *7*(4), 248–249. https://doi.org/10.1038/nmeth0410-248

Andersson, R., & Sandelin, A. (2020). Determinants of enhancer and promoter activities of regulatory elements. *Nature Reviews Genetics*, *21*(2), Article 2. https://doi.org/10.1038/s41576-019-0173-8

Bögershausen, N., & Wollnik, B. (2018). Mutational Landscapes and Phenotypic Spectrum of SWI/SNF-Related Intellectual Disability Disorders. *Frontiers in Molecular Neuroscience*, *11*, 252. https://doi.org/10.3389/fnmol.2018.00252

Bultman, S., Gebuhr, T., Yee, D., La Mantia, C., Nicholson, J., Gilliam, A., Randazzo, F., Metzger, D., Chambon, P., Crabtree, G., & Magnuson, T. (2000). A Brg1 null mutation in the mouse reveals functional differences among mammalian SWI/SNF complexes. *Molecular Cell*, *6*(6), 1287–1295. https://doi.org/10.1016/s1097-2765(00)00127-1

Chen, S., Francioli, L. C., Goodrich, J. K., Collins, R. L., Kanai, M., Wang, Q., Alföldi, J., Watts, N. A., Vittal, C., Gauthier, L. D., Poterba, T., Wilson, M. W., Tarasova, Y., Phu, W., Grant, R., Yohannes, M. T., Koenig, Z., Farjoun, Y., Banks, E., … Karczewski, K. J. (2024). A genomic mutational constraint map using variation in 76,156 human genomes. *Nature*, *625*(7993), 92–100. https://doi.org/10.1038/s41586-023-06045-0

Collins, F. S., & Fink, L. (1995). The Human Genome Project. *Alcohol Health and Research World*, *19*(3), 190–195.

Hodges, C., Kirkland, J. G., & Crabtree, G. R. (2016). The Many Roles of BAF (mSWI/SNF) and PBAF Complexes in Cancer. *Cold Spring Harbor Perspectives in Medicine*, *6*(8), a026930. https://doi.org/10.1101/cshperspect.a026930

Iurlaro, M., Stadler, M. B., Masoni, F., Jagani, Z., Galli, G. G., & Schübeler, D. (2021). Mammalian SWI/SNF continuously restores local accessibility to chromatin. *Nature Genetics*, *53*(3), 279–287. https://doi.org/10.1038/s41588-020-00768-w

Kadoch, C., & Crabtree, G. R. (2015). Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics. *Science Advances*, *1*(5), e1500447. https://doi.org/10.1126/sciadv.1500447

Kargbo, R. B. (2020). SMARCA2/4 PROTAC for Targeted Protein Degradation and Cancer Therapy. *ACS Medicinal Chemistry Letters*, *11*(10), 1797–1798. https://doi.org/10.1021/acsmedchemlett.0c00347

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., & Haussler, and D. (2002). The Human Genome Browser at UCSC. *Genome Research*, *12*(6), 996–1006. https://doi.org/10.1101/gr.229102

Kircher, M., Witten, D. M., Jain, P., O'Roak, B. J., Cooper, G. M., & Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics*, *46*(3), 310–315. https://doi.org/10.1038/ng.2892

Klemm, S. L., Shipony, Z., & Greenleaf, W. J. (2019). Chromatin accessibility and the regulatory epigenome. *Nature Reviews Genetics*, *20*(4), Article 4. https://doi.org/10.1038/s41576-018-0089-8

Koressaar, T., & Remm, M. (2007). Enhancements and modifications of primer design program Primer3. *Bioinformatics*, *23*(10), 1289–1291. https://doi.org/10.1093/bioinformatics/btm091

Lee, H.-T., Oh, S., Ro, D. H., Yoo, H., & Kwon, Y.-W. (2020). The Key Role of DNA Methylation and Histone Acetylation in Epigenetics of Atherosclerosis. *Journal of Lipid and Atherosclerosis*, *9*(3), 419. https://doi.org/10.12997/jla.2020.9.3.419

Marchetto, M. C., Hrvoj-Mihic, B., Kerman, B. E., Yu, D. X., Vadodaria, K. C., Linker, S. B., Narvaiza, I., Santos, R., Denli, A. M., Mendes, A. P., Oefner, R., Cook, J., McHenry, L., Grasmick, J. M., Heard, K., Fredlender, C., Randolph-Moore, L., Kshirsagar, R., Xenitopoulos, R., … Gage, F. H. (n.d.). Species-specific maturation profiles of human, chimpanzee and bonobo neural cells. *eLife*, *8*, e37527. https://doi.org/10.7554/eLife.37527

Martin, B. J. E., Ablondi, E. F., Goglia, C., Mimoso, C. A., Espinel-Cabrera, P. R., & Adelman, K. (2023). Global identification of SWI/SNF targets reveals compensation by EP400. *Cell*, *186*(24), 5290-5307.e26. https://doi.org/10.1016/j.cell.2023.10.006

Minchin, S., & Lodge, J. (2019). Understanding biochemistry: Structure and function of nucleic acids. *Essays in Biochemistry*, *63*(4), 433–456. https://doi.org/10.1042/EBC20180038

Noonan, J. P., Coop, G., Kudaravalli, S., Smith, D., Krause, J., Alessi, J., Chen, F., Platt, D., Pääbo, S., Pritchard, J. K., & Rubin, E. M. (2006). Sequencing and Analysis of Neanderthal Genomic DNA. *Science (New York, N.Y.)*, *314*(5802), 1113–1118. https://doi.org/10.1126/science.1131412

Papillon, J. P. N., Nakajima, K., Adair, C. D., Hempel, J., Jouk, A. O., Karki, R. G., Mathieu, S., Möbitz, H., Ntaganda, R., Smith, T., Visser, M., Hill, S. E., Hurtado, F. K., Chenail, G., Bhang, H.-E. C., Bric, A., Xiang, K., Bushold, G., Gilbert, T., … Jagani, Z. (2018). Discovery of Orally Active Inhibitors of Brahma Homolog (BRM)/SMARCA2 ATPase Activity for the Treatment of Brahma Related Gene 1 (BRG1)/SMARCA4-Mutant Cancers. *Journal of Medicinal Chemistry*, *61*(22), 10155–10172. https://doi.org/10.1021/acs.jmedchem.8b01318

Prado-Martinez, J., Sudmant, P. H., Kidd, J. M., Li, H., Kelley, J. L., Lorente-Galdos, B., Veeramah, K. R., Woerner, A. E., O'Connor, T. D., Santpere, G., Cagan, A., Theunert, C., Casals, F., Laayouni, H., Munch, K., Hobolth, A., Halager, A. E., Malig, M., Hernandez-Rodriguez, J., … Marques-Bonet, T. (2013). Great ape genetic diversity and population history. *Nature*, *499*(7459), 471–475. https://doi.org/10.1038/nature12228

Prüfer, K., Munch, K., Hellmann, I., Akagi, K., Miller, J. R., Walenz, B., Koren, S., Sutton, G., Kodira, C., Winer, R., Knight, J. R., Mullikin, J. C., Meader, S. J., Ponting, C. P., Lunter, G., Higashino, S., Hobolth, A., Dutheil, J., Karakoç, E., … Pääbo, S. (2012). The bonobo genome compared with the chimpanzee and human genomes. *Nature*, *486*(7404), 527–531. https://doi.org/10.1038/nature11128

Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., Heinze, A., Renaud, G., Sudmant, P. H., de Filippo, C., Li, H., Mallick, S., Dannemann, M., Fu, Q., Kircher, M., Kuhlwilm, M., Lachmann, M., Meyer, M., Ongyerth, M., … Pääbo, S. (2014). The complete genome sequence of a Neandertal from the Altai Mountains. *Nature*, *505*(7481), 43–49. https://doi.org/10.1038/nature12886

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

Ronan, J. L., Wu, W., & Crabtree, G. R. (2013). From neural development to cognition: Unexpected roles for chromatin. *Nature Reviews Genetics*, *14*(5), 347–359. https://doi.org/10.1038/nrg3413

Schick, S., Grosche, S., Kohl, K. E., Drpic, D., Jaeger, M. G., Marella, N. C., Imrichova, H., Lin, J.-M. G., Hofstätter, G., Schuster, M., Rendeiro, A. F., Koren, A., Petronczki, M., Bock, C., Müller, A. C., Winter, G. E., & Kubicek, S. (2021). Acute BAF perturbation causes immediate changes in chromatin accessibility. *Nature Genetics*, *53*(3), 269–278. https://doi.org/10.1038/s41588-021-00777-3

Sim, N.-L., Kumar, P., Hu, J., Henikoff, S., Schneider, G., & Ng, P. C. (2012). SIFT web server: Predicting effects of amino acid substitutions on proteins. *Nucleic Acids Research*, *40*(Web Server issue), W452-457. https://doi.org/10.1093/nar/gks539

Simpson, B., Tupper, C., & Al Aboud, N. M. (2023). Genetics, DNA Packaging. In *StatPearls*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK534207/

Soshnikova, N. V., Tatarskiy, E. V., Tatarskiy, V. V., Klimenko, N. S., Shtil, A. A., Nikiforov, M. A., & Georgieva, S. G. (2021). PHF10 subunit of PBAF complex mediates transcriptional activation by MYC. *Oncogene*, *40*(42), Article 42. https://doi.org/10.1038/s41388-021-01994-0

Valencia, A. M., Sankar, A., van der Sluijs, P. J., Satterstrom, F. K., Fu, J., Talkowski, M. E., Vergano, S. A. S., Santen, G. W. E., & Kadoch, C. (2023). Landscape of mSWI/SNF chromatin remodeling complex perturbations in neurodevelopmental

disorders. *Nature Genetics*, *55*(8), 1400–1412. https://doi.org/10.1038/s41588-023-01451-6

Waterson, R. H., Lander, E. S., Wilson, R. K., & The Chimpanzee Sequencing and Analysis Consortium. (2005). Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature*, *437*(7055), 69–87. https://doi.org/10.1038/nature04072

Yan, F., Powell, D. R., Curtis, D. J., & Wong, N. C. (2020). From reads to insight: A hitchhiker's guide to ATAC-seq data analysis. *Genome Biology*, *21*(1), 22. https://doi.org/10.1186/s13059-020-1929-3