Integration of CTCF Loops, Methylome, and Transcriptome in Differentiating LUHMES as a Model for Imprinting Dynamics of the 15q11-q13 Locus in Human Neurons

By

ORANGEL GUTIERREZ FUGON
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Integrative Genetics and Genomics

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

David J Segal (Chair)

Janine LaSalle

Kyle Fink

Committee in Charge

2023

Acknowledgements

To everyone who believed in me, my mentors, my dad Saul, my tia Nora and my tia Marta.

To my cohort and all the friends I made here, who provided support throughout this experience.

To my rock, my loving sister Yaremi.

But especially to the one person without who's unconditional love I would not be the man I am, my mom Mireya Fugon, R.I.P. I love you and miss you so much.

Contents

Abstract

Human cell line models, including LUHMES, are important for investigating developmental transcriptional dynamics within imprinted regions, particularly the 15q11-q13 Angelman (AS) & Prader-Willi (PWS) syndrome locus. AS is caused by a loss of maternal *UBE3A*, which is paternally silent in neurons. Silencing is mediated by the antisense *UBE3A-ATS*, which extends from *SNRPN* through *UBE3A* in neurons, as opposed to termination at *PWAR1* in non-neurons. Quantitative (qPCR) analysis confirmed the exclusive & progressive increase in *UBE3A-ATS* in differentiating LUHMES neurons, validating their use for studying *UBE3A* silencing. Genome-wide transcriptome analyses revealed changes to 11,834 genes during neuronal differentiation, including the upregulation of genes within the 15q11-q13 locus.

To identify dynamic changes to chromatin loops related to transcriptional changes we performed a HiChIP analysis validated by 4C which identified two neuron-specific CTCF chromatin loops between *MAGEL2-SNRPN* & *PWAR1-UBE3A*. To determine if allele-specific methylation patterns may be associated with CTCF loop anchors, whole genome long-read nanopore sequencing was performed. We identified a paternally hypomethylated DMR near the *SNRPN* loop anchor exclusive to neurons & a paternally hypermethylated DMR near the *PWAR1* CTCF anchor exclusive to undifferentiated cells, suggesting its role in regulating the cell specific boundary that occurs there. Additionally, DMRs near CTCF binding sites, observed in both cell types, may also influence chromatin loop formation & gene regulation within this locus.

This study provides an integrated view of the epigenetic landscape at the 15q11-q13 locus during LUHMES cell differentiation, underscoring the complex interplay of transcription, chromatin looping, & methylation, offering valuable insights for future therapeutic approaches for AS & PWS.

Chapter 1: Background

1.1 Introduction

Angelman syndrome (AS) is a severe neurogenetic disorder that affects approximately 1 in 15,000 births, characterized by developmental delay, seizures, speech impairments, movement difficulties, and a unique happy demeanor (Angelman, 1965). AS arises from a functional loss of the *UBE3A* gene located within the 15q11-q13 region, a 6-Mb segment of the genome implicated in three neurogenetic disorders: Prader-Willi syndrome (PWS), AS, and 15q duplication syndrome (Christian et al., 1999). This locus has also been associated with autism spectrum disorder (Cook et al., 1997). This region is renowned for its intricate gene regulation paradigms, especially genomic imprinting, a process where genes are preferentially expressed from only one parental allele. Within this domain, at least 14 imprinted transcripts exist, with the majority being expressed exclusively from the paternally inherited allele in somatic tissues (Runte et al., 2000). Notably, *UBE3A* and *ATP10A* stand out as the sole imprinted genes in this region expressed from the maternally inherited allele (Meguro et al., 2001).

UBE3A is biallelically expressed in non-neuronal cells, but in neurons, it undergoes paternal imprinting, leading to exclusive maternal expression (Jiang et al., 1998). Most AS cases emerge from a *de novo* maternal allele deletion spanning approximately 6 million base pairs (Sadikovic et al., 2014). Due to the loss of maternal *UBE3A* and the imprinting of the paternal allele in neurons, affected individuals experience a lack of UBE3A production in their brain (Matsuura et al., 1997). *UBE3A* encodes the ubiquitin ligase E3A, a pivotal enzyme for synaptic development (Hanayama et al., 2010; Yashiro et al., 2010).

UBE3A is integral to the ubiquitin proteasome system, a cellular machinery responsible for tagging proteins for degradation, activation, or relocalization. Mutations or deletions in *UBE3A* lead to the accumulation of its protein substrates in the brain, disrupting its regular function and manifesting the characteristic symptoms of AS (Kishino et al., 1997). In mouse models, several vital substrates of Ube3a have been identified that influence neuronal function, including dendritic spine morphology and experience-dependent synaptic plasticity (Wang et al., 2019). Ube3a modulates excitatory synapse development by controlling the degradation of Arc, a synaptic protein that promotes the internalization of AMPA receptors (Dindot et al., 2018).

The 15q11-q13 region's imprinted domain is flanked by non-imprinted genes, including *TUBCGP5, CYFIP1, NIPA2, NIPA1, GABRB3, GABRA5, GABRG3, OCA2*, and *HERC2*. Most individuals with PWS and AS have a significant deletion of the 15q11-q13 region. Depending on their specific deletion breakpoints, they may also lack one copy of some or all the non-imprinted genes in the region. AS and PWS arise due to a truncating mutation from misalignment of HERC2 duplicons during recombination at the pachytene stage of prophase I (Burnside et al., 2011). Each duplicon creates a breakpoint in order from upstream to downstream BP1, BP2, BP3. BP1 is involved in 37% of the truncations, BP2 in 60%, and BP3 in 95%. Type I deletion involves BP1 and BP2, while Type II deletion involves BP2 and BP3, with the latter being the more commonly observed truncation (Amos-Landgraf et al., 1999).

1.2 Gene and protein functions

UBE3A is a HECT E3 ubiquitin ligase that orchestrates protein ubiquitination, leading to degradation through the ubiquitin proteasome system. Initially identified to target p53 in the presence of an E6 viral cofactor, *UBE3A's* role in the brain suggests an E6-independent ubiquitination of p53 (Huibregtse et al., 1993). Beyond p53, UBE3A regulates p27, a cyclin-dependent kinase inhibitor, and Arc, an immediate early gene pivotal for learning, memory, and homeostatic plasticity (Mishra et al., 2009; Dindot et al., 2018). Furthermore, UBE3A modulates Pbl, a Rho-GEF, hinting at Pbl being a UBE3A substrate (Reiter et al., 2006). Within the nucleus, UBE3A targets RING1B, another E3 ubiquitin ligase within the polycomb regulatory complex 2 (PRC2) (Mortensen et al., 2015). The loss of *UBE3A* results in altered p27 turnover and heightened p27 protein levels in various neuronal types (Mishra et al., 2009). *In vivo* binding of UBE3A to Arc and its *in vitro* ubiquitination implies UBE3A's necessity for Arc turnover in the brain during heightened synaptic activity (Dindot et al., 2018).

This genomic region houses several protein-coding transcripts, including *MKRN3, MAGEL2* (MAGE-like protein 2)*, NDN* (necdin), *SNRPN* and *UBE3A*. While MKRN3 is a potential ubiquitin ligase, *MAGEL2* resembles *NDN*, a melanoma antigen gene-encoding protein (Liu et al., 2017; Wijesuriya et al., 2017). *NDN* is thought to function as a growth suppressor in both neuronal and non-neuronal cells, engaging with the transactivation domain of E2F1 to promote cell cycle cessation (Taniura et al., 1998). The processing of RNA in this locus includes mechanisms such as alternative splicing, polyadenylation, and the exonucleolytic removal of sequences from introns.

MAGEL2's role in *SNRPN* gene enhancement remains uncertain. Deletion of the shortest regions overlap (SRO) for PWS on exon 1 of *SNRPN* can lead to methylation and reduced expression of *MAGEL2* and *NDN* (Buiting et al., 1995; Horsthemke & Wagstaff, 2008). This PWS SRO's influence on *MAGEL2* expression is believed to employ mechanisms like chromatin remodeling and histone modifications rather than direct enhancer activity (Horsthemke & Wagstaff, 2008; Soejima & Wagstaff 2005). *MAGEL2* also plays an important role in circadian rhythm, appetite control, fertility and neurodevelopment (Lee & Potts, 2017). Loss of function of *MAGEL2* has a phenotypic overlap with PWS and ASD which has been denoted as Schaaf-Yang Syndrome (Fountain et al., 2017; Schaaf et al., 2013).

The *SNRPN* transcript is crucial for gene expression regulation in the chromosome 15q11-q13 region, and it produces both exons encoding a protein with splicing functions, as well as diverse noncoding RNAs, including the *UBE3A* Antisense Transcript (*UBE3A-ATS*), which represses paternal *UBE3A* in the brain (Rougeulle et al., 1998). This long noncoding transcript also generates 72 C/D box snoRNAs, SNORD116 and SNORD115 which are conserved between humans and mice but are considered "orphans" because they lack known targets of more typical snoRNAs that modify ribosomal RNAs and/or spliceosomal RNAs (Falaleeva et al., 2015). They have also been shown to be associated with diurnal cycle, modifying the expression of multiple genes and even each other (Coulson et al., 2018).

1.3 Regulation of *UBE3A* imprinting and convergent transcription dynamics

The mechanism underlying the silencing of the *UBE3A* gene on the paternal allele in neurons appears to be a complex interplay between long non-coding RNAs (lncRNAs), convergent transcription, and the potential collision of transcriptional machinery. At the heart of this mechanism is a lncRNA that originates from the *SNRPN* promoter and extends towards *UBE3A*. This lncRNA is crucial for the silencing of the *UBE3A* gene in neurons. *UBE3A-ATS* is the terminal part of this lncRNA and is transcribed in the opposite direction to the *UBE3A* gene, a phenomenon termed convergent transcription (Landers et al., 2004).

RNA polymerase II, the enzyme responsible for transcribing DNA into RNA, moves along the DNA template strand and synthesizes the corresponding RNA. In the context of the *UBE3A* locus, convergent transcription implies that the RNA polymerases transcribing *UBE3A* and *UBE3A-ATS* could potentially meet, leading to a collision. This collision model suggests that the RNA polymerase transcribing *UBE3A-ATS* could either collide with the machinery transcribing *UBE3A* or could stall at the *UBE3A* promoter, thereby preventing its transcription and leading to its silencing (Mabb et al., 2011).

Another proposed mechanism is that *UBE3A-ATS* could recruit chromatin modifying enzymes, such as histone methyltransferases and histone deacetylases, to the paternal allele and induce repressive chromatin modifications, such as H3K9me3 and H3K27me3, that lead to transcriptional repression of the paternal *UBE3A* allele (Chamberlain, 2012). Alternatively, *UBE3A-ATS* could form RNA-DNA hybrids (R-loops) with the paternal DNA strand, which

could cause replication fork stalling and subsequent DNA methylation of the paternal allele (Powell et al., 2013).

1.4 CTCF and methylation

CTCF (CCCTC-binding factor) is a multifaceted transcriptional regulator that binds to specific DNA sequences, playing a pivotal role in chromatin architecture and gene expression regulation. It is found in various genomic regions, including gene promoters, enhancers, and insulators (Ruiz-Velasco 2017; Kurukuti et al., 2006). CTCF binding to DNA facilitates the formation of chromatin loops, bringing distant regulatory elements, such as enhancers, into proximity with gene promoters (Splinter et al., 2006; Yusufzai, et al., 2004). This spatial organization is crucial for the coordinated regulation of gene expression, ensuring temporal and spatial specificity in different cell types and under varying physiological conditions (Calderon et al., 2022). Furthermore, CTCF contributes to genome integrity by demarcating chromatin domains, thereby segregating active and inactive genomic regions and preventing the spread of active chromatin marks (Rao et al., 2014). This segregation is essential for maintaining the genomic landscape's stability and functionality.

The CTCF binding motif allows for bidirectionality binding (de Wit et al., 2005). Chromatin loops are formed preferentially by two convergent CTCFs and a cohesin ring which initially binds and then begins to extrude chromatin but tends to stop when it encounters them in this orientation (de Wit et al., 2005; Guo et al., 2015). This has been shown to be a cyclical and

dynamic process with CTCFs binding and unbinding in a matter of several seconds while cohesin can remain bound to chromatin for several minutes (Anders et al, 2018).

CpG methylation, involving the addition of a methyl group to cytosine within a CpG dinucleotide, significantly influences CTCF DNA binding. Methylation within CTCF binding sites can diminish or inhibit its binding, impacting chromatin loop formation and gene regulation (Bell et al., 2000). Additionally, methylation near these sites can alter chromatin conformation, affecting gene expression by modifying regulatory element accessibility (Fuks, 2005).

CTCF binding sites are often located in CpG islands, typically unmethylated genomic regions (Wiehle et al., 2019). Methylation in these areas can disrupt CTCF binding, leading to gene expression changes (Prickett et al., 2013). This interaction between CpG methylation and CTCF binding underscores the complexity of epigenetic regulation in gene expression.

In neurological disorders, aberrant methylation patterns can lead to gene silencing or activation, contributing to diseases like Rett Syndrome and Fragile X Syndrome (Lewis et al., 1992; Naumann et al., 2009). Similarly, in cancer, methylation changes can silence tumor suppressor genes or overexpress oncogenes, driving disease progression (Feinberg et al., 2006). Understanding methylation's role in these contexts is vital for developing targeted therapies.

In summary, CTCF's role in chromatin looping and gene expression, coupled with the influence of CpG methylation, highlights their importance in maintaining genomic stability and function. Their dysregulation can lead to various diseases, emphasizing the need for further research.

1.5 A Bipartite Boundary

Within the 15q11-q13 locus, a boundary region exists in non-neuronal cells, which contains the *IPW* and *PWAR1* genes. Notably, this region also houses a tandem CTCF binding site, suggesting a potential role for CTCF in the function of this boundary (Hsiao et al., 2019). CTCF, a multifunctional protein, is known to act as an insulator, preventing the spread of heterochromatin and ensuring that enhancers interact with their appropriate promoters (Yusufzai, et al., 2004). Additionally, CTCF can also facilitate the interaction between enhancers and promoters to activate gene transcription (Splinter et al., 2006). The binding of CTCF to its target sites is sensitive to DNA methylation; when these sites are methylated, CTCF binding is less likely. (Jones et al., 2001).

In non-neuronal cells, this boundary region effectively halts the progression of the *UBE3A-ATS*, preventing it from reaching the *UBE3A* gene and silencing it. However, in neurons, this boundary is absent or non-functional, allowing the *UBE3A-ATS* to extend past *PWAR1* and silence the *UBE3A* gene on the paternal allele (Martin et al., 2013).

Understanding the intricate mechanisms, especially the function of the boundary region in non-neuronal cells, could provide insights into how *UBE3A-ATS* expression is regulated. If

researchers can replicate the epigenetic landscape observed in non-neuronal cells within neurons, it might be possible to inhibit *UBE3A-ATS* and reactivate the paternal *UBE3A*, offering therapeutic potential for conditions like AS.

In summary, *UBE3A* also plays a role in gene expression regulation. Its loss of function correlates with altered expression of genes pivotal for synapse formation and function, further contributing to AS symptoms. Overall, *UBE3A*'s role in protein degradation and gene expression regulation is indispensable for the brain's proper functioning. The functional loss of *UBE3A* in AS leads to protein accumulation and altered gene expression, disrupting the brain's regular activities, and manifesting the disorder's characteristic symptoms (Wang et al., 2019; Dindot et al., 2018). The intricate epigenetic landscape regulated by CTCF and CpG methylation of the 15q11-q13 region, combined with the critical role of *UBE3A* in neuronal function, underscores the complexity of AS. Understanding the molecular mechanisms underlying this disorder is essential for developing targeted therapeutic strategies and providing better care for affected individuals.

1.6 Murine models provide valuable but limited AS perspectives

Through studies in both mouse and human systems, much has been learned about the genomic organization and regulation of imprinted expression in the 15q11-q13 region. Imprinted genes are usually clustered and utilize a variety of mechanisms to establish and maintain their imprinted status, including differential DNA methylation, differential histone modification, and antisense transcription (MacLean et al., 2011; Williamson et al., 2011). While there is strong

conservation of gene structure and regulation of imprinting between the human chromosome 15q11-q13 region and the mouse chromosome 7C region, there are some differences in gene content and expression patterns. Some protein-coding genes and noncoding RNAs are unique to either human or mouse, and there are differences in the tissues where certain genes are expressed. However, the core elements of genomic imprinting and the bipartite imprinting center appear to be conserved between the two species (Grabiel et al., 1998; Chamberlain & Brannan, 2001).

Another difference is that the size of the AS region is larger in humans than in mice. The human AS region spans approximately 4 Mb, while the mouse AS region is around 1.5 Mb (Runte et al., 2001). This size difference is due to differences in the repetitive DNA sequences and the number of copies of transposable elements in the two genomes.

Mouse models have been a cornerstone in the study of Angelman Syndrome (AS), providing valuable insights into its genetic and phenotypic intricacies. One of the most frequently utilized model is known as Ube3a^{tm1Alb1}. This model is characterized by the removal of the fifth exon associated with isoform 2 of the *UBE3A* gene, leading to a dysfunctional protein. It has been effective in mirroring several key symptoms including motor dysfunction, vocalization, seizures, and repetitive behaviors (Jiang et al., 1998). However, it's important to note that this model does not effectively mimic the profound cognitive impairments often seen in AS. Other mouse models have also been created that recapitulate this cognitive dysfunction (Syding et al., 2022)

Motor deficits were another significant observation. These deficits manifested as abnormalities in gait, motor coordination, and motor learning. While these motor challenges have been largely attributed to a loss of cerebellar *UBE3A*, the exact role of the cerebellum remains speculative (Heck et al., 2008). The involvement of other neural circuits, such as proprioceptive, spinal, and basal ganglia circuits, in these motor deficits is significant. Furthermore, a loss of approximately 25% of dopaminergic neurons in the substantia nigra was observed in one of the mouse models, leading to deficits in dopamine-sensitive motor tasks (Mulherkar & Jane, 2010).

However, it's essential to approach these findings with caution. The convergent transcript responsible for paternal *UBE3A* silencing in mice is spliced differently and terminates at a distinct region in non-neurons, underscoring the genetic differences between mice and humans (Chamberlain & Brannan, 2001). This divergence highlights the limitations of mouse models in fully capturing the complexity of AS in humans.

Additionally, larger chromosomal deletions, particularly in the 15q11-q13 region, are prevalent in many AS cases (Burnside et al., 2011). These deletions often span genes neighboring *UBE3A*, such as *GABRB3* and *ATP10A*. The interplay between these genes and their collective role in AS remains an area of active research. While mouse models have been instrumental, they may not fully encapsulate the genetic and phenotypic complexity of AS, emphasizing the need for a multifaceted research approach to gain a more comprehensive understanding of the underlying molecular mechanisms and potential therapeutic targets.

1.7 Exploring cellular and animal models for AS research

The quest to understand the molecular and cellular underpinnings of AS has led to the development of diverse research models. Each model offers a unique lens through which the intricacies of AS can be examined, providing invaluable insights into its pathogenesis and potential therapeutic avenues.

Induced pluripotent stem cells (iPSCs) have been a cornerstone in AS research. By reprogramming patient-derived fibroblasts to a pluripotent state and subsequently differentiating them into neurons, this model offers a dynamic platform to study AS's molecular intricacies within a controlled environment. However, iPSCs present challenges in terms of ensuring consistent differentiation into the desired cell types (Takahashi et al., 2007).

Mouse models, specifically the *Ube3a*-knockout mouse primary cell culture, have been instrumental in understanding AS. By specifically deleting the *Ube3a* gene in mouse embryonic stem cells and guiding their differentiation into neurons, this model simulates AS's progression, offering insights into its impact on neural pathways (Jiang et al., 1998).

The fruit fly, Drosophila melanogaster, serves as another valuable model for studying AS. Due to its genetic tractability and well-characterized nervous system, a Drosophila model for AS has been developed. Flies with mutations in the fly counterpart of *UBE3A*, *dube3a*, display abnormal locomotive behavior, circadian rhythms, and defective long-term memory (Wu et al., 2008).

The SH-SY5Y neuroblastoma cell line is a popular in vitro model for studying neurodevelopmental disorders, including AS. These cells are derived from neuroblastoma and could present an aberrant epigenome. However, their primary advantage lies in their ease of use in laboratory settings due to their ease of growth and transfection (Xicoy et al., 2017).

Primary culture cells offer a closer in vivo resemblance. However, their genetic and functional characteristics can change over time, especially with prolonged culture.

The pig model has the potential to be particularly insightful. Pigs, with their close physiological resemblance to humans, have been employed to create AS models using CRISPR-mediated genome editing by Scott Dindot's team (unpublished work). These models have exhibited behavioral and cognitive abnormalities similar to human AS patients, emphasizing their potential as a tool for studying the disorder.

Among these models, the LUHMES cell culture model holds significant promise. Originating from the human mesencephalon, these cells differentiate quickly into neurons (Scholz et al., 2011). Their expression of *UBE3A-ATS* can position them as a prime candidate for in-depth AS research, especially in understanding *UBE3A* silencing.

While each model has advanced our understanding of AS, the journey to fully comprehend this disorder continues. Each model, with its strengths and limitations, contributes a piece to the puzzle, with the LUHMES cell culture model potentially offering a path to future breakthroughs.

1.8 Summary

The current theory of paternal *UBE3A* imprinting suggests a collision model of convergent transcription. In this model, polymerase II from the lncRNA originating at *SNRPN* (*UBE3A-ATS*) travels downstream, potentially affecting *UBE3A* transcription. Notably, CTCF, a potential regulator of gene expression, may play a role in governing *UBE3A-ATS* transcription beyond *PWAR1*, impacting *UBE3A* imprinting.

Addressing a critical need in the study of AS, the presence of an imprinted yet intact copy of *UBE3A* creates a unique opportunity for potential therapeutic interventions by unsilencing the paternal allele. However, the mechanisms governing tissue-specific imprinting control at this locus remain elusive. To advance research in this area, a model that is easy to cultivate and quick to differentiate is essential. The immortalized human dopaminergic neuronal precursor cell line LUHMES emerges as a potential solution, expressing *UBE3A-ATS* within just seven days. LUHMES cells' rapid differentiation into mature neurons and non-cancer cell origin provides advantages for investigating the locus's boundary region and the role of CTCF in *UBE3A-ATS* expression.

In conclusion, deciphering the intricacies of AS hinges on unraveling the complexities of *UBE3A* imprinting and its epigenetic regulation. The LUHMES cell culture model holds the potential to provide insights into this intricate disorder, expediting the development of therapeutic strategies.

Chapter 2: Intended Manuscript "Integration of CTCF loops, methylome, and transcriptome in differentiating LUHMES as a model for imprinting dynamics in human neurons"

2.1 Introduction

Human *in vitro* models play a crucial role in advancing our understanding of neurodevelopmental disorders. These models offer a controlled environment to investigate the intricate interplay of genetics and epigenetics, shedding light on the molecular mechanisms underlying these disorders. However, for *in vitro* models of imprinted neurodevelopmental disorders associated with human 15q11.2-13.3 deletions and duplications, there are additional considerations due to the developmental transcriptional dynamics of this locus in early postnatal neuronal maturation (Leung et al., 2011).

Angelman syndrome (AS) is a severe neurogenetic disorder affecting approximately 1 in 15,000 births. It is characterized by developmental delay, seizures, speech impairments, movement difficulties, and a happy demeanor (Angelman, 1965). AS is caused by a functional loss of *UBE3A* (Matsuura et al., 1997) located within the 15q11-q13 region. This gene is biallelically expressed in non-neurons but in neurons it is paternally imprinted resulting in exclusive maternal expression (Jiang et al., 1998). Most AS cases arise from a *de novo* maternal allele deletion spanning about 6 million base pairs (Sadikovic et al., 2014). Because of the loss of the maternal *UBE3A* and the paternal allele's imprinting in neurons, AS individuals lack UBE3A within their brain (Matsuura et al., 1997). *UBE3A* codes for a ubiquitin ligase E3A protein which is essential for synaptic development (Hanayama et al., 2010; Yashiro et al., 2010).

SNRPN is located upstream to *UBE3A* and encodes a protein regulator of alternative splicing (Huntriss et al., 1993). The *SNRPN* protein coding region is at the 5' end of a longer 700 kb transcript that includes an extensively spliced long non-coding RNA (lncRNA) (Runte et al., 2001). In neurons and non-neurons, paternal expression of this lncRNA begins at the *SNRPN* promoter and extends past *SNORD116 and SNORD115*, a repetitive region of small nucleolar RNAs (snoRNA) that are processed from the larger host gene transcript (*SNHG14*). Deletion of the shortest region of overlap (SRO) on the paternal *SNORD116* is the minimal deletion associated with Prader-Willi syndrome (PWS) (Sahoo et al., 2008; de Smith et al., 2009). In non-neurons the transcript terminates at the non-coding *PWAR1* (Martins-taylor et al., 2018). However, in neurons, it continues beyond *PWAR1* through the *SNORD115* cluster and further extends antisense to *UBE3A (UBE3A-ATS)*. This antisense transcript has been shown to be responsible for the silencing of the paternal allele in neurons (Rougeulle et al., 1998; Figure 1).

2.1.1 Figure 1: Genetic landscape at the AS/PWS locus

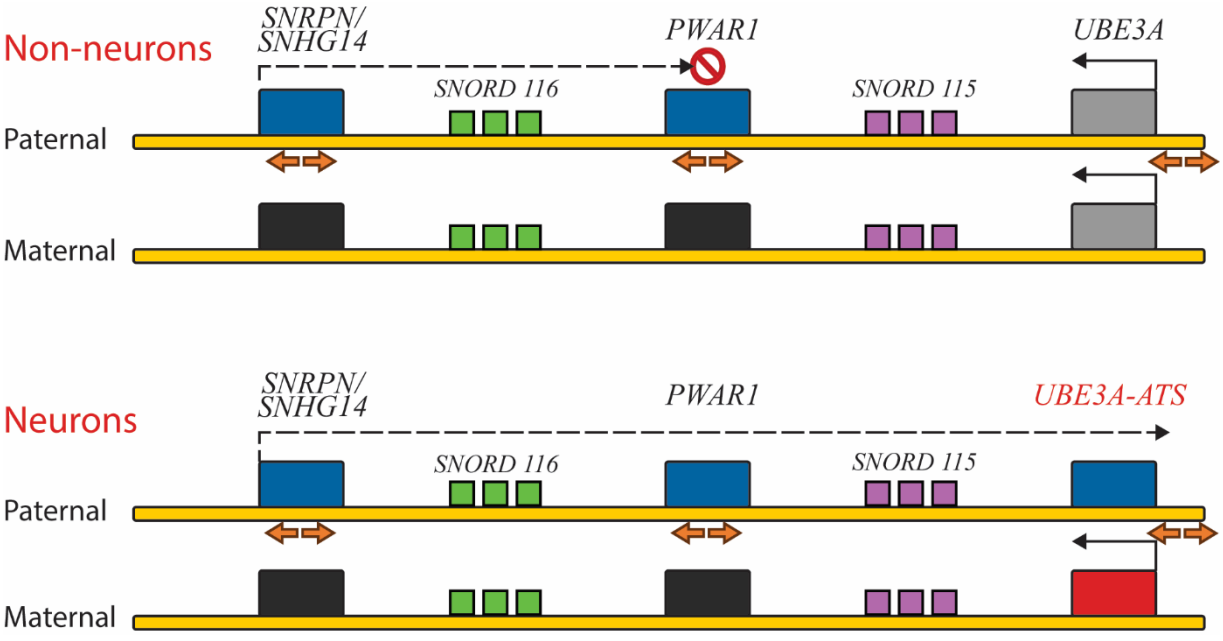

Figure 1. Simplified map of the human 15q11-q13 imprinted locus showing transcription termination of *SNHG14* lncRNA in non-neurons (top) compared to extension through *UBE3A-ATS* in neurons (bottom). Red indicates exclusively maternally expressed genes, blue is exclusively paternally expressed genes, gray is biallelically expressed and black is repressed. Arrows indicate divergent CTCF binding sites. *SNORD116* repeats are shown in green, while *SNORD115* repeats are shown in purple.

Non-neuronal cells from a PWS patient with a *SNORD116* deletion that included *PWAR1* were shown to express *UBE3A-ATS* suggesting the existence of a boundary region (Martins-taylor et al., 2014). The presence of binding sites for the insulator protein CTCF (CCCTC-Binding Factor) at *PWAR1* suggested that this boundary may serve as the barrier to transcriptional extension in non-neurons (Hsiao et al., 2019). CTCF associates with cohesin to form chromatin loops which have been shown to regulate tissue and allele specific differential gene expression (Hansen et al., 2017; Kurukuti et al., 2006). Reduced CTCF binding correlates with CpG hypermethylation at its binding site. (Renda et al., 2007). The CTCF binding motif allows for bidirectionality (de Wit et al., 2005). Chromatin loops are formed preferentially by two convergent CTCFs and a cohesin ring. Cohesin initially binds and begins to extrude chromatin but tends to stop when it encounters them in this orientation (de Wit et al., 2005; Guo et al., 2015). This has been shown to be a cyclical and dynamic process with CTCFs binding and unbinding in a matter of several seconds while cohesin can remain bound to chromatin for several minutes (Anders et al, 2018).

Previous studies have suggested a neuronal transcriptional collision mechanism in which the *UBE3A-ATS* silences paternal *UBE3A* in neurons by outcompeting the *UBE3A* sense transcript, but the exact mechanism is poorly understood (Faghihi et al., 2009; Mabb et al., 2011).

A major challenge to the field is that no *in vitro* model can fully replicate the dynamic processes that occur during neurodevelopment in the human brain. Differentiation protocols might not accurately recapitulate the complex maturation steps that *UBE3A-ATS* expressing neurons undergo *in vivo*. Moreover, epigenetic modifications crucial for the regulation of *UBE3A* expression may not be fully established or maintained in these *in vitro* systems. Models for studying the AS/PWS locus include SH-SY5Y cells and human induced pluripotent stem cells (iPSCs) from AS patients. SH-SY5Y are derived from cancer cells and thus may have an aberrant epigenetic profile. While patient-derived iPSCs hold great promise, their full differentiation to mature neurons is a challenging and costly process that can extend beyond 7 weeks (Hsiao et al., 2019). Despite their valuable insights, these models might not fully capture the intricate epigenetic complexities inherent in the 15q11.2-q13.3 locus and other disease loci with complex neuronal expression patterns.

In contrast, the human LUHMES (Lund human mesencephalic) cell line may be an ideal model to study neurodevelopmental disorders with an epigenetic component. LUHMES are human embryonic neuronal precursor cells capable of sustained proliferation, which is attributed to the presence of a tetracycline-inducible (Tet-off) v-myc transgene. When subjected to tetracycline along with glial cell-derived neurotrophic factor (gDNF) and dibutyryl cAMP, these cells can undergo differentiation into postmitotic dopaminergic neurons displaying the presence of β-tubulin, synaptophysin and the enzyme tyrosine hydroxylase. Furthermore, they showcase spontaneous electrical activity inherent to neurons (Scholz et al., 2011). Compared to pluripotent stem cell lines, they are relatively easy to grow and differentiate into neurons within 1 week.

However, this short life span can limit some uses and the cell line can be difficult to transfect. They have also shown good survivability with mechanical stress and FACS sorting.

CpG methylation plays an important role in the regulation of gene expression in imprinting control regions. However, less investigated is the role of chromatin loops and how they may regulate with CpG methylation the gene expression in imprinted regions such as the AS/PWS locus.

In this study, we conducted an integrated analysis of the LUHMES neuronal model system, encompassing genetic, epigenetic, and transcriptomic approaches. Our assessment revealed the temporal expression patterns of *UBE3A-ATS* and 11,834 transcripts genome-wide during differentiation of LUHMES to neurons. Furthermore, we identified a differential expression of multiple genes within the AS/PWS imprinted locus following neuronal differentiation and a distinct strand specific expression profile. Notably, we uncovered two CTCF loop interactions unique to LUHMES neurons from *MAGEL2* to *SNRPN* and from *PWAR1* to *UBE3A*. We found a hypomethylated paternal DMR on the *SNRPN* anchor exclusive to neurons. We also saw a hypermethylated paternal DMR near the *PWAR1* CTCF anchor in undifferentiated cells, suggesting its role in regulating the cell specific boundary.

2.2 Results

We hypothesized that LUHMES may be a particularly useful AS model and sought to further characterize its morphological, genetic, transcriptional, and epigenetic characteristics. LUHMES

cells showed an epithelial-like morphology in the undifferentiated state but demonstrate morphological characteristics of neurons including long neurites resembling mid-brain axonal networks within 7 days in differentiation media (Figure 2).

2.2.1 Figure 2: Microscopy of Undifferentiated and Differentiated LUHMES

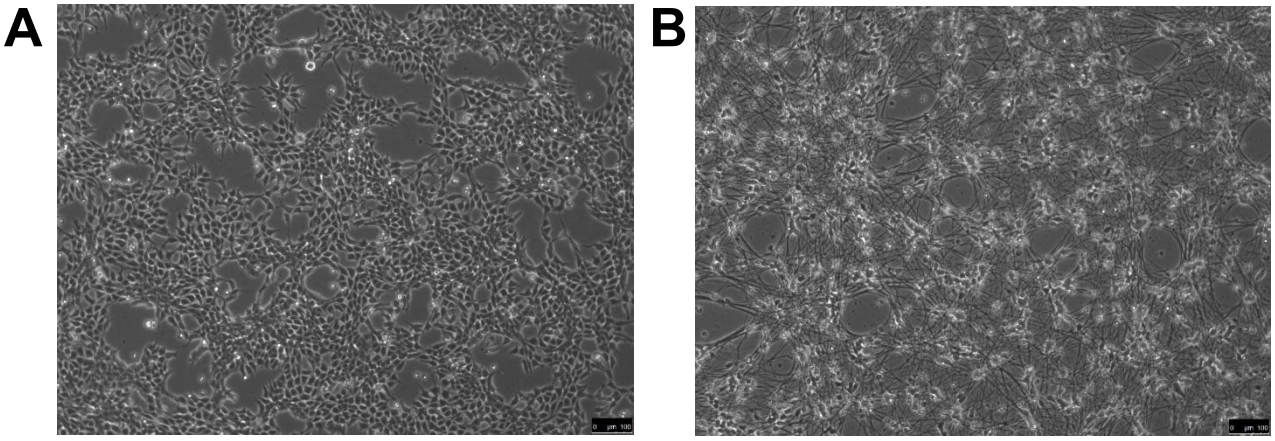

Figure 2. Brightfield Microscopy at 10X of **A.** Undifferentiated LUHMES. **B.** 7 day differentiated LUHMES neurons.

2.2.2 *UBE3A-ATS* is progressively induced during differentiation of LUHMES

To evaluate the relevance of the LUHMES differentiation system for the postnatal neuronal dynamics of the *UBE3A* locus in AS, we evaluated the expression levels of the *UBE3A-ATS* transcript by quantitative PCR (qPCR) across several cell types and human brain tissue. We used the $2^{-\Delta\Delta Ct}$ method to calculate relative *UBE3A-ATS* transcript levels in HEK293T cells, undifferentiated LUHMES cells, 7 days differentiated LUHMES neurons and adult cerebral cortex tissue (Figure 3A).

In the HEK293T cells and undifferentiated LUHMES cells, the *UBE3A-ATS* transcript was below the level of detection. In contrast, the differentiated LUHMES neurons showed high levels of *UBE3A-ATS*, which was comparable to that observed in adult cerebral cortex (Figure 3A).

We then used qPCR to characterize the temporal expression of the antisense transcript over a seven-day period in differentiation media (Figure 3B). On Day 1 and 2, *UBE3A-ATS* transcript levels were relatively low. However, by Day 4, there was a substantial increase in expression. This upward trend continued throughout the seven-day period, with the most substantial increases observed between Days 5 and 7. Together, these results demonstrate that LUHMES neurons are a valid model for the transcriptional changes in *UBE3A-ATS* expression known to occur during early postnatal neuronal maturation in the brain.

2.2.3 Figure 3: Expression of *UBE3A-ATS* in differentiated LUHMES

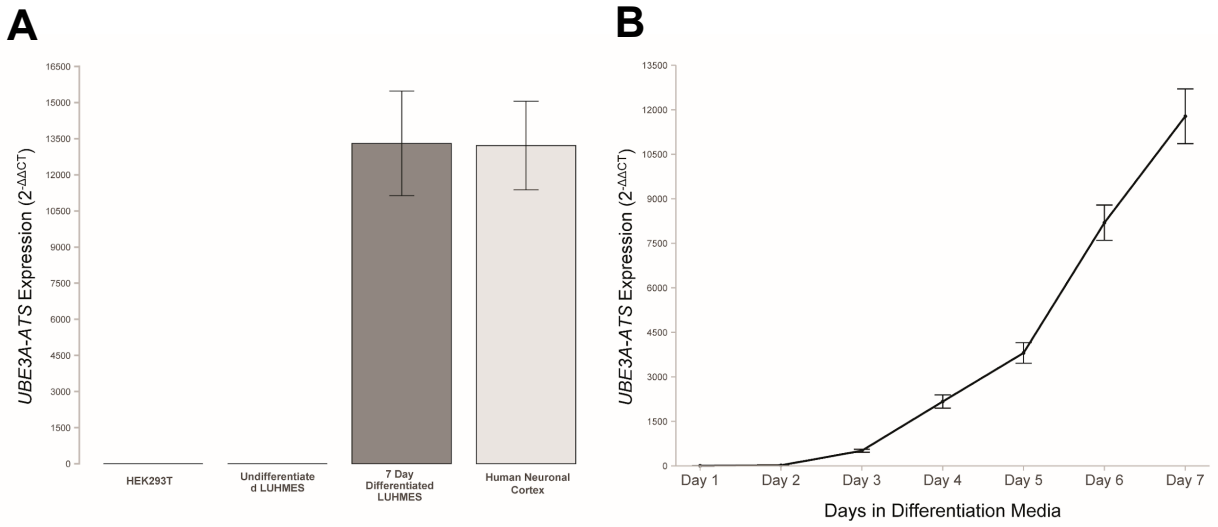

Figure 3. A. *UBE3A-ATS* is expressed in differentiated LUHMES neurons and brain cortex, but not in undifferentiated LUHMES and HEK293T cells. **B.** LUHMES *UBE3A-ATS* transcript levels progressively increased throughout 7 days in neuronal differentiation media.

2.2.4 Global transcriptomic differences between undifferentiated and LUHMES neurons

To further characterize the transcriptional changes in 6 days differentiated LUHMES neurons compared to undifferentiated, we performed RNA-seq in triplicates.

After correcting for genome-wide significance, when looking at differentially expressed genes in LUHMES neurons, we found 5,379 genes upregulated and 6,455 genes downregulated compared to undifferentiated LUHMES (Figure 4A; S1A).

In LUHMES neurons, the top ten differentially expressed genes, based on the lowest adjusted P values, were *ALCAM, MAP2, RTN1, NCAM1, CNTN2, AKAP6, KIF5A, SCD5, ROBO2,* and *NRG1*. All these genes showed significant upregulation in LUHMES neurons compared to undifferentiated LUHMES cells, as indicated by negative log fold change (logFC) values ranging from -4.85 (*ROBO2*) to -8.66 (*CNTN2*). The adjusted P values for these genes ranged from 4.27E-13 to 8.98E-13, indicating highly significant differential expression (Figure 4A).

The top ten differentially expressed genes that were downregulated in neurons were *H1-5, H2AC11, ASS1, H2BC18, NCAPD2, CCNB1, SMC4, SUSD2, HMGA2, and CENPF* with logFC values ranging from 4.56 (*NCAPD2*) to 7.53 (*H1-5*). The adjusted P values for these genes ranged from 4.94E-13 to 1.72E-12, again indicating highly significant differential expression (Figure 4A).

2.2.5 LUHMES neurons show an upregulation of key genes associated with the AS locus

Within the AS/PWS locus, several genes showed significant upregulation in LUHMES neurons compared to undifferentiated LUHMES cells. Notably, *MAGEL2, SNRPN, SNHG14, PWAR1,* and several small nucleolar RNAs (snoRNAs) within the *SNORD116* cluster showed significant upregulation, with logFC values ranging from -0.05 (*SNORD116-13*) to -5.20 (*SNORD116-24*).

The *UBE3A* gene, which is of particular interest in the context of AS, showed a slight upregulation, but this was more variable and not statistically significant (logFC = -0.08, adjusted P value = 0.45) (Figure 4B).

Only the plus strand transcriptional profile was distinct between these two cell states (Figure 4C). In neurons we saw an increase in plus strand transcription begin upstream of the *SNRPN* 5' splice site and in undifferentiated cells we saw an abrupt decrease after *PWAR1* (Figure 9B;C). In neurons transcription continues with a marked increase over the *UBE3A* gene body and again downstream (Figure 4C, 9D).

2.2.6 Figure 4: RNA-seq data

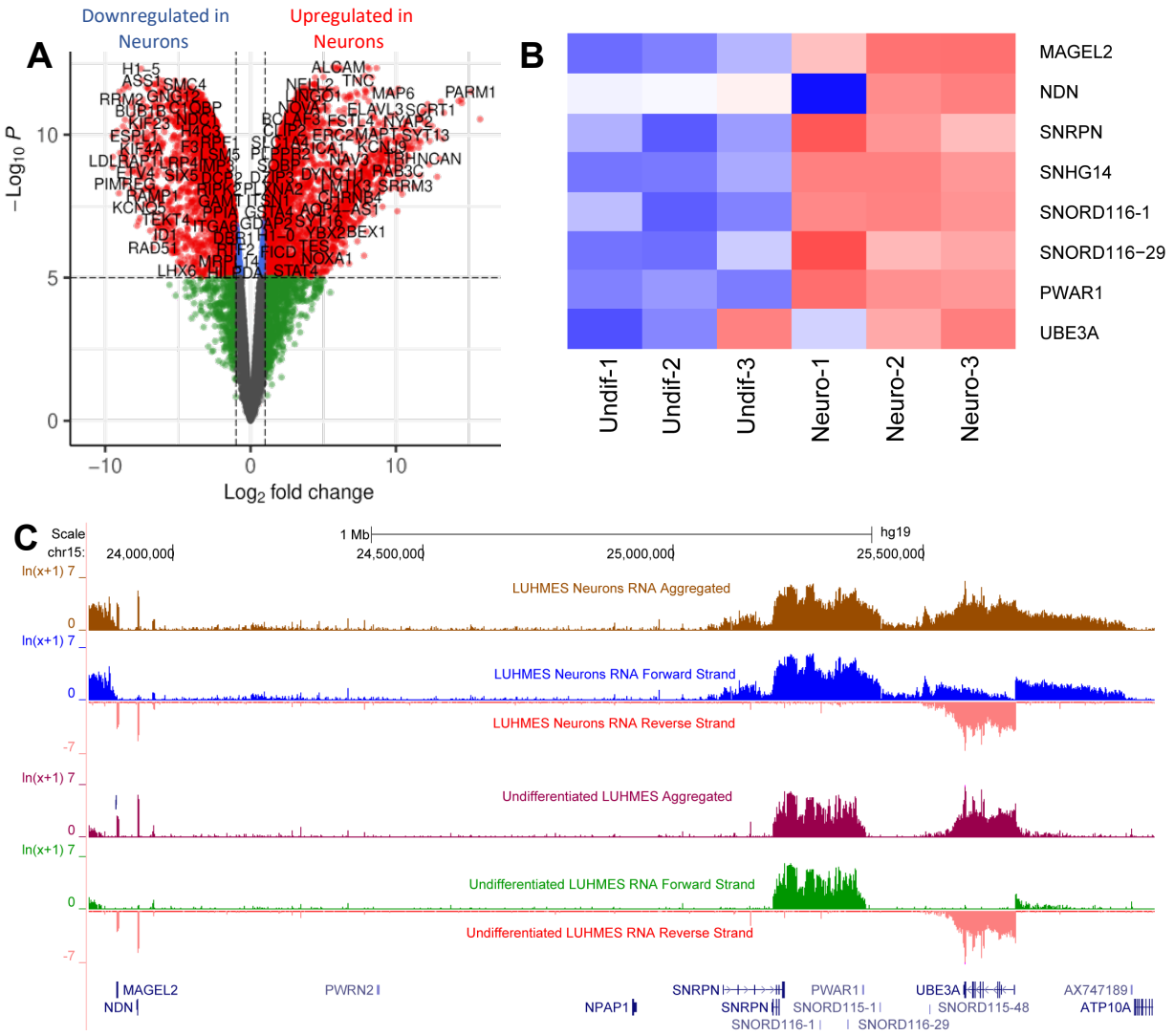

Figure 4: A. Volcano plot of differential gene expression on the minus strand comparing undifferentiated vs differentiated LUHMES in -Log10 P-value vs Log2 fold change. Gray = Nonsignificant, Green = Log2 fold change, -Log10 p-value < 5, Blue = -Log10 p-value > 5 without fold change, Red = -Log10 p-value greater than 5 and Log2 fold change. Log2 fold change and -Log10 P-value greater than 5 seen in red in the right upper quadrant indicates genes that are upregulated in LUHMES neurons. **B.** Heatmap of differentially expressed genes in AS locus based on Z score, Red = Upregulated in Neurons, Blue = Downregulated, shown in triplicates. Only first and last *SNORD116* are shown. **C.** LUHMES aggregated and strand specific RNAseq throughout the AS/PWS locus (chr15:23,832,378-25,962,021).

2.2.7 Gene ontology analysis

We then performed a gene ontology (GO) analysis to identify the biological processes that were enriched in LUHMES neurons and undifferentiated LUHMES cells.

In LUHMES neurons, a reactome pathway analysis revealed enrichment for the dopamine neurotransmitter release cycle pathway (p=2.79E-09). This finding is consistent with the expected dopaminergic nature of LUHMES neurons and further supports their neuronal identity (Figure 5A). The GO cellular component showed that the top process enriched in neurons are related to neuron projection (p=1.81E-7) and axonal development (p=7.23E-17) (Figure 5B)

Additionally, the top five enriched biological processes were nervous system development (p=1.62E-17), axonogenesis (p=2.05E-15), synapse organization (p=3.08E-15), axon guidance (p=1.24E-13), and modulation of chemical synaptic transmission (p=1.00E-11). These processes are all critical for neuronal function and development, suggesting that the genes upregulated in LUHMES neurons are involved in these key biological processes (Figure 5C).

In contrast, the top five enriched biological processes downregulated in LUHMES neurons were ribosome biogenesis (p=2.37E-76), gene expression (p=1.19E-72), translation (p=2.47E-72), rRNA processing (p=3.93E-72), and cellular macromolecule biosynthetic process (p=1.75E-71). These processes are fundamental for cellular function and growth which non dividing cells should downregulate (Figure 5D).

These results provide a comprehensive overview of the biological processes and pathways that are enriched in LUHMES neurons following differentiation and neuronal maturation. The identified processes and pathways may play key roles in the differentiation of LUHMES cells into neurons and warrant further investigation.

2.2.8 Figure 5: GO Terms

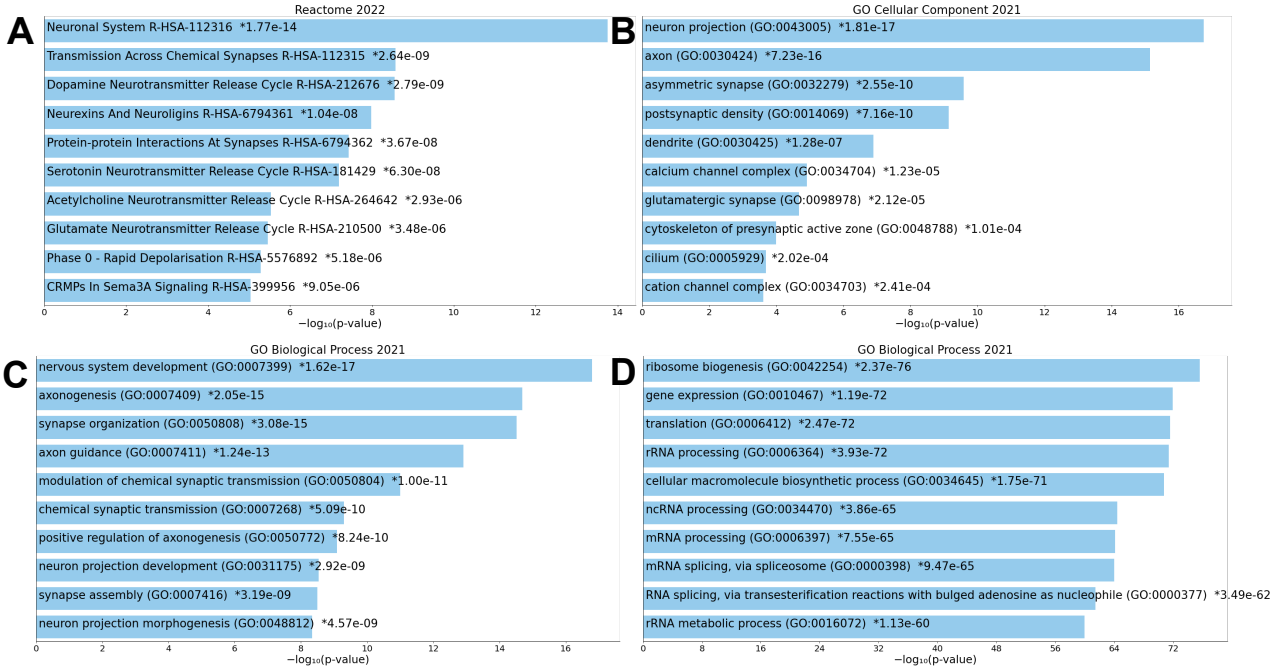

Figure 5: A. GO terms reactome 2020 enriched in LUHMES neurons **B.** GO terms cellular component 2021 enriched in neurons. **C**. Go biological process 2021 enriched in LUHMES neurons. **D.** GO biological process 2021 downregulated in LUHMES neurons.

2.2.9 Chromatin loop analysis revealed neuron specific CTCF loops in LUHMES

We next employed HiChIP analysis to investigate differential chromatin loop formations involving CTCF in undifferentiated and 6-day differentiated LUHMES neurons. CTCF is a key regulator of chromatin architecture and its role in the formation of chromatin loops is crucial for gene regulation. To understand the role of CTCF loop dynamics in the AS/PWS locus, we focused on a region spanning chr15:23,832,378-25,962,021 (hg19). Specifically in the

differentiated LUHMES neurons, we observed a significant long-range chromatin loop interaction spanning approximately 1.2 Mb between the *MAGEL2* gene (chr15:23,890,148-23,895,147) and a region about 100 kb upstream of *SNRPN* (chr15:25,090,148-25,095,147) (Figure 6A;6B). This interaction was given a score of 6 by our stringent analysis. In comparison, *MAGEL2* also interacts with a cluster of loops present in both cell types with values that range from single digits to 169 (chr15:23,890,148-24,105,147). Another notable neuron specific chromatin loop interaction was observed between the *PWAR1* gene (chr15:25,380,148-25,385,147), and a region located approximately 64 kb downstream of the *UBE3A* (chr15:25,745,148-25,750,147) with a value of 9 (Figure 6A). When using our loose filtering method, we also observed another interaction originating from the same *PWAR1* bin and landing on the *UBE3A* promoter at (chr15:25,680,148-25,685,147) that was unique to neurons with a score of 6 (Figure 6B). In turn that bin seems to show a nearby interaction within the 3' *UBE3A* body (Figure 6B). Using the same filtering method, genome wide we were able to separate 36,816 interactions unique to neurons, 74,469 unique to undifferentiated cells and 26,162 were shared between them (Figure S1B). We also observed some overall differences between the two cell types when looking at their contact matrix (Figure S4).

These findings provide novel insight into the dynamic changes in chromatin architecture that occur in the AS/PWS locus during the differentiation of LUHMES cells into neurons. The neuron-specific chromatin loops coincide with increased expression of multiple paternal transcripts, including *MAGEL2, NDN, SNHG14*, *SNRPN, PWAR1*, and *UBE3A-ATS* (Figure 9). These dynamic changes in neuronal chromatin structure associating with paternally expressed

transcripts suggest their involvement in paternal silencing of *UBE3A*, although these experiments do not directly determine the allele-specificity of CTCF binding.

2.2.10 Figure 6: HiChIP data on the WashU epigenome browser

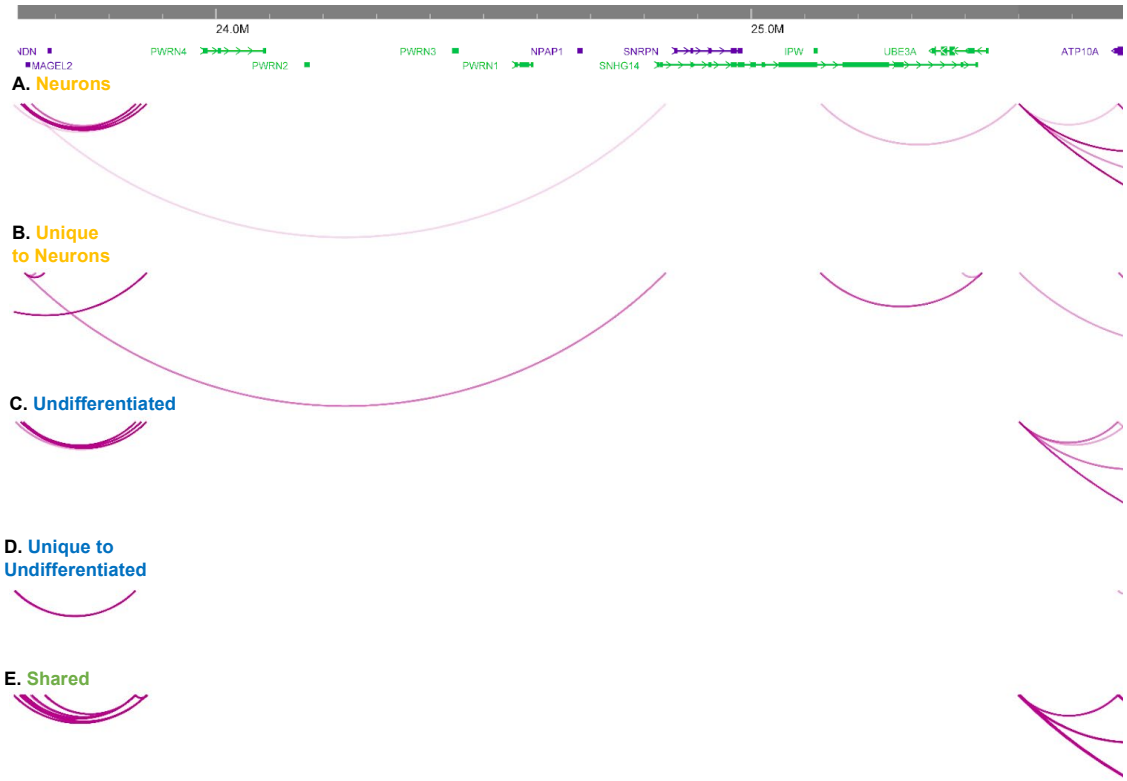

Figure 6: WashU epigenome browser depiction of the CTCF HiChIP demonstrating long range chromatin interactions for LUHMES neurons and undifferentiated LUHMES. HiChIP loops are shown as purple arches, with greater intensity reflecting greater significance. **A.** Loops detected in Neurons with stringent filtering using 5 kb bins at 0.05 FDR **B.** Loops unique to neurons using a less stringent analysis and FDR 0.1. **C.** Loops in Undifferentiated cells using 5 kb bins at 0.05 FDR. **D.** Loops in undifferentiated LUHMES using less stringent analysis 0.1 FDR. **E.** Only loops shared between the two cell types at 0.1 FDR.

2.2.11 4C validation of HiChIP findings

To validate the chromatin loops identified in our HiChIP analysis, we performed a 4C experiment using viewpoints from *SNRPN* (ICR), *IPW*, and *UBE3A*. This experiment allowed us to confirm the presence of these loops in LUHMES cells (Figure 7). A loop observed exclusively in neurons from *MAGEL2* to *SNRPN* in the HiChIP data was replicated by an

interaction from the *SNRPN* viewpoint at chr15:25,092,529 to a region which included *MAGEL2* from chr15: 23,885,650 to 23,902,345 (Figure 7A). In contrast, from the same viewpoint undifferentiated LUHMES, a different region of interaction was detected spanning 97,966 bp from chr15:25,318,254 to 25,332,219 (Figure 7B). Looking at the neurons from the *PWAR1* viewpoint chr15: 25,382,560 we observed two nearby interactions with the beginning *SNORD115* cluster from chr15:25,411,021-25,514,112 and from a region encompassing *UBE3A* from chr15:25,514,112-25,689,001 further supporting the interaction we saw using the less stringent HiChIP analysis (7C). In contrast, this interaction was not present in undifferentiated LUHMES (Figure 7D). Using a viewpoint from the *UBE3A* promoter at chr15:25,684,119 confirmed that it interacted with the *PWAR1* region from chr15:25,371,923-25,384,018 (Figure 7E). In contrast, the undifferentiated LUHMES with the same viewpoint interacted with three different regions. The first 2 close together between the *SNORD115* cluster (chr15:25549549-25,561,569) and the final 5' exon of *UBE3A* (chr15:25562286-25583229) with a third downstream close to *ATP10A* from chr15: 25907922-25926082 (Figure 7F).

2.2.12 Figure 7: 4C validation of loops

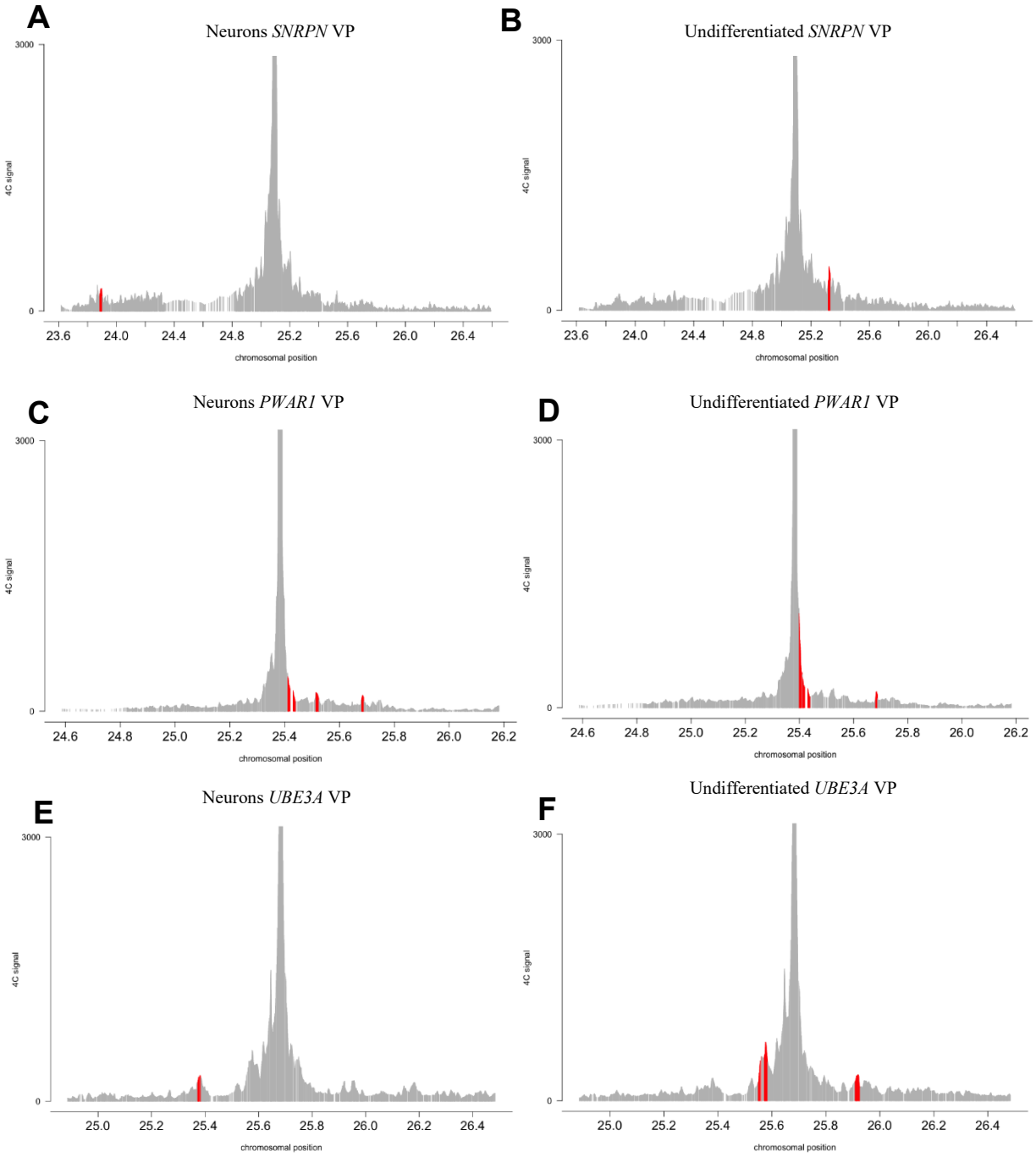

Figure 7: Results from a 4C analysis, in red are called interaction peaks (hg19) **A.** LUHMES Neurons using *SNRPN* viewpoint (chr15:25092529.) **B.** Undifferentiated LUHMES using the same SNRPN viewpoint. **C.** Neurons using the *PWAR1* viewpoint at chr15:25382560. **D.** Undifferentiated using the same *PWAR1* viewpoint (chr15:25137413). **E.** LUHMES neurons using a viewpoint located at the promoter of *UBE3A* (chr15: 25684119). **F.** Undifferentiated LUHMES using the same viewpoint.

2.2.13 CpG methylation and integration with CTCF loops and expression

We used Oxford Nanopore (ONP) sequencing to examine CpG methylation within the AS/PWS, with a particular focus on the CTCF motif sequences where unique loops were found in neurons. This analysis provides us with insights into the relationship between DNA methylation, CTCF loops and gene expression that could potentially regulate imprinting of *UBE3A* and other genes within the AS/PWS. We used ONP's pipeline, modkit pileup, to call methylation and visualize the data (Figure S2). We observed little difference in global methylation landscape patterns between the undifferentiated LUHMES and neurons. ONP's long reads provided the advantage of allowing phasing of the methylome using nanomethphase (Akbari et al., 2021), which is of particular importance for imprinted loci. UCSC browser track hubs were created for visualization together with our LUHMES HiChIP and RNAseq data as well as other genome annotations (Figure 8;9). We were able to assign parentage for each haplotype based on the well characterized hypomethylation of the PWS-ICR (Figure 8B).

A distinct pattern of CpG methylation landscape was observed in an allele specific manner in both neurons and undifferentiated cells (Figure 8A). Narrow regions of maternal hypermethylation are seen within broader regions of paternal hypermethylation, a pattern that was previously observed by whole genome bisulfite sequencing in postmortem PWS, AS, and Dup15q brain samples (Dunaway et al., 2016). When comparing the paternal allele to the maternal allele this regions appears to be a hotbed for DMRs particularly between both loops (Figure 8A;8B).

For both cells types we observed paternally hypomethylated regions overlapping with the *MAGEL2* and *NDN* promoters with a region of hypermethylation in between (Figure 9A). However, a paternally hypomethylated region overlapping the *SNRPN* interacting bin was exclusive to neurons (Figure 8B;9B). A downstream paternal hypomethylation at a CpG island close to the PW SRO is associated with the beginning of *SNRPN* transcription in undifferentiated cells. In contrast we found a paternally hypermethylated DMR exclusive to undifferentiated LUHMES upstream of the *PWAR1* anchor after which transcription decreases dramatically (Figure 9C). All loop anchors were adjacent to increased plus strand transcription in neurons (Figure 8;9). The loop anchor located close to *UBE3A* is about 61 kb 5' of its biallelically hypomethylated CpG island promoter. A differential methylation analysis was also performed comparing the paternal allele in neurons vs the paternal allele in undifferentiated cells, however this resulted in few occurrences throughout a large portion of chromosome 15. This was also true when comparing the maternal allele in neurons to the maternal in undifferentiated cells (Figure S3).

2.2.14 Figure 8: Integration of phased methylation data on UCSC genome browser

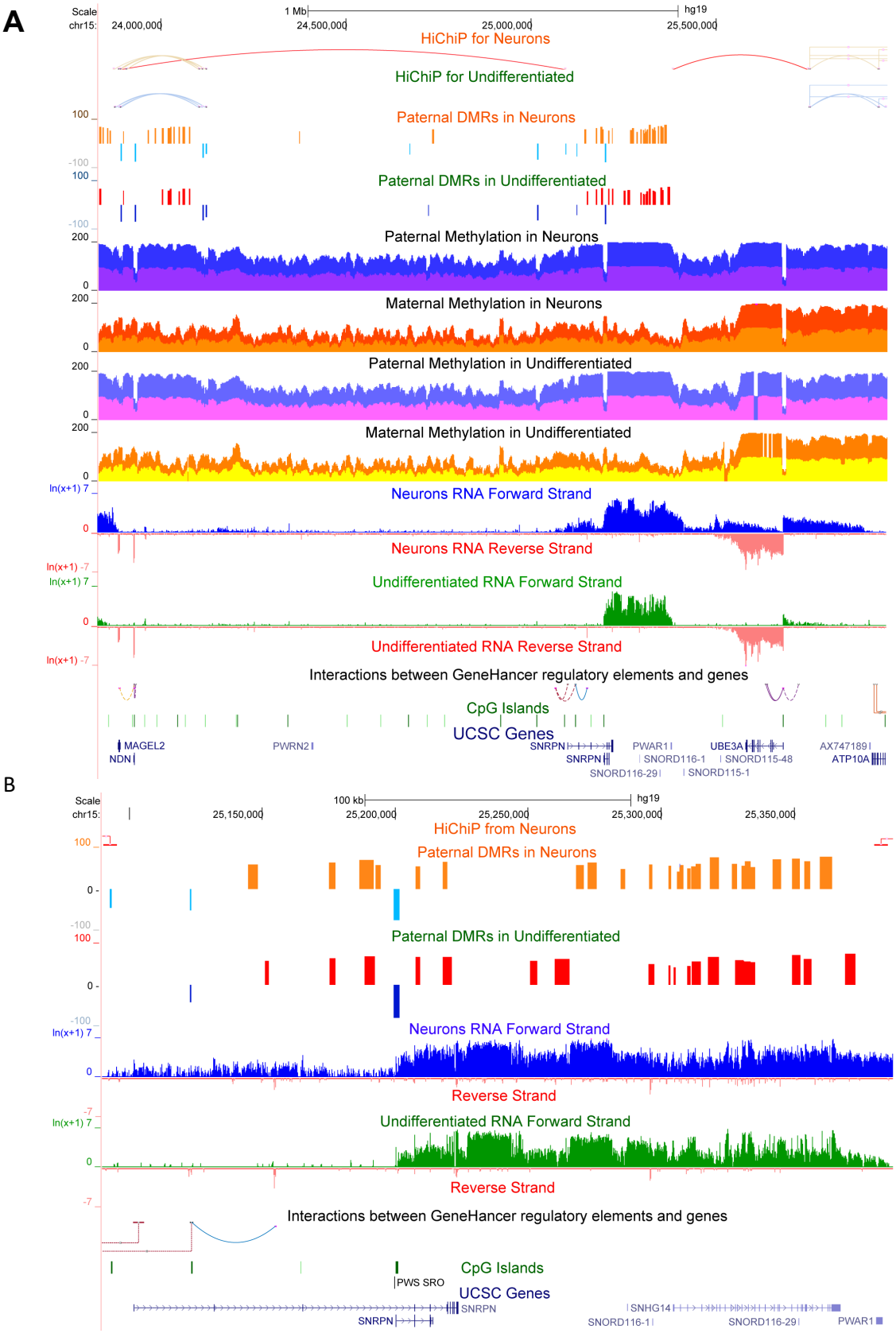

Figure 9: Methylation, HiChIP and RNA-seq data with various UCSC genome browser tracks (hg19) interacting 5kb loop bins are shown in red. For neurons paternally hypermethylated DMRs are shown in orange and paternally hypomethylated hypomethylation. In undifferentiated LUHMES paternally hypermethylated DMRs are shown in red with dark blue representing hypomethylation. DMRs represent differences in percent methylation **A.** CTCF loops in neurons and undifferentiated LUHMES; methylation profile for both alleles in both cell types, replicates were stacked and shown in different colors for contrast with values representing the sum of their percent methylation (max 200); Strand specific RNAseq in neurons and undifferentiated LUHMES; Genehancer regulatory elements, CpG island and UCSC Genes are also included (chr15:23,832,378-25,962,021) **B.** The paternal DMR cluster between the *SNRPN* to *UBE3A* loop anchors (chr15:25,089,681-25,387,210)

2.2.15 Figure 9: DMRs and Transcription at CTCF loop anchors

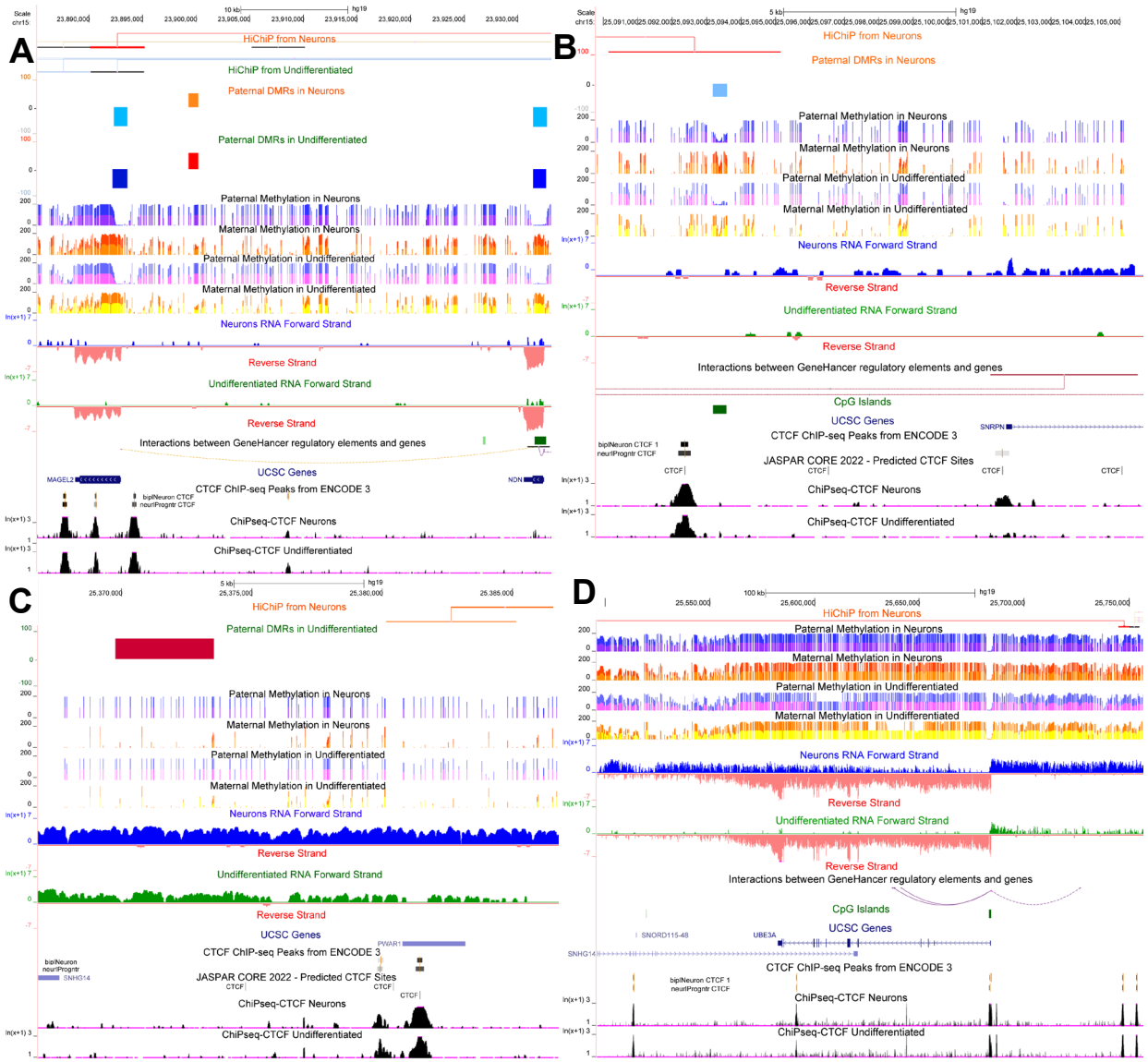

Figure 9: CTCF loops, methylation and transcription from the same LUHMES sample, CTCF ChIP-seq from LUHMES from Pierce et al., 2018 (hg19): Same tracks shown in figure 8 with some negative data omitted **A.** A close look at the region encompassing the *MAGEL2* anchor and *NDN* (chr15:23,885,189-23,933,105) **B.** *SNRPN* anchor that overlaps with a neuron specific paternally hypomethylated DMR (chr15:25,089,792-25,105,663) **C.** *PWAR1* loop anchor and the paternally hypermethylated DMR exclusive to undifferentiated cells (chr15:25,366,804-25,386,789) **D.** *UBE3A* anchor region showing a hypermethylated profile but no paternal DMRs in either cell type (chr15:25,496,109-25,756,788).

2.2.16 Linked-read along with long read sequencing assign parentage to 15q11-q13

For modeling diseases that involve imprinting it is also vital to assign parental SNPs to each allele. To be able to target a specific allele for knockout, heterozygous SNPs could be used to design CRISPR-Cas9 gRNAs. These SNPs could also be used to target one allele when piloting therapeutic interventions such as ASO's, genetic and epigenetic editing. Furthermore, SNP signatures can also be used when assessing an intervention's effect on allele-specific expression.

To assign parental origin to the alleles in the AS/PWS region, we performed 10X-linked read sequencing. However, due to a lack of heterozygosity, the continuity of the haplotype blocks was disrupted downstream of *SNORD115*, including exons 10-14 of *UBE3A* (Figure S5). This meant that the assignment of haplotype 1 on one side of this gap did not necessarily correspond to the same allele for haplotype 1 on the other side of the gap. To overcome this challenge, we utilized the ONP long read sequencing that allowed us to span the gap region and assign SNPs to either maternal or paternal alleles.

2.3 Discussion

In this comprehensive study, we characterized and integrated genome-wide DNA methylation with CTCF loops and RNA expression of LUHMES cells to shed light on their relationship, with a particular emphasis on evaluating their potential as a model for AS. Our findings provide a robust foundation for the use of this cell line as a valid model, using the undifferentiated as a non-neuron and differentiated as a neuron, especially in the context of allele specific imprinting.

We demonstrate the progressive induction of *UBE3A-ATS* during their differentiation, with an upregulation of genes pivotal to neuronal function, and differential expression of genes in the AS/PWS locus including *MAGEL2, SNRPN, SNHG14*, and several snoRNAs within the *SNORD116* cluster, providing a compelling case for the relevance of LUHMES neurons in 15q11.2-13.3 disorder research.

Only the plus strand transcriptional profile was distinct between these two cell states (Figure 4C, 8A). In neurons we saw an increase in plus strand transcription begin before *SNRPN* with an abrupt decrease after *PWAR1* in undifferentiated cells (Figure 8A; 9C). Transcription continues in neurons with a marked increase over the *UBE3A* gene body and again downstream of it. This region was also highly methylated on all haplotypes (Figure 9D).

We identified neuron specific convergent CTCF loops between *MAGEL2* and *SNRPN* as well as one from *PWAR1* to *UBE3A*. The *MAGEL2-SNRPN* loop was not as frequently seen as nearby interactions but persisted even when using stringent filtering and seen again in our 4C from a

separate time point. Therefore, we can't rule out a possible enhancement of *SNRPN* lncRNA transcription. This interaction may also regulate the paternal specific expression of *MAGEL2* by the PWS ICR although it's located over 100 kb away from where the *SNRPN* anchor.

When comparing the paternal to maternal alleles we found several DMRs including a large cluster between *SNRPN* and *PWAR1*, and hypomethylated DMRs within a few kb of the *MAGEL2* and *SNRPN* CTCF binding sites in both cell types. Interestingly though, the DMR close to *PWAR1* was unique in that it was hypermethylated and it was only seen in the undifferentiated LUHMES. This could point to the criticality of the *PWAR1* binding site as part of the boundary region where the lncRNA stops in non-neurons. This DMR and the two chromatin interactions could represent the necessary and or sufficient conditions for favoring progression of transcription from *SNRPN* through *PWAR1* and *UBE3A-ATS*, silencing the paternal *UBE3A* allele in neurons.

One of the prevailing models for how *UBE3A* is silenced by *UBE3-ATS* is the collision model which proposes that RNA polymerase II from the convergent transcripts disrupt expression, leading to the repression of one or both genes (Mabb et al., 2011). Our analyses suggest a fresh model of how the lncRNA that begins at *SNRPN* could regulate *UBE3A-ATS* expression and therefore silence the paternal allele in neurons. Since *MAGEL2* shares enhancer elements with the nearby gene *NDN* it could be acting in the same manner or recruiting transcription factors to *SNRPN* by way of this novel loop (Stelzer et al., 2016; Fishilevich et al., 2017). This *MAGEL2* to *SNRPN* chromatin interaction or in combination with the loop from *PWAR1* to *UBE3A* could

perturb the delicate balance of where RNA polymerase II collision occurs between these convergent transcripts (Figure 10).

Further studies are warranted to determine the functional implications of these chromatin interactions and DMRs in the context of *UBE3A* regulation and AS. One way could be through ablation of the binding sites but a more nuance approach could through epigenetic editing directed towards these DMRs to determine if *UBE3A-ATS* expression is perturbed. A Cas9 fused to DNA demethylator TET1 directed towards the hypermethylated DMR close to *PWAR1* in undifferentiated cells could get it to express *UB3A-ATS*. You can also use a DNA methylator dCas9-DNMT3AL to methylate the CTCF binding site trying to reduce *UB3A-ATS* in neurons. The fact that dCas9 also has a large footprint and a stronger bond energy with DNA a no ATF control is at least work a try. For the use of this model to advance a maternal knockout must be created. By assigning parentage to heterozygous SNPs in the AS/PWS locus we facilitate the use of this cell line for more allele specific research including knockouts. Optimizing transfection protocols for this cell line can be tricky, usually showing poor transfection efficiency and high death rates.

Our findings underscore the crucial role of chromatin looping and methylation in dictating the tissue-specific imprinting pattern at the *UBE3A* locus and how this cell line could provide novel insights into potential epigenetic strategies for unsilencing the paternal allele as a therapeutic approach for AS.

2.3.1 Figure 10: Proposed epigenetic model of the AS/PWS locus

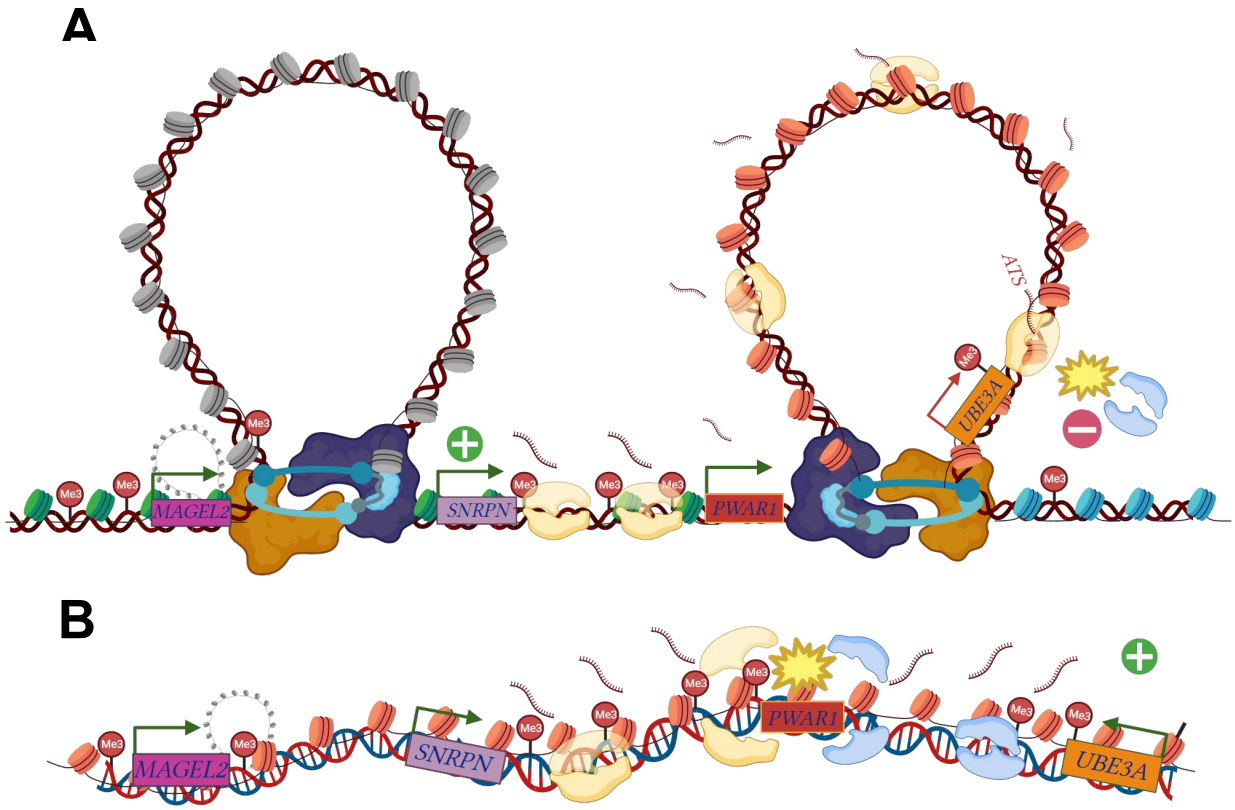

Figure 10. Proposed model of the epigenetic landscape on the paternal allele at the human 15q11-q13 locus **A.** LUHMES Neurons and **B.** Undifferentiated cells. Boxes represent labeled gene regions, arrows demonstrating direction of transcription with green representing transcribed genes. Green plus icons and red minus icons also highlight differential expression between the two cell types. CpG hypermethylated is represented with Me3 lollipops. Yellow star denotes proposed sites of polymerase II collision. The loop proximal to *MAGEL2* is shown as in gray. Note the hypermethylation seen close to *PWAR1* in the undifferentiated state. Not to scale, created with BioRender.com

2.4 Methods

2.4.1 Cell culture

LUHMES were purchased from ATCC catalog number CRL-2927 (2021). They were cultured and differentiated as described in Scholz *et al* 2011 with minor optimizations. We found that only using hydrophilic Nunclon® Δ surface treated flask allowed for appropriate cell adherence (Sigma, Cat. No. F7552-1CS. We also modified the additional coating of poly-L-ornithine (Sigma, Cat. No. P-3655-10MG) and fibronectin (MilliporeSigma, Cat. No. 341631-1MG) by mixing 50 ug/mL of poly-L-ornithine and 1 ug/mL of fibronectin, covering the flask and incubating overnight at 37 C. We found that 2 rinses with water was enough for the subsequent wash. For maintenance media we used DMEM/F-12, GlutaMAX™ (Gibco-Invitrogen, Cat. No. 10565-018) with 1% N2 supplement (Gibco-Invitrogen, Cat. No. 17502-048). Undifferentiated LUHMES were cultured to 80% confluency by removing media, washing with Ca++/Mg++ free Dulbecco's phosphate-buffered saline (D-PBS) and incubating in warm 0.025% trypsin in D-PBS at 37 C for 2 minutes followed by light scraping. For differentiation we used the maintenance media with the addition of 2 ng/ml of human recombinant gDNF (Thermo, Cat. No. PHC7045), 1 mM of dibutryl cAMP (Sigma, Cat. No. D0627) and 1 ug/ml of tetracycline (Sigma, Cat. No. T7660-5G). Cells were place in differentiation media to 50-70% confluency and the first day was considered day 0. Differentiation media was changed every other day but leaving approximately 20% of the prior media. Since LUHMES neurons may become detached on day 7 those cells were harvested on day 6. Differentiated LUHMES were harvested in the same manner and undifferentiated. Technical replicates used for the HiChIP, RNA-seq and ONP were all grown from the same aliquot of frozen (passage 4) LUHMES. HEK 293T (CRL-3216) were

purchased from ATCC and manufacturers recommendations were followed for growth and subculturing.

2.4.2 qPCR

qPCR assays were carried out on a Bio-Rad CFX384 real time system. Thermo-Fisher's TaqMan gene expression probes for *UBE3A-ATS* (FAM hs01372957_m1) were used with PPIA (VIC hs99999904) as the housekeeping gene. To calculate values the $2^{-\Delta\Delta Ct}$ method was used where ΔΔCt was calculated by subtracting ΔCt of control (*PPIA*) from ΔCt of experimental (*UBE3A-ATS*). Postmortem human cortex (#1406) was obtained from the Maryland Tissue Bank.

2.4.3 10x linked-read sequencing

High molecular weight DNA from LUHMES was isolated for 10x linked read sequencing using the Qigen Puregene Cell Kit (Cat. No. 158767) and the manufacturer's protocol was followed. Genomic DNA was adjusted to a concentration specified by the manufacturer and loaded on a Chromium Genome Chip. Whole genome sequencing libraries were prepared using Chromium Genome Library & Gel Bead Kit v.2 (10X Genomics, cat. 120258) and Chromium controller according to manufacturer's instructions with one modification. Briefly, gDNA was combined with Master Mix, Genome Gel Beads, and partitioning oil to create Gel Bead-in-Emulsions (GEMs) on a Chromium Genome Chip. The GEMs were isothermally amplified and barcoded DNA fragments were recovered for Illumina library construction. The post-GEM DNA was quantified prior using a Bioanalyzer 2100 with an Agilent High sensitivity DNA kit (Agilent, cat. 5067-4626). Prior to Illumina library construction, the GEM amplification product

was sheared on an E220 Focused Ultrasonicator (Covaris, Woburn, MA) to approximately 375 bp (50 seconds at peak power = 175, duty factor = 10, and cycle/burst = 200). Then, the sheared GEMs were converted to a sequencing library following the 10X standard operating procedure. The library was quantified by qPCR with a Kapa Library Quant kit (Kapa Biosystems-Roche) and sequenced on a partial lane of NovaSeq6000 (Illumina, San Diego, CA) with paired-end 150 bp reads. Longranger under vcmode gatk using hg19 as reference was used to bioinformatically process the data. Loupe browser 2.1.2 was used to view haplotype blocks.

2.4.4 RNA-seq

All replicates for RNA-seq were harvested from the same passage and time point as those used for the HiChIP, and ONP sequencing. RNA was isolated using Qiagen RNeasy Plus Mini Kit (Cat. No. 74134). To capture non-coding RNAs, expression was studied after ribosomal RNA depletion. Strand-specific and dual-barcode indexed RNA-seq libraries were generated from 450 ng total RNA each using the Kapa RNA-seq Hyper kit (Kapa Biosystems-Roche, Basel, Switzerland) and both the QIAseq FastSelect–5S/16S/23S ribodepletion and FastSelect rRNA Plant reagents (Qiagen, Hilden Germany) in combination, following the instructions of the manufacturers. The fragment size distribution of the libraries was verified in an automated electrophoresis platform on the TapeStation (Agilent, Santa Clara, CA). The libraries were quantified by fluorometry on a Qubit instrument (Life Technologies, Carlsbad, CA) and pooled in equimolar ratios. The pool was quantified by qPCR with a Kapa Library Quant kit (Kapa Biosystems) and sequenced on an Illumina NovaSeq 6000 (Illumina, San Diego, CA) with paired-end 150 bp reads. Results were processed using Babraham Bioinformatics Trim Galore!, STAR, Samtools, and MultiQC using gtf file GRCh38.109 (Dobin et al., 2013; Danecek et al.,

2021; Ewels et al., 2016). Differential gene expression and visualization was performed with lima-voom in R (Smyth, G. K. 2004, Law *et al* 2014). All differentially expressed genes with an FDR < 0.05 were used as input to https://maayanlab.cloud/Enrichr/ for GO enrichment analysis. For visualization in UCSC genome browser pileup bam files were concatenated from all replicates for each condition. CrossMap was used to liftover coordinates to hg19 using UCSC chain files (Zhao et al., 2013). Strands were split, values were transformed by LOG (ln(1+x)) and we used a max range of 7 to highlight smaller peaks.

2.4.5 HiChIP

Technical replicates for LUHMES neurons and undifferentiated cells were harvested and flash frozen. Chromatin was fixed with disuccinimidyl glutarate (DSG) and formaldehyde in the nucleus. Fixed chromatin was digested *in situ* with micrococcal nuclease (MNase) and then extracted upon cell lysis. Chromatin fragments were incubated with the CTCF antibody overnight for chromatin immunoprecipitation. The antibody-protein-DNA complex was pulled down with protein A/G-coated beads. Chromatin ends were repaired and ligated to a biotinylated bridge adapter followed by proximity ligation of adapter-containing ends. After proximity ligation, crosslinks were reversed, the DNA was purified from proteins and converted into a sequencing library. The sequencing library was generated using Illumina-compatible adapters. Biotin-containing fragments were isolated using streptavidin beads before PCR enrichment of the library. The FitHiChIP pipeline was used to analyze the data using 5 kb bins with the stringent settings at FDR:0.05, merged nearby loops (Sourya *et al* 2019). To identify loop interactions unique to a single condition, data were also analyzed using a less exclusive modified workflow where GenomicRanges and macs2 were used to first filter out only the peaks present

in two or more replicates before running through FitHiChIP using the loose "all to all", merged nearby loops and FDR:0.1 settings (Zhang et al., 2008, Lawrence et al., 2013). Loops were categorized as shared or unique using bedtools pairToPair with -type both imposed to ensure that both loop anchors were shared (Quinlan & Hall, 2010). Reads were initially mapped to GRCh38 and long-range interaction files were plotted in the WashU epigenome browser. To allow viewing of loops in the UCSC genome browser together with previous HiChIP assays, JASPAR scores and other useful tracks, coordinates were lifted to hg19 using the liftOver utility.

2.4.6 4C

We used 3 technical replicates for each viewpoint (VP) and condition harvested from a separate LUHMES cell thaw passage 4 with 10 million cells each. The 4C protocol was adapted from Krijger *et al* (2020) with the following modifications: Invitrogen MagMax DNAbinding beads (Cat. No. 4489112) were substituted for the Nucleomag beads. Primary restriction enzyme digest was performed using DpnII (Cat. No. R0543S) and secondary digestion with CviQI (Cat. No. R0639S) from NEB. Before sequencing a final cleanup using SPRIselect beads from Beckman Coulter (Cat No. B23317) was used. The fragment size distribution of the library pool was verified via micro-capillary gel electrophoresis on a Bioanalyzer 2100 (Agilent, Santa Clara, CA). The pool was bead cleaned twice to remove the adapter-dimer at 129 bp. Then, the library was quantified by qPCR with a Kapa Library Quant kit (Kapa Biosystems/Roche, Basel, Switzerland). The library was sequenced on one flow cell of Aviti sequencer (Element Biosciences, San Diego, CA) with single-end 150 bp reads. Pipe4C was used for the initial bioinformatic analysis followed by peak calling with PeakC using the default settings aligned to hg19 (Krijger et al., 2020; Geeven et al., 2018; Davies et al., 2015). Viewpoint (VP) primers

included: *SNRPN* VP reading primer (FP) 5'-TGTAATCCCAACACACTGG-3' and non-reading primer (RP) 5'-TGTTGTCTCTCATTTTCCTCA-3'. For the *PWAR1* VP FP 5'-TCATAGCTGAAACCATGAGA-3' and RP 5'-TAGACGAACATTGCTGTGAC-3' were used. For the *UBE3A* viewpoint FP 5'-ACCATCTTGGGAGACACAC-3' and RP 5'-TCCTCATCTTGGTGGTAAAG-3' were utilized.

2.4.7 Oxford Nanopore Sequencing

We used 2 technical replicates from passage 4 for each condition that were from the same harvest as the HiChIP and RNAseq. Flash frozen cultured cell pellets containing 5 million cells were used for high molecular weight genomic DNA (gDNA) isolation. Two ml of lysis buffer containing 100 mM NaCl, 10 mM Tris-HCl pH 8.0, 25 mM EDTA, 0.5% (w/v) SDS and 100µg/ml Proteinase K was added to the frozen cell pellet. Reaction was incubated at room temperature for a few hours until the solution was homogenous. The lysate was then treated with 20µg/ml RNase A at 37^0C for 30 minutes and cleaned with equal volumes of phenol/chloroform using phase lock gels (Quantabio Cat # 2302830). The DNA was precipitated by adding 0.4X volume of 5M ammonium acetate and 3X volume of ice-cold ethanol. The DNA pellet was washed twice with 70% ethanol and resuspended in an elution buffer (10mM Tris, pH 8.0). Purity of gDNA was accessed using NanoDrop ND-1000 spectrophotometer. DNA was quantified with Qbit 2.0 Fluorometer (Thermo Fisher Scientific, MA). Integrity of the HMW gDNA was verified on a Femto pulse system (Agilent Technologies, Santa Clara, CA) where majority of the DNA was observed in fragments above 100 Kb. Sequencing libraries were prepared from 1.5µg of high molecular weight gDNA using the ligation sequencing kit SQK-LSK114 (Oxford Nanopore Technologies, Oxford, UK) following instructions of the

manufacturer with the exception of extended incubation times for DNA damage repair, end repair, ligation and bead elutions. 30 fmol of the final library was loaded on the PromethION flowcell R10.4.1 (Oxford Nanopore Technologies, Oxford, UK) and run was set up on a PromethION P24 device using MinKNOW 22.12.5. To improve the yield, the flow cell was washed with a flow cell wash kit EXP-WSH004 (Oxford Nanopore Technologies, Oxford, UK) at approximately 24 and 48 hrs. after the start of the run and the fresh library was loaded. Basecalling was performed after the run using guppy 6.5.7. For the non-phased methylation data 5mc-5hmc calls were also made with guppy and the ONP pipeline modkit-pileup was followed with the --cpg option using GRCh38.p13 fna, and then the UCSC liftOver tool was used to convert to hg19 coordinates. For the phased data minimap2 was used for alignment to hg19 (Li, 2018). f5c was used to call-methylation using the --pore r10 option (Gamaarachchi et al 2020). Clair3 was used to call variants using model r1041_e82_400bps_sup_g615 and whatshap was used for phasing (Zheng et al., 2022; Martin et al., unpublished). Nanomethphase was used to phase the methylome and DSS was used for differential methylation analysis (Akbari et al., 2021; Wu & Feng, 2020). Differentially methylation analysis was performed comparing paternal vs maternal, in both cell types using two replicates for each condition. DMRs were called using the paternal allele as the treatment group and the maternal as the control from the same sample. As a result, positive values represent regions where methylation is higher in the paternal allele while negative represents higher methylation on the maternal allele, with those values indicating their percent differences. In addition, differential methylation analysis was performed comparing the paternal alleles in neurons and undifferentiated LUHMES. bedGraphToBigWig was used to prepare visualization for UCSC genome browser (Kent et al., 2002).

2.5 Supplemental figures

2.5.1 Figure S1: Genome wide differential loops and expression

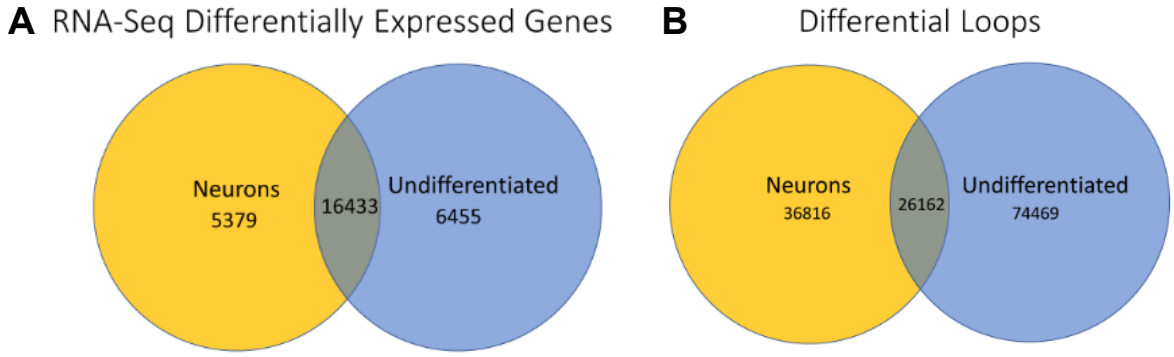

Figure S1: A. Genome wide differential expression and differential loops **A.** Genes expressing differential expression <FDR: 0.05 upregulated in Neurons (5,379) and undifferentiated LUHMES (6,455), no significant differences were seen in 16,432 genes. **B.** Differential loops determined by less stringent analysis described in HiChIP methods.

2.5.2 Figure S2: Integration of non-phased methylation data

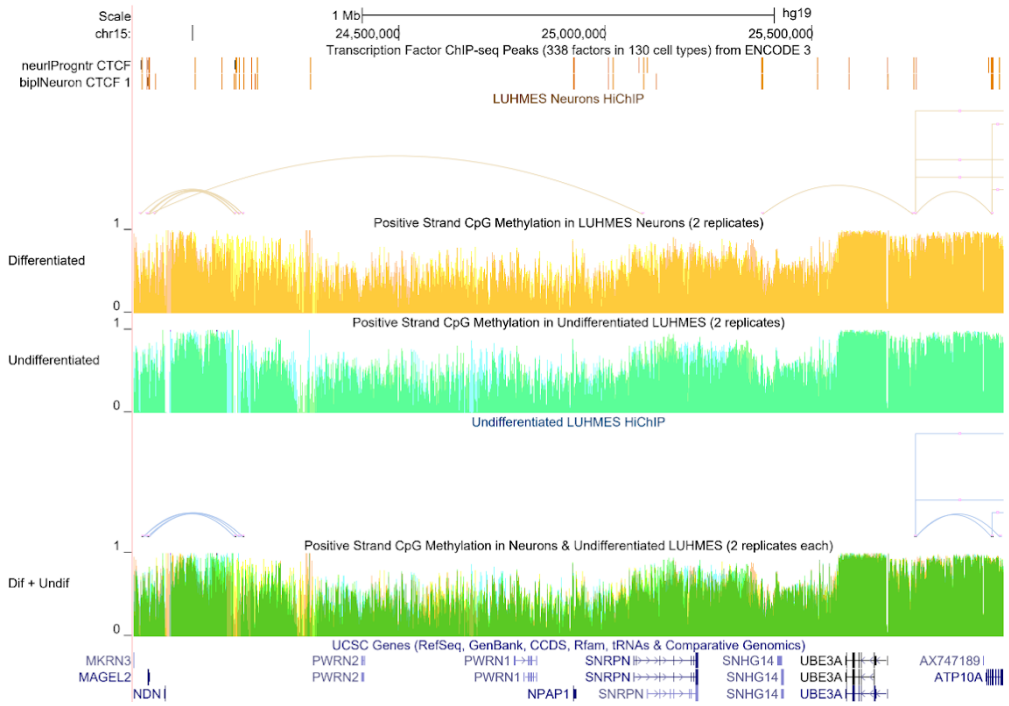

Figure S2: Integrated view of LUHMES HiChIP, and CpG methylation from the positive strand (hg19). UCSC tracks display two overlaid replicates in different colors for differentiated and undifferentiated cells. Additionally, a combined track labeled Dif + Undif overlays all four samples (chr15:23,832,378-25,962,021)

2.5.3 Figure S3: DMRs for paternal neurons vs undifferentiated LUHMES

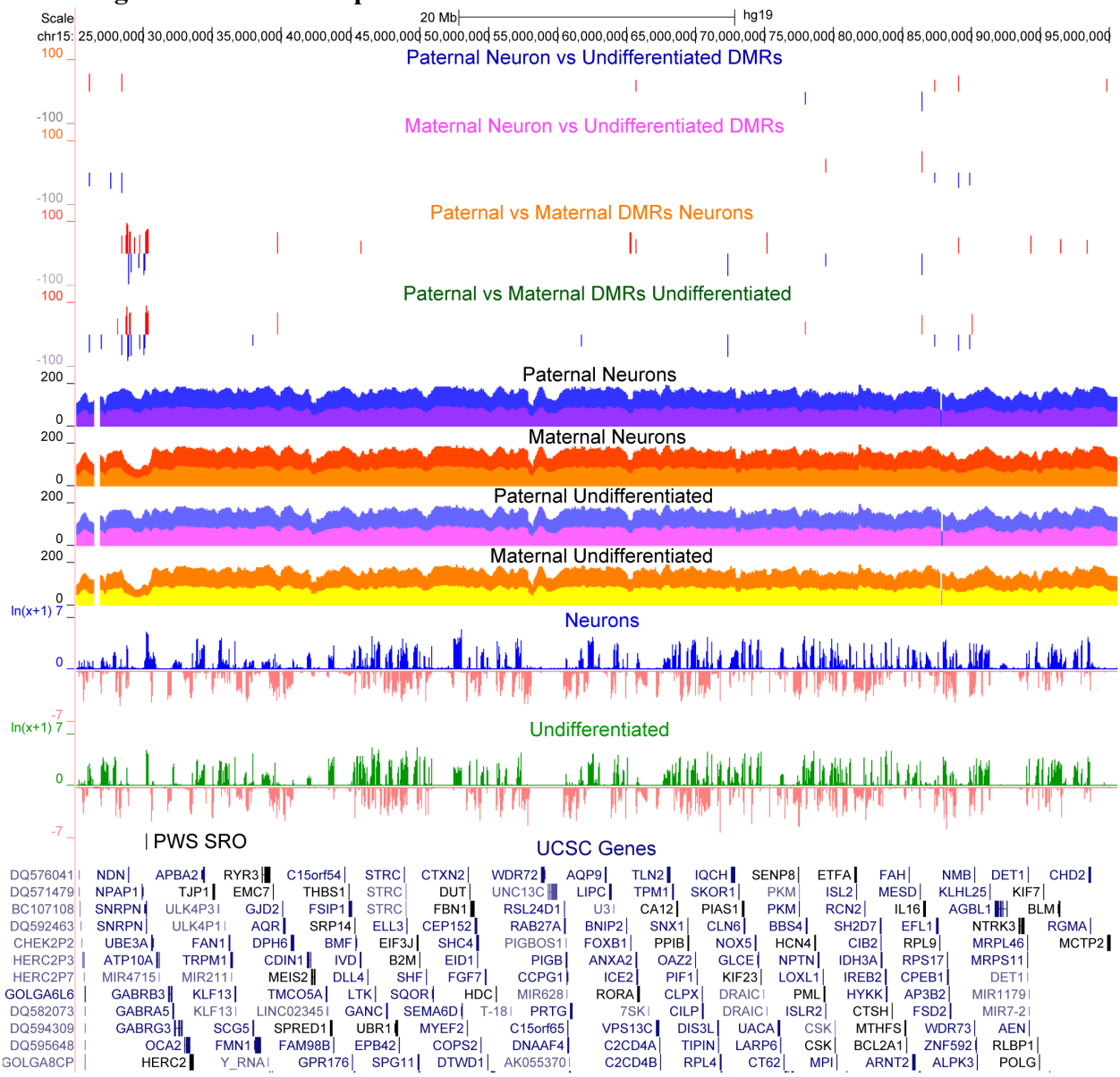

Figure S3: A. chr15:20,141,900-95,560,739 shows few DMRs when looking a large portion of chr15 when comparing the paternal allele in neurons vs the paternal allele in undifferentiated cells. Few were also seen when comparing the maternal allele in neurons vs the maternal allele in undifferentiated cells. We can also view how clustered the paternal vs maternal DMRs are for both cell types in the AS/PWS region.

2.5.4 Figure S4: HiChIP contact matrix

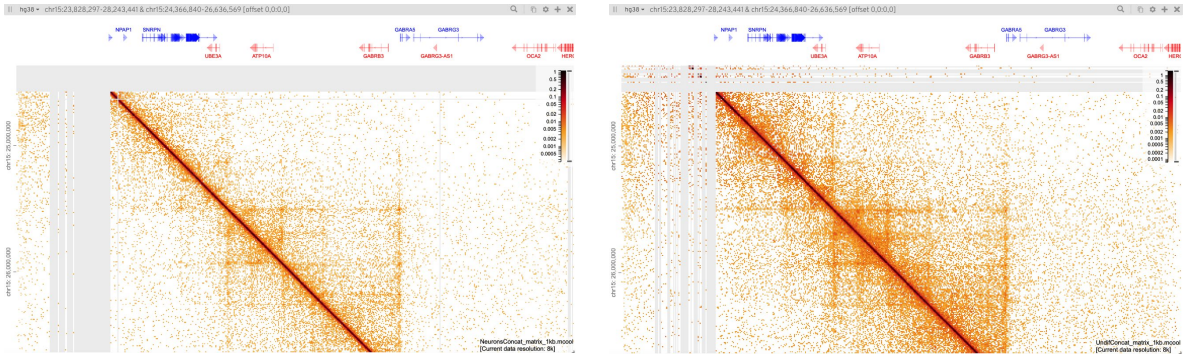

Figure S4: HiChIP Contact matrix generated using HiGlass (Kerpedjiev et al., 2018) of region circumjacent to *UBE3A* in **A.** Neurons **B.** Undifferentiated LUHMES

2.5.5 Figure S5: Assigning allele parentage using 10x-linked reads

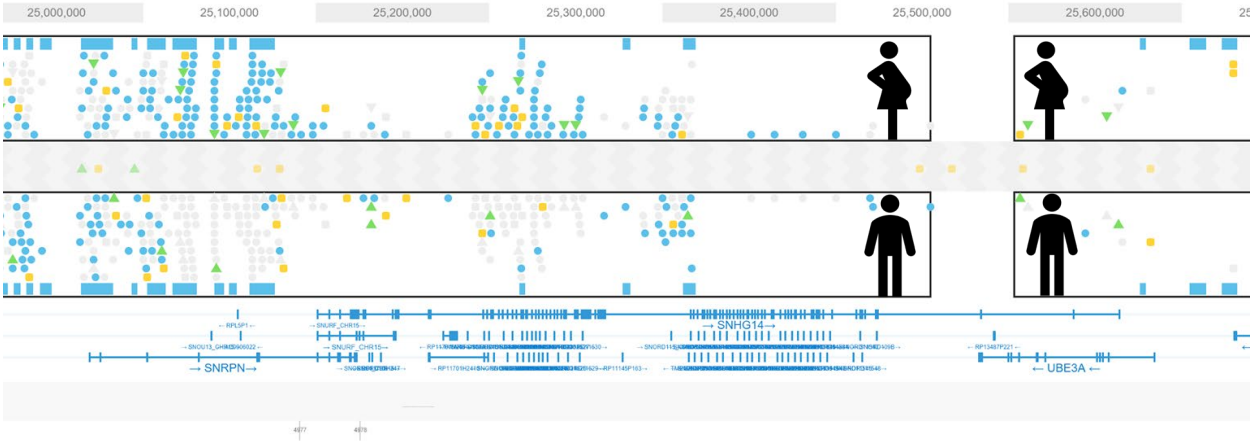

Figure S5: LUHMES cells 10x linked read phasing showing SNPs grouped into four unlinked haplotype blocks. Parentage was determined by phased SNPs from 10x reads combined with Oxford nanopore long read sequencing. Blue circles = Heterozygous SNPs, Green Triangles = Heterozygous insertions, Orange squares = Heterozygous deletions.

2.6 Author's Contribution

Osman Sharifi provided substantial bioinformatic assistance, critical advice and helped in 4C library clean up. Nicholas Heath contributed in the 4C design and experiment. Daniela Soto provided substantial bioinformatic support and contributed to the analysis and visualization of linked-read SNP data. J Antonio Gomez aided in LUHMES sample harvest. Aron Mendiola contributed bioinformatically to the differential expression analysis. Dag Yasui performed the HMW DNA collection for the linked-read analysis and provided critical advice. Henriette O'geen, Ulrika Beitnere, Marketa Tomkova and Greg Dillon provided critical advice. David J Segal and Janine LaSalle conceived the study, contributed substantially on study design, data interpretation and the manuscript for this project.

The library preparations and sequencing were carried out at the UC Davis Genome Center DNA Technologies and Expression Analysis Core, supported by NIH Shared Instrumentation Grant 1S10OD010786-01. This work was supported by NIH NIGMS Grant T32 GM007377, NIH R01HD098038, Foundation for Angelman Syndrome Therapeutics (FAST) and Biogen.

Chapter 3: Expanded discussion and conclusions

In this comprehensive study, we characterized and integrated genome-wide DNA methylation with CTCF loops and RNA expression of LUHMES cells to shed light on their relationship, with a particular emphasis on evaluating their potential as a model for AS. Our findings provide a robust foundation for the use of this cell line as a valid model, using the undifferentiated as a non-neuron and differentiated as a neuron, especially in the context of allele specific imprinting.

As in previous studies we observed the morphological transition of LUHMES cells from an epithelial-like state to a neuron-like state characterized by long neurites resembling midbrain axonal networks (Scholz et al., 2011). Our transcriptional data demonstrating the progressive induction of *UBE3A-ATS* during neuronal differentiation confirms that LUHMES neurons exhibit this molecular signature of mature neurons more rapidly than AS iPSCs (Chamberlain et al., 2010).

RNA-seq analysis revealed a profound shift in the transcriptional landscape of LUHMES cells upon differentiation. Differentiated LUHMES demonstrated an upregulation of genes pivotal to neuronal function and the downregulation of genes associated with the cellular processes seen in dividing cells which bolsters their validity as a non-neuron and neuron comparative model. The differential expression of genes within the AS/PWS locus, especially the significant upregulation of genes like *MAGEL2, SNRPN, SNHG14*, and several snoRNAs within the *SNORD116* cluster, provides a compelling case for the relevance of LUHMES neurons in 15q11.2-13.3 disorder research.

Only the plus strand transcriptional profile was distinct between these two cell states. In neurons we saw an increase in plus strand transcription begin before *SNRPN* with an abrupt decrease after *PWAR1* in undifferentiated cells (Figure 8A; 9C). Transcription continues in neurons with a marked increase over the *UBE3A* gene body and again downstream of it. This region was also highly methylated on all haplotypes (Figure 9D).

The chromatin architecture, as revealed by our HiChIP analysis, offers novel insights into the dynamic interplay of genes within the AS/PWS locus during neuronal differentiation. The identification of neuron specific long-range CTCF loops, especially those involving key genes like *MAGEL2, SNRPN, PWAR1,* and *UBE3A*, underscores the intricate regulatory mechanisms at play during neuronal imprinting. Although our 4C experiments were not specific for CTCF loops, they still validated our HiChIP findings, emphasizing the specificity of these chromatin interactions in neurons versus undifferentiated states. The JASPAR Core 2022 CTCF binding motifs showed the expected convergent orientation. Although our HiChIP analysis was limited by the absence of phasing in the future this could be possible for this cell line now that SNP and indel parentage has been assigned.

Our methylation analysis suggests that there are no dramatic differences between the differentiated and non-differentiated LUHMES at the level of global methylation landscape at the 15q11-q13 region, in fact few paternal neuron vs undifferentiated DMRs were seen throughout chromosome 15. Our use of long read sequencing allowed us to phase the methylome and revealed a stark contrast between the paternal and maternal alleles. By

integrating genome-wide DNA methylation with the rest of our data we can shed light on the relationship between methylation, CTCF loops and expression taken from the same sample, possibly reducing batch effects. We also see the value in viewing our LUHMES data in the context of any other region, transcription factor binding site or track on the UCSC genome browser. We hope this simplifies the use of this cell line to study imprinted loci and that our data can aid in formulating novel ideas of how *UBE3A* silencing is regulated.

A distinct pattern of CpG methylation landscape was observed in an allele specific manner in both neurons and undifferentiated cells (Figure 8A). Narrow regions of maternal hypermethylation are seen within broader regions of paternal hypermethylation, a pattern that was previously observed by whole genome bisulfite sequencing in postmortem PWS, AS, and Dup15q brain samples (Dunaway et al., 2016). When comparing the paternal allele to the maternal allele this regions appears to be a hotbed for DMRs particularly between both loops (Figure 8A;B). Of particular interest was the observation that a large cluster of DMRs are present outside of the neuron specific loops we observed. Even though this cluster was seen in both cell types we do see some variations between them although this was not seen in differential methylation analysis using paternal neuron vs paternal undifferentiated LUHMES. Hypermethylation of the paternal allele in this cluster was also associated with active transcription as is usually seen in the gene bodies (Rauluseviciute et al., 2020).

For both cells types we observed paternally hypomethylated regions overlapping with the *MAGEL2* and *NDN* promoters with a region of hypermethylation in between (Figure 9A). A hypomethylated region overlapping the *SNRPN* interacting bin was exclusive to neurons (Figure

9B). However, a downstream hypomethylation at a CpG island close to the PW SRO is associated with the beginning of *SNRPN* transcription in undifferentiated cells. In contrast we found a hypermethylated DMR exclusive to undifferentiated LUHMES upstream of the *PWAR1* anchor after which transcription decreases dramatically. All three other loop anchors were adjacent to increased plus strand transcription in neurons (Figure 8;9). The loop anchor close to *UBE3A* is located about 61 kb 5' of the promoter, which is a biallelically hypomethylated CpG island. A differential methylation analysis was also performed comparing paternal alleles from neurons vs undifferentiated cells, however this resulted in few occurrences throughout the entire chromosome (Figure S3).

Since CTCF binding is known to be inhibited by methylation at its binding site, we thought we might find hypomethylated DMRs that directly overlap the motifs within the neuron specific loop anchors assuming that these loops are exclusive to one allele. We found no DMRs in direct overlap, but we identified hypomethylated DMRs within a few kb of the *MAGEL2* and *SNRPN* motifs, which also aligns with CpG islands. Interestingly though, the DMR close to *PWAR1* was unique in that it was hypermethylated and it was only seen in the undifferentiated LUHMES. This could point to the criticality of the *PWAR1* binding site as part of the boundary region where the lncRNA stops in non-neurons. It's possible that this could be the key to regulating the imprinting of *UBE3A* in neurons, but more research needs to be done.

The *MAGEL2-SNRPN* loop was not as frequently present in the LUHMES neurons as the loop closer to *MAGEL2* that was detected in both cell types. However, we are confident that they are real interactions occurring in a subset of cells at the time of harvest since they persisted even

when using stringent filtering to reduce for false positives. It was observed again in our 4C experiment from a separate thaw, passage and time point. Furthermore, even if the interaction is not that frequent, it could still be enough to enhance the transcription of the *SNRPN* lncRNA. We can only speculate in what direction this loop affects transcription. It could be going from *MAGEL* to *SNRPN* or vice versa. This interaction could also be related to the fact that the PWS ICR controls paternal allele expression of *MAGEL2* (Horsthemke & Wagstaff, 2008; Soejima & Wagstaff 2005). However, its location is over 100 kb away from where the *SNRPN* loop anchor lands. The mechanism behind this would require further investigation but the enhancement could be due to positioning at proximity in 3-dimensional space inside an active compartment with recruitment of transcription factors.

One of the prevailing models for how *UBE3A* is silenced by *UBE3-ATS* is the collision model which proposes that RNA polymerase II from the convergent transcripts disrupt expression, leading to the repression of one or both genes (Mabb et al., 2011).

Our integrated analyses suggest a fresh model of how the lncRNA that begins at *SNRPN* could regulate *UBE3A-ATS* expression and therefore silence the paternal allele in neurons. As was shown in previous studies *MAGEL2* appears to only be expressed from the paternal allele. Since *MAGEL2* shares enhancer elements with the nearby gene *NDN* it is possible that it is acting in the same manner or recruiting transcription factors to *SNRPN* by way of this novel loop (Stelzer et al., 2016; Fishilevich et al., 2017). This *MAGEL2* to *SNRPN* chromatin interaction or in combination with the loop from *PWAR1* to *UBE3A* could play a role in how paternal imprinting is regulated on the paternal allele in neurons.

The *PWAR1-UBE3A* loop might aid in the silencing on the paternal allele by allowing the extension of the lncRNA past the boundary region. This could because it insulates the region from transcription from the minus direction making it more permissive from the *SNRPN* lncRNA positive direction. Although, transcription from both undifferentiated and differentiated LUHMES from the minus strand looks similar. In paternal neurons one of these loops or the combination of both could perturb the delicate balance of where RNA polymerase II collision occurs between these convergent transcripts (Figure 9).

In its undifferentiated state the rate of transcription from the *SNRPN* lncRNA could be that which makes it collide at the boundary region which includes *PWAR1* since transcription is known to stop there in undifferentiated cells and was observed in our RNA-seq analysis. With *MAGEL2* acting as an enhancer the *SNRPN* lncRNA, that may cause the collision to occur further downstream not allowing the expression of *UBE3A* (Figure 10). However, we should use caution in assuming causality in this direction since the PWS imprinting center is known to regulate *MAGEL2* expression (Li et al., 2021).

3.2Conclusions

Our findings unequivocally establish LUHMES cells as a potent model for studying the intricate molecular and epigenetic mechanisms of imprinting at the 15q11-q13 locus, especially in the context of disorders like AS. To facilitate this, we have additionally created UCSC genome browser track hubs that can be used in the future to provide an integrated genome wide insight into the epigenetic landscape underlying neuronal differentiation and the changes that occur in

methylation, CTCF loops and gene expression from the same sample. We also succeeded in assigning parentage to heterozygous SNPs in the AS/PWS locus for this cell line, opening avenues for allele specific knockouts, as well as genetic and epigenetic editing.

Two chromatin interactions were observed in LUHMES neurons but not in undifferentiated cells. Firstly, *MAGEL2* appears to interact with a region the overlaps a paternally hypomethylated DMR upstream of *SNRPN*, potentially augmenting its transcription. Secondly, a loop from *PWAR1* to a section downstream of *UBE3A* was identified, which might aid in its silencing on the paternal allele by allowing the extension of the lnRNA past the boundary region. Additionally, the *PWAR1* CTCF anchor binding site had a nearby hypermethylated paternal DMR exclusive to undifferentiated LUHMES.

These newly identified phased DMRs and chromatin interactions could have a significant influence on the transcriptional landscape of the 15q11-q13 locus. They may represent the necessary and or sufficient conditions for favoring progression of transcription from *SNRPN* through *UBE3A-ATS* consequently, silencing the paternal *UBE3A* allele in neurons.

Further studies are warranted to determine the functional implications of these chromatin interactions and DMRs in the context of *UBE3A* regulation and AS. One way could be through ablation of the binding sites but a more nuance approach could be through epigenetic editing directed towards these DMRs to determine if *UBE3A-ATS* expression is disrupted (Chapter 4). By assigning parentage to heterozygous SNPs in the AS/PWS locus we facilitate the use of this cell line for more allele specific research including knockouts (Chapter 4).

Our findings underscore the crucial role of chromatin looping and methylation in dictating the tissue-specific imprinting pattern at the *UBE3A* locus and how this cell line could provide novel insights into potential epigenetic strategies for unsilencing the paternal allele as a therapeutic approach for AS.

Epigenetic editing could be directed towards these CTCF binding sites and paternal DMRs to see if *UBE3A-ATS* expression is perturbed. However, performing allele specific methylation would be challenging therefore a maternal *UBE3A* knockout will be necessary to be conclusive (Chapter 4). In the next chapter, we lay the groundwork for achieving this path using our heterozygous SNP data as well as exact protocols followed.

Chapter 4: Future direction and appendix

In this work we focused on describing the LUHMES cell culture model focusing on the AS locus. Since LUHMES are not derived from neoplasia and presents an epigenome that's closer to wild type. Using the undifferentiated LUHMES as a model for non-neurons and differentiated LUHMES to represent neurons reduces confounding variables and approximates a more direct comparison.

Several studies published by the Segal lab using an AS murine model have been successful at unsilencing paternal *Ube3a* using epigenetic editing tools targeted to the *Snrpn* promoter (Bailus et al., 2016). Epigenetic editing offers great potential as a therapeutic platform because it does not alter the genome permanently. However, our lab's previous strategy also affects critical paternally expressed genes upstream of *PWAR1*. The design of an improved strategy for *UBE3A* unsilencing to treat AS is partly limited by the lack of understanding of how tissue-specific imprinting is controlled at this locus.

Artificial transcription factors (ATFs) are engineered chimeric proteins that consist of a DNA binding domain fused to an effector domain that can activate or repress transcription (Bailus & Segal, 2014). The Segal lab has had some success in blocking the expression of the *Ube3a-ats* and unsilencing the paternal *Ube3a* utilizing ATFs with repressing effector domains targeted to the promoter of *Snrpn* (Bailus et al., 2016). However, as a therapeutic approach it would be more favorable to target a region further downstream in order to leave transcription of the snoRNA's intact due to their physiological relevance. dCas9 directed methylation of CTCF

binding with subsequent changes in gene expression has already been accomplished (Tarjan et al., 2019).

An imprinted locus with an intact but silent *UBE3A* provides a unique opportunity for an epigenetic intervention. Using our newly characterized model we sat out to pilot the use of epigenetic editing in this cell line. One of the first hurtles we encountered was the fact that *UBE3A* is expressed from the maternal allele in this line. With the use of our heterozygous SNP data, we were able to design a few different strategies for creating a maternal knockout. We first optimized a transient transfection protocol for delivery of our gene editor using the Neon electroporation system. After achieving this *UBE3A* knockout we can create two stable cell lines. One expressing an ATF dCas9-DNMT3AL for methylation and dCas9-TET1 for demethylation.

Since LUHMES are transfected during passage before placing in differentiation media our plasmid would not be expressed by day 3 however *UBE3A-ATS* does increase significantly until after day 4. Because of this and to be able to differentiate between keeping a loop from occurring during differentiation or undoing a loop we felt an inducible promoter was warranted. However, LUHMES already use a TET-off system for the myc gene to differentiate into neurons. Tamoxifen inducible promoters are an option but could interfere with myc gene expression therefore a cumate switch was chosen along with a GFP reporter. We chose a piggyBac backbone making the size of our plasmids around 11 kb so we expected our transfection efficiency to be low, but we planned on compensating by using FACS sorting. However, we

found that when single cell sorted in 96 well plates LUHMES tended not to survive. We increased the sorting to 10 cells a well with a plan to perform dilutions further down the line. When placing 10 cells per well LUHMES tended to have about 3 survivors. We believe that the optimal way to get one survivor per well might be by placing 3 cells per well or subsequent serial dilutions to reduce the possibility of a mixed population of cells.

For subsequent stable transfection of the inducible methylation and demethylation cell lines we designed gRNAs plasmids to target the neuron exclusive CTCF binding sites. We used a strategy of placing pairs of gRNAs for both anchors of a loop. By using RFP and BFP we would be able to segregate the transfected cells in spectrums that don't overlap with the GFP included in the stable lines or with each other. We felt that we could simplify our plasmid and keep the size to a minimum by not using an inducible promoter for the gRNA with the tradeoff that these would be constitutively made after integration but should be quickly degraded. We can then determine if a set or multiple sets of gRNAs are affecting the epigenetic landscape of LUHMES neurons and the expression of *UBE3A-ATS* and *UBE3A*. Here we lay out our design strategy for achieving these goals.

4.1 Creating a LUHMES maternal knockout

To create a maternal knockout of LUHMES several gRNA strategies were examined. To cut at a known heterozygous SNP on maternal allele upstream of *SNRPN*: CTACTCATGACAGAAAGAAAAGg and a second cut downstream of *UBE3A* at another heterozygous SNP AAATACAGGcCTGGCACGTGGGG. As an alternative that is a bit closer but does not encompass *SNRPN*. One upstream to *UBE3A* known to contain a heterozygous SNP for the maternal allele GTTCCCCCAGATGGTGACCATgG and the other downstream of *UBE3A* on a PAM site TAGTCTCTCCCCCACGTGCCAGg. It was also surmised that a simple frameshift strategy for the exons had the most potential to succeed. Since Isoform I is the most abundant in human brain, we initially targeted with gRNAs AGCCGAATGTAAGTGTAACTTGG and GCTTCAATGTCGTCAGACTGAGG. We also created a gRNA for a region on Isoform I that is shared with Isoform 2 and 3 (KRAAK region after the TSS). CTACTACCACCAGTTAACTGAGG. We additionally examined the strategy of using a previously established gRNA AATGTAAGTGTAACTTGGTT (Sirois et al., 2020). However, we also examined using homology directed repair (HDR) with a single stranded oligo that changes the TSS in Isoform I into a stop codon (UAG). HDR works best if the donor strand is complementary to the non-target strand, and DNMT3AL works best at 36 bp PAM distal (before gRNA sequence) or 91 bp proximal to PAM (after NGG). Using those guidelines, we designed the HDR oligo, TTTGGAATATGGAATATTTTGCTAACTGTTTCTCAATTGCATTTTACAGATCAGGAGA ACCTCAGTCTGACGACATTGAAGCTAGtCGAtaGTAAGTGTAACTTGGTTGAGACTGT GGTTCTTATTT (Figure I). We do however realize that homozygous cuts only occur in a part

of the transfected population of cells but because of the exclusive expression of *UBE3A* from maternal neurons we could perform a simple qPCR to find the right clone (Figure II).

Figure I. Locations of KO sites and HDR on UBE3A

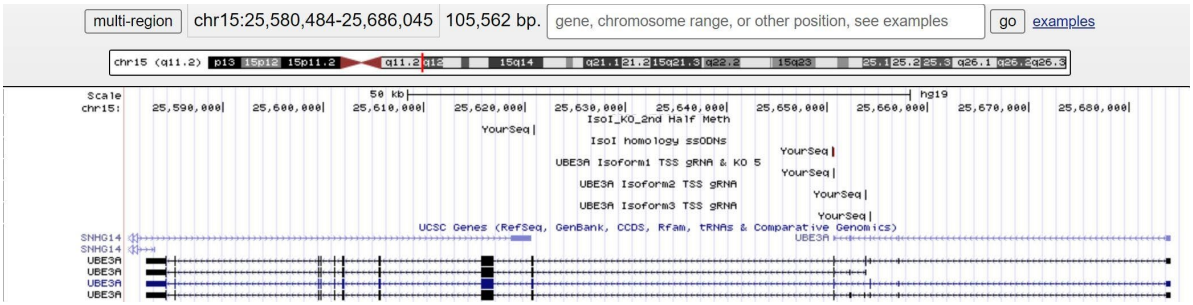

Figure I. UCSC genome browser view showing location of *UBE3A* Isoform TSS, gRNA for KO, and HDR repair template.

Figure II. Identifying heterozygous maternal knockout

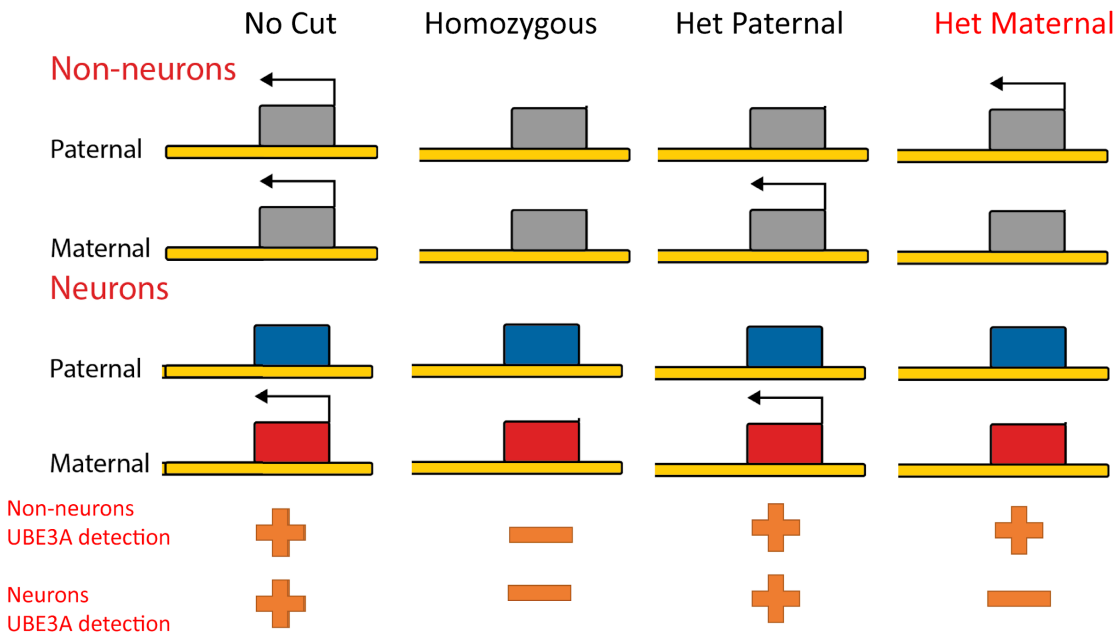

Figure II: Pattern of expression for *UBE3A* in undifferentiated LUHMES and neurons depending on if there was no KO, a homozygous KO, heterozygous paternal and the desired heterozygous maternal.

4.2 Epigenetic editing using stable cells lines

Using transient transfection has that disadvantage that plasmids will only reliably be expressed for about 72 hrs with even less time for gRNAs. We sought out to create a number of stable cell lines for implementation after establishing a maternal knockout. In order to be able to turn the system on after day 4 of differentiation, stable transfection with a cumate on system. We first designed a fusion protein consisting of a catalytically inactive Cas9 (dCas9) fused to a DNA methylator DNMT3AL and one fused to a demethylator TET1. We had success in using the piggyBac system in LUHMES before, so we used that as our backbone with a GFP fusion. The constitutively off cumate switch was integrated into the design since LUHMES already utilize a TET off system for differentiation.

Figure III. Cumate switch Piggy-Bac stable transfection plasmid design

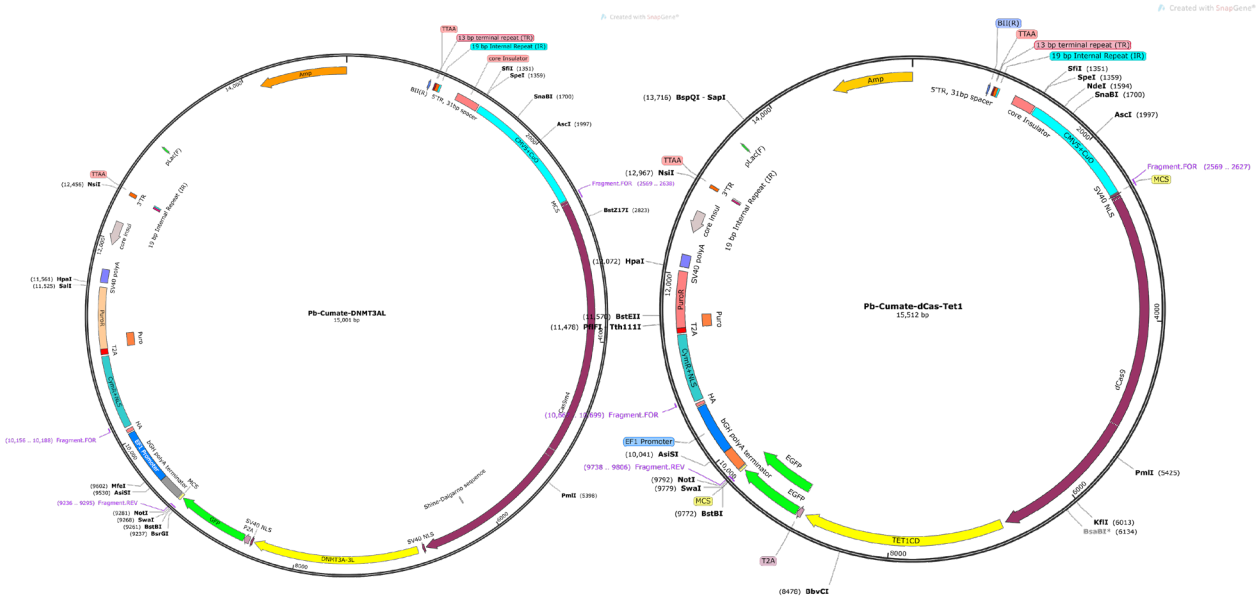

Figure III. Design for ATF including a constitutively off cumate switch controlling a dCas9 fused to either DNMT3AL or TET1 with a GFP reporter on a piggyBac backbone.

After these two stable cell lines are established, we would follow with a secondary stable transfection with our gRNAs directed to each pair of CTCF anchor points. Employing a combination of Red Fluorescent Protein (RFP) and Blue Fluorescent Protein (BFP) will allow FACS sorting for specific loops and gRNA multiplexing (Figure IV). When relaxing some of filtering criteria of FitHiChIP to an FDR:0.1 we saw some of the same loops we saw in our secondary bioinformatic analysis. Most notably one that land much closer to the *UBE3A* promoter.

The following guidelines were used to create gRNAs. PAM Should be ~27 bp upstream of target CpG for CTCF binding site (if it contains a CpG). Since gRNA is ~23 bp (with PAM) then the beginning of gRNA can be ~50bp upstream of CpG with farther than 30 being better than closer. Since our results show this to be the most interesting loop, an example of a gRNA design for the *PWAR1* to *UBE3A* loops was carried out. For the upstream *PWAR1* bin chr15:25,380,147-25,385,147 (hg19) we suspect that the CTCF motif involved is GATCCTGCAGAGGGAGCCA but since it has no CpG we aligned our target to its center resulting in ATGTTGAAGAGCGTTCCCTGTGG which is 38 bp away from the PAM site. To target the CpG closest to the motif GGGAGCCATAGCTAGTGTAGAGG was designed which is 36 bp away from the PAM. We also hypothesize that just directing dCas9 to the PAM closest to the CTCF peak could cause a steric hindrance to CTCF binding especially due to the large footprint of Cas9. To target the downstream anchor close to *UBE3A* chr15:25745147-25750147, if we direct our guide towards the center, we can use AGGGACAGACATTGTTTAAAGGG. However, to target the CTCF motif upstream to this we designed CTCACATTCCATCCAACAGATGG which is 43 bps away from the motif. Since there was

also a motif downstream, we designed GAGACACTTGTGAAGTTAGAAGG which is 25 bps away. To possibly create steric hindrance, we used CTCACATTCCATCCAACAGATGG since this partially covers the binding motif.

Figure IV. gRNA for targeting CTCF binding sites

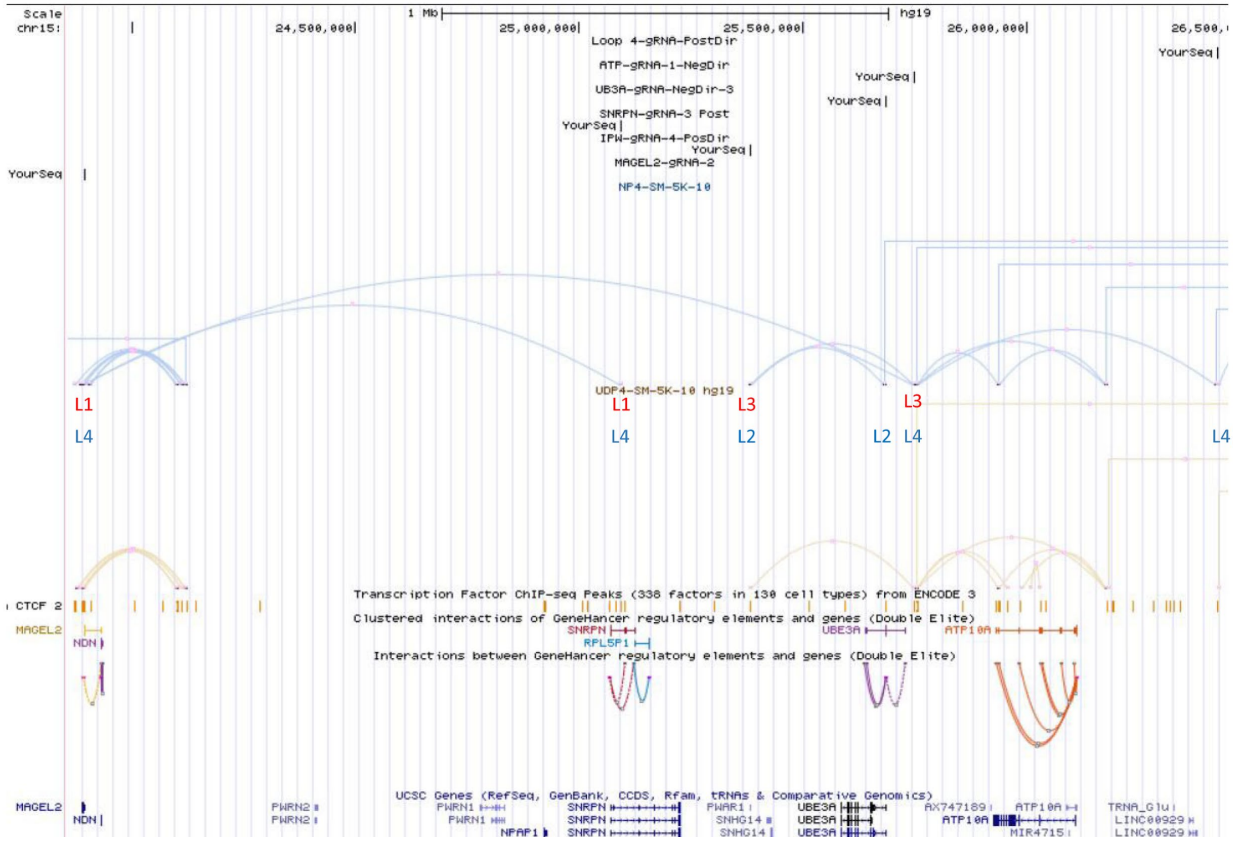

Figure V. UCSC genome browser view of neurons (NP4) and undifferentiated cells (UDP4) along with gRNA pairs designed. Those in red were designed with RFP and those in blue in BFP.

4.3LUHMES transient transfection

LUHMES cells have been transfected with lentivirus successfully in several studies (Lotharius et al., 2002; Schildknecht et al., 2013). We had some success using chemical transfection with the Mirus X2 system, but we sat out on creating a protocol with the Thermo Fisher Neon electroporation system.

4.3.1 Electroporation with Neon Transfection System Protocol

Notes:

- This protocol all comes down to speed, the buffer is toxic to cells, and so is being left in a small amount of media in a tube. Do 1 condition at a time, no more than 1 plate at a time. Tips can be re-used twice for the same condition, do not use 3 or more times
- For first day use media with no antibiotics
- Higher success rate when starting differentiation immediately after electroporation
- 100 ul tips work just as well regardless of Neon manual and avoids multiple electroporations per well. 6-12 well plates work best, this protocol is for 100 ul tips
- Total DNA cannot exceed 10 % of tip volume (<10 ul), max thermos recommendations for plasmid <5kb (5-30 ug), need more for larger plasmids
- Plasmids <5kb transfect only about 1-5% of cells so usually paired with FACS sorting
- Adjust quantity of recommended DNA for 100 ul to what it would be in 115 ul (extra 15)
- Some research has shown that leaving larger plasmids in the tip for about 10 minutes after electroporation before putting in media may increase survival rates

Protocol:

1. Add 3 mL of Buffer E to neon tube and leave ~1mL of buffer R at room temperature
2. Prep complete growth media/ Dif media in plate well (no Ab) & maintenance T-75 flask
3. Label round bottom Eppendorf tubes one for spinning cells and one with same name but with -D added for DNA (example 1,2,3 and 1-D, 2-D, 3-D)
4. Add plasmid + gRNA to -D tubes and mix (avoid creating foam or bubbles),
5. Set up electroporator and load conditions (Saved under Oran-Parameters-Voltage)
6. Split maintenance culture as usual right before electroporation, where 500-600 ul of 3 ml from 80% confluent cells is enough for 6 wells of a 6 well plate
7. Load max number of cells according to Neon protocols for that tip size 0.4 x 10^6 (Cell concentration is very important and too much is just as bad as not enough) in 100 ul. Limit to 2 Eppendorf tubes at a time. Adjust for extra R buffer (.574 x 10^6 for 113 ul)
8. Spin cells in both tubes at 300 g's x 5' to pellet
9. Add volume corresponding to 1.6 x 10^6 to maintenance T-75
10. Store the rest of resuspended cells in non nunc non coated T25 flask in incubator at 37 C
11. Aspirate media from 2 Eppendorfs that were just spun using P200, resuspended pellet in 113-180 ul of dPBS
12. Spin both tubes at 300 g's x 5', aspirate dPBS from only one of the tubes, resuspended it in about 113-115 ul of Buffer R, spin another 2 tubes with media+cells at the same time

13. Using P1000 set to 140 ul transfer R buffer + cells to -D tube containing DNA, mix gently
14. Using Neon tip, prime with R buffer and aspirate DNA+cells carefully avoiding bubbles, this step is crucial since it will cause arching and kill cells, can keep trying for up to 15'
15. Electroporate using the one of Neon neuro cell conditions
    a. T98-70 best for Undif and Dif SH-SY5Y have worked best
    b. Protocols with one pulse work better, other ones that have worked well are
        i. T98-80, SK-85, SH-70, SK-80, U87
16. Eject contents directly into warm differentiation media, use a p1000 several times to ensure no clumps of death, agitate plate, go on to second tube of set, place plate back in incubator, bring out T25 with cells, mix with P1000, prep next pair of tubes, aspirate media from just spun (2nd set of tubes), resuspend in dPBS, repeat spin with both sets

Repeat until all 6-12 tubes are done.

Figure V. Electroporation of LUHMES neurons

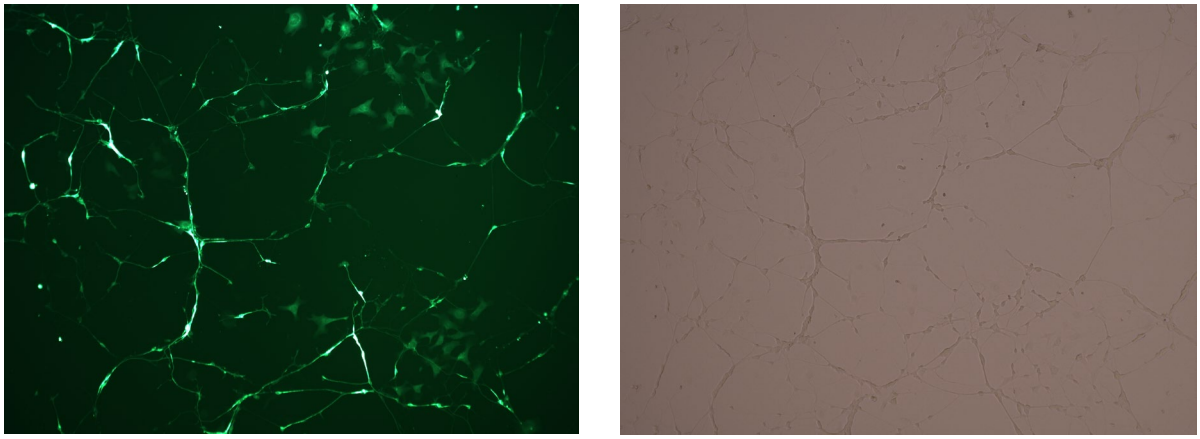

Figure V. 3 Day LUHMES Neurons using 3 µg of GFP Plasmid Neon Electroporation Protocol for T98-80 Cells **A.** View under EVOS imaging system using GFP cube **B.** Same region under brightfield.

4.3.2 Coating Flasks and Plates

Notes: Only use Nunclon delta treated plastics, no other brands or type of Nunclon (i.e Vita). This surface is hydrophilic and LUHMES will die after 1-3 days on any other surface.

Protocol:

1. Thaw a poly-L-ornithine and fibronectin aliquot at 37 C
    a. 50 µg/mL poly-L-ornithine (Sigma, Cat. No. P-3655-10MG)
        i. Start with 5 bottles (of 10mg) reconstitute each one with 2mL H2O (total 10 mL) and add to 989 mL H2O (can also use 1 bottle of 50mg)
        ii. If use 50mg bottle of Poly-O can reconstitute in 5 mL H2O, add to 994 mL H2O in stericup
    b. 1 µg/mL Bovine Fibronectin (MilliporeSigma, Cat. No. 341631-1MG)
        i. reconstitute 1mg with 1 mL of H2O then add to 999 mL of prepared 50mg of poly-o and filter through .22 um 1L stericup
    c. Make aliquots up to 42ml in a 50 ml tube and store at -20 C (overfilled tubes crack when freezing)
2. **Completely coat surface by tilting** (Add 7 ml per a T-75 flask, 2 ml per a well in a 6 well)
    a. If coating evaporates overnight and surface is exposed, it will not be usable
3. Place into 37 C incubator overnight (3-hour protocol for emergencies only, will stress cells)
4. Rinse twice with 10mL sterile filtered ddH20 and let air dry completely with open cap/lid, flasks upright. (Remaining water drops increase contamination risk). Make sure to not touch aspirating pipette to any of the coated surface, best to aspirate in corner of flask)
5. Store at 4 C in sterile bag covered with aluminum foil, make sure to date and fully seal bag
    a. Flasks last for up to a month
    b. Plates last for up to a week, parafilm sides

4.3.3 Thawing Cells

Notes: While you want to freeze cells as slowly as possible, thawing must be done as soon as possible

1. Sterilize a small flask (150 or 200 ml) from lab with alcohol and let air dry completely. Warm T-25 flask and 20 mL of media)
2. Remove cells from storage, cover completely with dry ice until ready or leave in small amount of liquid nitrogen.
3. Once flask is 100% dry, fill with sterile ddH20 and microwave for 30-60 seconds, then sterilize outside with ethanol
4. Spray cell vial with ethanol, wipe and immediately submerge pellet entirely into hot water using cryo-forceps without wetting the cap. Move in circular motion to move contents from vial walls then when it is almost completely defrosted and only a small ice crystal remains (< 1 min) remove it from water.
5. Quickly sterilize vial and hands, then open and carefully pipette contents using a p1000 into a 15 ml conical tube
6. **Slowly** add 10 ml of pre-warmed wash media **drop by drop** into the 15 ml conical tube and mix by pipetting gently 1-2 times
7. Spin at 190 x g for 6 minutes, add bFGF for media volume (4 ul for 10 ml)
8. Aspirate the supernatant, and resuspend the pellet in 1 ml of growth media using a p1000, pipette several times, and then add 2 ml of media with a 5 ml serological pipette
9. Transfer into remaining 7 mL of growth media and place in T-25 flask
10. Change the media the next day if there is an appreciable level of cell death

Note: After thawing, LUHMES will grow slow, look amorphic and take up to 3 passages to recover**. Media change needed 1-2 days after thawing** and when it starts to turn yellow. Often cells will be at 80% confluency in T-25 after a couple of days, at that point passage to T-75 flask.

4.3.4 Passaging Cells

Notes before passaging:

- Subculture before cells reach 80% confluency, ~ 5-6 million cells/mL in T-75 (since resuspended in ~2.5mL = 12.5-13 million total cells per flask)
    - If there are floaters/neurospheres lifting off, try double filtering but may need to thaw new cells. If cells are getting clumpy/growing on top of each other only slightly, try one passage with extra pipetting, and thaw new cells if persist.

- Warm aliquots of base/wash medium (25 ml for a T-75 flask) 1 hour before
- Warm freshly diluted 0.025% trypsin solution to 37°C 30 minutes to an hour before
    - Prepared by diluting 0.05% trypsin 1:1 in Ca++/Mg++ free Dulbecco's phosphate-buffered saline (D-PBS) to final concentration 0.025% Trypsin-0.1g/l EDTA
    - Regular strength trypsin or trypsin alternatives will kill a lot of the cells
    - Can use Accutase at room temperature, apply on cells for 5 minutes, pipette up and down with 20-200ul pipette to remove cells and add all of contents (including accutase and cells to new flask/well)

T75 optimal final volume 15 mL + 6 mL quench + 3 mL resusp + 1 mL spillage = 25 mL (Prewarm)

Material to place in hood: 18cm scraper, 40 micron filter, 15 and 50 mL tubes, 25mL serological pipette, 3 x 10mL pipette, 3 x 5mL, plastic aspirator tip

Protocol: (For T-75 flask but can adjust in relative surface area to any flask/well)

1. Aspirate culture medium.
2. Briefly rinse the cell layer with 10 mL d-PBS.
3. Add 4 mL of 0.025% Trypsin to flask and incubate for 2’45’’ at 37°C, take out wash media
4. Scrape surface to detach all cells windshield wiper motion down neck to base and back again towards neck of flask (~ 30 seconds)
5. Remove cap off of 15mL tube, using 10mL serological pipette add 6 mL of pre-warmed wash medium and mix gently a few times to break up and rinse off remaining cells, transfer cell suspension to the open 15 ml conical tube.
6. Spin at approximately 190 g’s for 6 minutes. While spinning, defrost bFGF, add 7.6 ul to 19 mL of remaining media) and leave media in hood. Return bFGF to -20. Prep a new microcentrifuge tube with 10 ul of trypan blue dye for cell counter
7. Aspirate supernatant careful to not get close to pellet. Using a 1 mL pipet, add 1.0 mL complete growth medium and pipet the pellet up and down to resuspend the cells (avoid creating foam or bubbles).
8. Place 40 µm cell strainer on a new 50 mL tube.
9. Add an additional 2.0 mL of complete growth medium using 5 mL serologic pipette and dissociate cells further by pipetting up and down.
10. Strain with a 40µm cell strainer, pipette last drops from bottom of strainer with p1000
11. Mix with P1000 and take 10 ul of cell suspension to place in tube with trypan blue

12. For maintenance in T-75 flask add volume corresponding to 1.6x10^6 of the cell suspension into 16 ml of remaining growth media and transfer to new pre-coated vented culture flask (1.6 divided by the number of millions of cells in 1 mL from countess). If in hurry just subcultivate at 1:10 = 250-300 ul)

Incubate cultures in 5% CO2/95% air at 37°C, passage every 3 days.

4.3.5 Differentiation

Notes:

- Only use low passage cells (<10-15 passages) that are healthy
- Start counting days at day 0, LUHMES are mature at day 7 and die shortly after
- Cell bodies will migrate but if they start clumping together to the point where they are on top of each other at any point, do not use, this is early apoptosis and will alter results

Differentiation Media:

- gDNF at 2500x (5 ng/ml) to get amount needed in ul divide vol of media in mL by 2.5 (For T-25 flask and 7.5 mL of media = 3 ul)
- cAMP (d0627-250mg sigma) at 100x (media in mL x 10). For 7.5 mL = 75 ul

For reconstituting cAMP use Tocris molarity calculator

https://www.tocris.com/resources/molarity-calculator

In calculator add 250mg to Mass, 100 mM to [] and M. Wt on bottle leaving volume empty, press calculate for amount of mL to reconstitute in.

Usually add around 5 mL to 250mg powder to make 100mM [] but need exact molecular weight from lot.

- Tetracycline at 2000x (to get amount needed in ul divide vol of media in mL by 2)

Vortex before use and keep away from light. For 7.5 mL = 3.75 ul

If using minimum media for T-25, 5 mL media, gDNF 2 ul, Tetra 2.5 ul, cAMP 50 ul (helps conserve cAMP)

Protocol:

1. Split cells and aim for 50% confluency to start or grow cells to that confluency and then change the media to differentiation media
    a. Seed 25x10^6 T225, 8.33x10^6 in T75,2.77x10^6 for T25)
    b. Do not use growth media (bFGF) to resuspend cells during split, and instead use wash media
2. Add calculated amount of cells (~5 x 10^6 in T-25) into 7.5 mL of fresh, pre-warmed differentiation media
3. Every other day remove 6 mL of old media and add 6 mL of new differentiation media (2.4 bDNF, 3 Tet, 60 cAMP).
    a. Be careful to remove and add media slowly and off to side of flask. Best to avoid using glass or plastic aspirating pipettes, they produce too much suction and will kill cells by detaching them especially close to full maturation.
    b. Neurons are extremely sensitive to temperature, pH, and light, so be quick and don't leave them exposed when aspirating or imaging.
4. Harvest cells on differentiation day 8 or one day earlier (first day is day 0) or neurons will peel off. Also be gentle when moving cells to microscope to avoid peeling.

4.3.6 Growth Media

Notes: Avoid adding any antibiotics although have had success using Penicillin Streptomycin 100X (for 500 mL of media add 5 mL of antibiotics).

- Base/Wash Media
    - 500 ml DMEM/F-12, GlutaMAX™ (Gibco-Invitrogen, Cat. No. 10565-018)
    - 1% (~5.5 ml, 1 bottle) N2 supplement (Gibco-Invitrogen, Cat. No. 17502-048)
- Growth Media
    - 40 ng/ml b-FGF (basic recombinant human Fibroblast Growth Factor; Gibco-Invitrogen Cat. No. 13256-029) added fresh at the last moment (no more than 1 week). Aliquots at 2500 X, so divide media in mL by 2.5 to get quantity in ul)
    - Per manufacturer instructions lyophilized 100 ug. Add 1mL of sterile DI H2O to make 100 ug/ml sol. Make 20 x 50 ul aliquot tubes. For 10 ug add 100ul, for 5 ug add 50 ul.

- Differentiation Media
    - 2 ng/ml human recombinant gDNF, aliquots at 2500 X, so divide media in mL by 2.5 (same as bFGF) to get quantity in ul. To make: reconstitute adding 2ml sterile H2O to 10ug stock (Thermo, Cat. No. PHC7045) to make 5 ug/mL (2500x)
    - 1 mM dibutyryl cAMP (Sigma, Cat. No. D0627)
        - Make sure it's dibutryl, regular cAMP won't do as it won't penetrate cell membrane
        - (see below) and sterile filter with low retention filter for 100 um/ml (100x stock). Add 10 ul per 1 ml media so multiply media in mL by 10 for quantity in ul
        - Calculator: https://www.tocris.com/resources/molarity-calculator
    - 1ug/ml tetracycline (Sigma T7660-5G or -25G)
        - Light sensitive and crashes out of solution, so vortex or pipette 10-20 times to completely resuspend each time. 2 mg/ml stock (2000x stock), so 1 ul for every 2 ml of media, calculate by dividing mL of media divided by 2. Each gram can be made with 500mL of autoclaved water covered with foil and filtered using .22-micron stericup in total darkness, cover stericup with foil too. This is still too much volume 50x50ml, alternately:
        - Make Superstock of .1 g/mL by mixing lyophilized 5G with 50mL H2O, then taking 1 mL of superstock and adding to 49 mL of H2O in 50mL Nalgene rapidflow .22-micron filter. Preferably store in "Lightsafe" tubes
- Freezing Media
    - 70% growth medium
    - 20% heat inactivated fetal bovine serum
    - 10% (v/v) DMSO

References

Allen, D. J., Makhov, A., Grilley, M., Taylor, J., Thresher, R., Modrich, P., & Griffith, J. D. (1997). MutS mediates heteroduplex loop formation by a translocation mechanism. *EMBO Journal*, *16*(14), 4467–4476. https://doi.org/10.1093/emboj/16.14.4467

Anders S. Hansen, Claudia Cattoglio, Xavier Darzacq & Robert Tjian (2018) Recent evidence that TADs and chromatin loops are dynamic structures, *Nucleus*, 9:1, 20-32.

Angelman, H. (1965). 'Puppet' Children A Report on Three Cases. *Developmental Medicine & Child Neurology*, *7*(6), 681–688. https://doi.org/10.1111/j.1469-8749.1965.tb07844.x

Akbari, V., Garant, J.-M., O'Neill, K., Pandoh, P., Moore, R., Marra, M. A., Hirst, M., & Jones, S. J. M. (2021). Megabase-scale methylation phasing using nanopore long reads and NanoMethPhase. *Genome Biology*, 22, 68. https://doi.org/10.1186/s13059-021-02306-1

Babraham Bioinformatics. Trim Galore!. Retrieved from https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/

Bailus, B. J., Pyles, B., McAlister, M. M., O'Geen, H., Lockwood, S. H., Adams, A. N., Nguyen, J. T., Yu, A., Berman, R. F., & Segal, D. J. (2016). Protein Delivery of an Artificial Transcription Factor Restores Widespread Ube3a Expression in an Angelman Syndrome Mouse Brain. *Molecular therapy : the journal of the American Society of Gene Therapy*, *24*(3), 548–555.

Bailus, B. J., & Segal, D. J. (2014). The prospect of molecular therapy for Angelman syndrome and other monogenic neurologic disorders. *BMC Neuroscience*, *15*(76), 1–7.

Bell, A. C., & Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature*, *405*(6785), 482–485. https://doi.org/10.1038/35013100

Buiting, K., Saitoh, S., Gross, S., Dittrich, B., Schwartz, S., Nicholls, R. D., & Horsthemke, B. (1995). Inherited microdeletions in the Angelman and Prader-Willi syndromes define an imprinting centre on human chromosome 15. Nature genetics, 9(4), 395–400. https://doi.org/10.1038/ng0495-395

Burnside, R. D., Pasion, R., Mikhail, F. M., Carroll, A. J., Robin, N. H., Youngs, E. L., Gadi, I. K., Keitges, E., Jaswaney, V. L., Papenhausen, P. R., Potluri, V. R., Risheg, H., Rush, B., Smith, J. L., Schwartz, S., Tepperberg, J. H., & Butler, M. G. (2011). Microdeletion/microduplication of proximal 15q11.2 between BP1 and BP2: a susceptibility region for neurological dysfunction including developmental and language delay. *Human genetics*, *130*(4), 517–528.

Calderon, L., Weiss, F. D., Beagan, J. A., Oliveira, M. S., Georgieva, R., Wang, Y. F., Carroll, T. S., Dharmalingam, G., Gong, W., Tossell, K., de Paola, V., Whilding, C., Ungless, M. A., Fisher, A. G., Phillips-Cremins, J. E., & Merkenschlager, M. (2022). Cohesin-dependence of neuronal gene expression relates to chromatin loop length. *eLife*, 11, e76539. https://doi.org/10.7554/eLife.76539

Chamberlain, S. J. (2012). Rnas of the human chromosome 15q11-q13 imprinted region. *WIREs RNA*, 4(2), 155-166. https://doi.org/10.1002/wrna.1150

Chamberlain, S.J. and Brannan, C.I. (2001) The Prader-Willi syndrome imprinting center activates the paternally expressed murine Ube3a antisense transcript but represses paternal Ube3a. Genomics, 73, 316–322.

Chamberlain S.J., Chen P.F., Ng K.Y., Bourgois-Rocha F., Lemtiri-Chlieh F., Levine E.S., Lalande M. (2010). Induced pluripotent stem cell models of the genomic imprinting disorders Angelman and Prader-Willi syndromes. *Proceedings of the National Academy of Sciences; 107* (41) 17668-17673

Chen, X., Zaro, J. L., & Shen, W. C. (2013). Fusion protein linkers: Property, design and functionality. *Advanced Drug Delivery Reviews*. https://doi.org/10.1016/j.addr.2012.09.039

Christian, S. L., Bhatt, N. K., Martin, S. A., Sutcliffe, J. S., Kubota, T., Huang, B., ... & Ledbetter, D. H. (1999). Large genomic duplicons map to sites of instability in the Prader-Willi/Angelman syndrome chromosome region (15q11-q13). Human Molecular Genetics, 8(6), 1025-1037. https://doi.org/10.1093/hmg/8.6.1025

Cook, E. H., Jr, Lindgren, V., Leventhal, B. L., Courchesne, R., Lincoln, A., Shulman, C., Lord, C., &

Courchesne, E. (1997). Autism or atypical autism in maternally but not paternally derived proximal 15q duplication. American journal of human genetics, 60(4), 928–934.
Coulson, R. L., Yasui, D. H., Dunaway, K. W., Laufer, B. I., Vogel Ciernia, A., Zhu, Y., Mordaunt, C. E., Totah, T. S., & LaSalle, J. M. (2018). Snord116-dependent diurnal rhythm of DNA methylation in mouse cortex. Nature communications, 9(1), 1616. https://doi.org/10.1038/s41467-018-03676-0
de Smith, A. J., Purmann, C., Walters, R. G., Ellis, R. J., Holder, S. E., Van Haelst, M. M., Brady, A. F., Fairbrother, U. L., Dattani, M., Keogh, J. M., Henning, E., Yeo, G. S. H., O'Rahilly, S., Froguel, P., Farooqi, I. S., Blakemore, A. I. F. (2009) A deletion of the HBII-85 class of small nucleolar RNAs (snoRNAs) is associated with hyperphagia, obesity and hypogonadism. *Human Molecular Genetics 18:* 3257-3265
de Wit, E., Vos, E. S. M., Holwerda, S. J. B., Valdes-Quezada, C., Verstegen, M. J. A. M., Teunissen, H., Splinter, E., Wijchers, P. J., Krijger, P. H. L., & de Laat, W. (2015). CTCF binding polarity determines chromatin looping. Molecular Cell, 60, 676-684.
Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, 10(2). https://doi.org/10.1093/gigascience/giab008
Davies, J. O. J., Telenius, J. M., McGowan, S. J., Roberts, N. A., Taylor, S., Higgs, D. R., & Hughes, J. R. (2015). Multiplexed analysis of chromosome conformation at vastly improved sensitivity. Nature Methods, 13(1), 74-80.
Dindot, S. V, Antalffy, B. A., Bhattacharjee, M. B., & Beaudet, A. L. (2018). The Angelman syndrome ubiquitin ligase localizes to the synapse and nucleus , and maternal deficiency results in abnormal dendritic spine morphology. *Human Molecular Genetics*, *17*(1), 111–118. https://doi.org/10.1093/hmg/ddm288
Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., ... & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1), 15-21. doi: 10.1093/bioinformatics/bts635
Dunaway, K. W., Islam, M. S., Coulson, R. L., Lopez, S. J., Vogel Ciernia, A., Chu, R. G., Yasui, D. H., Pessah, I. N., Lott, P., Mordaunt, C., Meguro-Horike, M., Horike, S. I., Korf, I., & LaSalle, J. M. (2016). Cumulative Impact of Polychlorinated Biphenyl and Large Chromosomal Duplications on DNA Methylation, Chromatin, and Expression of Autism Candidate Genes. *Cell reports*, *17*(11), 3035–3048. https://doi.org/10.1016/j.celrep.2016.11.058
Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. Bioinformatics, 32(19), 3047-3048. doi: 10.1093/bioinformatics/btw354
Faghihi, M. A., & Wahlestedt, C. (2009). Regulatory roles of natural antisense transcripts. *Nature reviews. Molecular cell biology*, *10*(9), 637–643. https://doi.org/10.1038/nrm2738
Falaleeva, M., Surface, J., Shen, M., de la Grange, P., & Stamm, S. (2015). SNORD116 and SNORD115 change expression of multiple genes and modify each other's activity. *Gene*, 572(2), 266–273.
Feinberg, A. P., Ohlsson, R., & Henikoff, S. (2006). The epigenetic progenitor origin of human cancer. *Nature Reviews Genetics*, 7, 21–33.
Fishilevich, S., Nudel, R., Rappaport, N., Hadar, R., Plaschkes, I., Iny Stein, T., Rosen, N., Kohn, A., Twik, M., Safran, M., Lancet, D., & Cohen, D. (2017). GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database: the journal of biological databases and curation*, *2017*, bax028. https://doi.org/10.1093/database/bax028
Fountain, M. D., Aten, E., Cho, M. T., Juusola, J., Walkiewicz, M. A., Ray, J. W., Xia, F., Yang, Y., Graham, B. H., Bacino, C. A., Potocki, L., van Haeringen, A., Ruivenkamp, C. A., Mancias, P., Northrup, H., Kukolich, M. K., Weiss, M. M., van Ravenswaaij-Arts, C. M., Mathijssen, I. B., Levesque, S., … Schaaf, C. P. (2017). The phenotypic spectrum of Schaaf-Yang syndrome: 18 new affected individuals from 14 families. *Genetics in medicine: official journal of the American College of Medical Genetics*, *19*(1), 45–52. https://doi.org/10.1038/gim.2016.53
Fuks F. (2005). DNA methylation and histone modifications: teaming up to silence genes. Current

opinion in genetics & development, 15(5), 490–495. https://doi.org/10.1016/j.gde.2005.08.002
Gabriel, J. M., Gray, T. A., Stubbs, L., Saitoh, S., Ohta, T., & Nicholls, R. D. (1998). Structure and function correlations at the imprinted mouse Snrpn locus. Mammalian genome : official journal of the International Mammalian Genome Society, 9(10), 788–793. https://doi.org/10.1007/s003359900868
Gamaarachchi, H., Lam, C. W., Jayatilaka, G., Samarakoon, H., Simpson, J. T., Smith, M. A., & Parameswaran, S. (2020). GPU accelerated adaptive banded event alignment for rapid comparative nanopore signal analysis. *BMC Bioinformatics, 21*(1), 343.
Geeven, G., Teunissen, H., de Laat, W., & de Wit, E. (2018). peakC: A flexible, non-parametric peak calling package for 4C and Capture-C data. *Nucleic Acids Research, 46*(15), e91. https://doi.org/10.1093/nar/gky443
Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D. U., Jung, I., Wu, H., Zhai, Y., Tang, Y., Lu, Y., Wu, Y., Jia, Z., Li, W., Zhang, M. Q., Ren, B., Krainer, A. R., Maniatis, T., & Wu, Q. (2015). CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell*, *162*(4), 900–910. https://doi.org/10.1016/j.cell.2015.07.038
Hanayama, R., Bloodgood, B. L., Kim, T.-K., Lipton, D. M., Maehr, R., Waldon, Z., … Chowdhury, S. (2010). The Angelman Syndrome Protein Ube3A Regulates Synapse Development by Ubiquitinating Arc. *Cell*, *140*(5), 704–716. https://doi.org/10.1016/j.cell.2010.01.026
Hansen, A. S., Pustova, I., Cattoglio, C., Tjian, R., & Darzacq, X. (2017). CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *ELife*, *6*. https://doi.org/10.7554/eLife.25776
Hao, N., Shearwin, K. E., & Dodd, I. B. (2017). Programmable DNA looping using engineered bivalent dCas9 complexes. *Nature Communications*, *8*(1). https://doi.org/10.1038/s41467-017-01873-x
Heck, D. H., Zhao, Y., Roy, S., LeDoux, M. S., & Reiter, L. T. (2008). Analysis of cerebellar function in Ube3a-deficient mice reveals novel genotype-specific behaviors. Human molecular genetics, 17(14), 2181–2189. https://doi.org/10.1093/hmg/ddn117
Holwerda, S. J. B., & Laat, W. de. (2013). CTCF: the protein, the binding partners, the binding sites and their chromatin loops. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1620). https://doi.org/10.1098/rstb.2012.0369
Horsthemke, B., & Wagstaff, J. (2008). Mechanisms of imprinting of the Prader-Willi/Angelman region. American Journal of Medical Genetics Part A, 146(16), 2041–2052.
Hsiao, J. S., Germain, N. D., Wilderman, A., Stoddard, C., Wojenski, L. A., Villafano, G. J., … Chamberlain, S. J. (2019). A bipartite boundary element restricts UBE3A imprinting to mature neurons. *Proceedings of the National Academy of Sciences*, *116*(6), 2181–2186. https://doi.org/10.1073/pnas.1815279116
Huang, H. S., Allen, J. A., Mabb, A. M., King, I. F., Miriyala, J., Taylor-Blake, B., … Philpot, B. D. (2012). Topoisomerase inhibitors unsilence the dormant allele of Ube3a in neurons. *Nature*, *481*(7380), 185–191. https://doi.org/10.1038/nature10726
Huibregtse, J. M., Scheffner, M., & Howley, P. M. (1993). *Cloning and expression of the cDNA for E6-AP, a protein that mediates the interaction of the human papillomavirus E6 oncoprotein with p53*. Molecular and cellular biology, 13(2), 775–784. https://doi.org/10.1128/mcb.13.2.775-784.1993
Huntriss, J. D., Latchman, D. S., & Williams, D. G. (1993). The snRNP core protein SmB and tissue-specific SmN protein are differentially distributed between snRNP particles. *Nucleic Acids Research*, *21*(17), 4047–4053. https://doi.org/10.1093/nar/21.17.4047
Jiang, Y.H.; Armstrong, D.; Albrecht, U.; Atkins, C.M.; Noebels, J.L.; Eichele, G.; Sweatt, J.D.; Beaudet, A.L. (1998) Mutation of the Angelman ubiquitin ligase in mice causes increased cytoplasmic p53 and deficits of con-textual learning and long-term potentiation. Neuron, 21, 799–811.
Jiang, Y.H., Tsai, T., Bressler, J., & Beaudet, A. L. (1998). Imprinting in Angelman and Prader-Willi syndromes. *Current Opinion in Genetics & Development*, *8*, 334–342.
Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., & Haussler, D. (2002). The human genome browser at UCSC. *Genome Research, 12*(6), 996-1006. doi: 10.1101/gr.229102

Kerpedjiev, P., Abdennur, N., Lekschas, F., McCallum, C., Dinkla, K., Strobelt, H., Luber, J. M., Ouellette, S. B., Azhir, A., Kumar, N., Hwang, J., Lee, S., Alver, B. H., Pfister, H., Mirny, L. A., Park, P. J., & Gehlenborg, N. (2018). HiGlass: web-based visual exploration and analysis of genome interaction maps. *Genome Biology,* 19, 125.

Kishino, T., Lalande, M., & Wagstaff, J. (1997). UBE3A/E6-AP mutations cause Angelman syndrome. Nature Genetics, 15(1), 70-73. https://doi.org/10.1038/ng0197-70

Kurukuti, S., Tiwari, V. K., Tavoosidana, G., Pugacheva, E., Murrell, A., Zhao, Z., … Ohlsson, R. (2006). CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. *Proceedings of the National Academy of Sciences*, *103*(28), 10684–10689.

Krijger, P. H. L., Geeven, G., Bianchi, V., Hilvering, C. R. E., & de Laat, W. (2020). 4C-seq from beginning to end: A detailed protocol for sample preparation and data analysis. (2020), *Methods, 170*, 7-32. https://doi.org/10.1016/j.ymeth.2019.07.014, https://www.sciencedirect.com/science/article/pii/S1046202318304742

Law, C. W., Chen, Y., Shi, W., & Smyth, G. K. (2014). Voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biology, 15(2), R29. doi: 10.1186/gb-2014-15-2-r29

Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan M, Carey V (2013). Software for Computing and Annotating Genomic Ranges. *PLoS Computational Biology, 9.* doi:10.1371/journal.pcbi.1003118.

Lee, A. K., & Potts, P. R. (2017). A Comprehensive Guide to the MAGE Family of Ubiquitin Ligases. Journal of molecular biology, 429(8), 1114–1142. https://doi.org/10.1016/j.jmb.2017.03.005

Lewis, J. D., Meehan, R. R., Henzel, W. J., Maurer-Fogy, I., Jeppesen, P., Klein, F., & Bird, A. (1992). Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell*, 69(6), 905–914. https://doi.org/10.1016/0092-8674(92)90610-o

Li, J., Chen, W., Li, D., Gu, S., Liu, X., Dong, Y., Jin, L., Zhang, C., & Li, S. (2021). Conservation of Imprinting and Methylation of MKRN3, MAGEL2 and NDN Genes in Cattle. *Animals: an open access journal from MDPI*, 11(7), 1985. https://doi.org/10.3390/ani11071985

Long Ranger Software. 10x Genomics. Retrieved from https://support.10xgenomics.com/single-cell-gene-expression/software/overview/welcome

Lotharius, J., Barg, S., Wiekop, P., Lundberg, C., Raymon, H. K., & Brundin, P. (2002). Effect of mutant alpha-synuclein on dopamine homeostasis in a new human mesencephalic cell line. The Journal of biological chemistry, 277(41), 38884–38894. https://doi.org/10.1074/jbc.M205518200

Leung K.N., Chamberlain S.J., Lalande M, LaSalle J.M. (2011). Neuronal chromatin dynamics of imprinting in development and disease. *J Cell Biochemistry, 112*(2), 365-73.

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18), 3094-3100. https://doi.org/10.1093/bioinformatics/bty191

Liu H., Kong X., Chen F. Mkrn3 functions as a novel ubiquitin E3 ligase to inhibit Nptx1 during puberty initiation. Oncotarget. 2017; 8: 85102-85109.

Mabb, A. M., Judson, M. C., Zylka, M. J., & Philpot, B. D. (2011). Angelman syndrome: insights into genomic imprinting and neurodevelopmental phenotypes. *Trends in neurosciences*, *34*(6), 293–303. https://doi.org/10.1016/j.tins.2011.04.001

MacLean, J. A., Bettegowda, A., Kim, B. J., Lou, C. H., Yang, S. M., Bhardwaj, A., ... & Wilkinson, M. F. (2011). The Rhox Homeobox Gene Cluster Is Imprinted and Selectively Targeted for Regulation by Histone H1 and DNA Methylation. Molecular and Cellular Biology, 31(3), 540-555

Martins-taylor, K., Hsiao, J. S., Chen, P., Glatt-deeley, H., De, A. J., Blakemore, A. I. F., … Chamberlain, S. J. (2014). Imprinted expression of UBE3A in non-neuronal cells from a Prader – Willi syndrome patient with an atypical deletion. *Human Molecular Genetics*, *23*(9), 2364–2373. https://doi.org/10.1093/hmg/ddt628

Matsuura, T., Sutcliffe, J., Fang, P., Galjaard, R.-J., Jiang, Y., Benton, C., … Beaudet, A. L. (1997). De novo truncating mutations in E6-AP ubiquitin-protein ligase gene (UBE3A) in Angelman syndrome.

Nature Genetics, *15*(January), 74–77.
Meguro, M., Kashiwagi, A., Mitsuya, K., Nakao, M., Kondo, I., Saitoh, S., & Oshimura, M. (2001). A novel maternally expressed gene, ATP10C, encodes a putative aminophospholipid translocase associated with Angelman syndrome. Nature Genetics, 28, 19–20.
Mishra, A., Godavarthi, S. K., & Jana, N. R. (2009). UBE3A/E6-AP regulates cell proliferation by promoting proteasomal degradation of p27. Neurobiology of disease, 36(1), 26–34. https://doi.org/10.1016/j.nbd.2009.06.010
Morgan, S. L., Mariano, N. C., Bermudez, A., Arruda, N. L., Wu, F., Luo, Y., … Wang, K. C. (2017). Manipulation of nuclear architecture through CRISPR-mediated chromosomal looping. *Nature Communications*, *8*. https://doi.org/10.1038/ncomms15993
Morita, S., Noguchi, H., Horii, T., Nakabayashi, K., Kimura, M., Okamura, K., … Hatada, I. (2016). Targeted DNA demethylation in vivo using dCas9-peptide repeat and scFv-TET1 catalytic domain fusions. *Nature Biotechnology*, *34*(10), 1060–1065. https://doi.org/10.1038/nbt.3658
Mortensen, F., Schneider, D., Barbic, T., Sladewska-Marquardt, A., Kühnle, S., Marx, A., & Scheffner, M. (2015). Role of ubiquitin and the HPV E6 oncoprotein in E6AP-mediated ubiquitination. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(32), 9872–9877. https://doi.org/10.1073/pnas.1505923112
Mulherkar SA, Jana NR. (2010). Loss of dopaminergic neurons and resulting behavioural deficits in mouse model of Angelman syndrome. Neurobiol Dis. 2010;40(3):586–592
Naumann, A., Hochstein, N., Weber, S., Fanning, E., & Doerfler, W. (2009). A distinct DNA-methylation boundary in the 5'- upstream sequence of the FMR1 promoter binds nuclear proteins and is lost in fragile X syndrome. American journal of human genetics, 85(5), 606–616. https://doi.org/10.1016/j.ajhg.2009.09.018
O'Geen, H., Bates, S. L., Carter, S. S., Nisson, K. A., Halmai, J., Fink, K. D., … Segal, D. J. (2019). Ezh2-dCas9 and KRAB-dCas9 enable engineering of epigenetic memory in a context-dependent manner. *Epigenetics and Chromatin*, *12*(1), 1–20. https://doi.org/10.1186/s13072-019-0275-8
Pflueger, C., Tan, D., Swain, T., Nguyen, T., Pflueger, J., Nefzger, C., … Lister, R. (2018). A modular dCas9-SunTag DNMT3A epigenome editing system overcomes pervasive off-target activity of direct fusion dCas9-DNMT3A constructs. *Genome Research*, *28*, 1193–1206. https://doi.org/10.1101/gr.233049.117.6
Pierce, S. E., Tyson, T., Booms, A., Prahl, J., & Coetzee, G. A. (2018). Parkinson's disease genetic risk in a midbrain neuronal cell line. *Neurobiology of Disease*, *114*(October 2017), 53–64. https://doi.org/10.1016/j.nbd.2018.02.007
Powell, W. T., Coulson, R. L., Gonzales, M. L., Crary, F. K., Wong, S. S., Adams, S., Ach, R. A., Tsang, P., Yamada, N. A., Yasui, D. H., Chédin, F., & LaSalle, J. M. (2013). R-loop formation at Snord116 mediates topotecan inhibition of Ube3a-antisense and allele-specific chromatin decondensation. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(34), 13938–13943. https://doi.org/10.1073/pnas.1305426110
Prickett, A. R., Barkas, N., McCole, R. B., Hughes, S., Amante, S. M., Schulz, R., & Oakey, R. J. (2013). Genome-wide and parental allele-specific analysis of CTCF and cohesin DNA binding in mouse brain reveals a tissue-specific binding pattern and an association with imprinted differentially methylated regions. *Genome research*, 23(10), 1624–1635. https://doi.org/10.1101/gr.150136.112
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
Rauluseviciute, I., Drabløs, F. & Rye, M.B. (2020) DNA hypermethylation associated with upregulated gene expression in prostate cancer demonstrates the diversity of epigenetic regulation. *BMC Med Genomics* 13, 6. https://doi.org/10.1186/s12920-020-0657-6
Reiter, L. T., Seagroves, T. N., Bowers, M., & Bier, E. (2006). Expression of the Rho-GEF Pbl/ECT2 is regulated by the UBE3A E3 ubiquitin ligase. *Human Molecular Genetics*, 15(18), 2825–2835.
Renda, M., Baglivo, I., Burgess-Beusse, B., Esposito, S., Fattorusso, R., Felsenfeld, G., & Pedone, P. V. (2007). Critical DNA binding interactions of the insulator protein CTCF: A small number of zinc

fingers mediate strong binding, and a single finger-DNA interaction controls binding at imprinted loci. *Journal of Biological Chemistry*, *282*(46), 33336–33345. https://doi.org/10.1074/jbc.M706213200

Rougeulle, C., Fontes, M., Colleaux, L., & Lalande, M. (1998). An imprinted antisense RNA overlaps UBE3A and a second maternally expressed transcript. *Nature*, *19*(May), 15–16.

Ruiz-Velasco, M., Kumar, M., Lai, M. C., Bhat, P., Solis-Pinson, A. B., Reyes, A., Kleinsorg, S., Noh, K. M., Gibson, T. J., & Zaugg, J. B. (2017). CTCF-Mediated Chromatin Loops between Promoter and Gene Body Regulate Alternative Splicing across Individuals. *Cell Systems*, 5, 628–637. https://doi.org/10.1016/j.cels.2017.10.018

Runte, M., Hüttenhofer, A., Groß, S., Kiefmann, M., Horsthemke, B., & Buiting, K. (2001). The IC-SNURF – SNRPN transcript serves as a host for multiple small nucleolar RNA species and as an antisense RNA for UBE3A. *Human Molecular Genetics*, *10*(23), 2687–2700.

Runte, M., Hüttenhofer, A., Groß, S., Kiefmann, M., Horsthemke, B., & Buiting, K. (2000). Identification of novel imprinted transcripts in the Prader-Willi syndrome and Angelman syndrome deletion region: further evidence for regional imprinting control. Human Molecular Genetics, 9(6), 837-847. https://doi.org/10.1093/hmg/9.6.837

Sahoo T., del Gaudio D., German J.R., Shinawi M., Peters S.U., Person R.E., Garnica A., Cheung S.W., Beaudet A.L. (2008). Prader-Willi phenotype caused by paternal deficiency for the HBII-85 C/D box small nucleolar RNA cluster. *Nature Genetics 40*(6):719-21

Sadikovic, B., Fernandes, P., Zhang, V. W., Ward, P. A., Miloslavskaya, I., Rhead, W., … Fang, P. (2014). Mutation update for UBE3A variants in angelman syndrome. *Human Mutation*, *35*(12), 1407–1417. https://doi.org/10.1002/humu.22687

Schaaf, C. P., Gonzalez-Garay, M. L., Xia, F., Potocki, L., Gripp, K. W., Zhang, B., Peters, B. A., McElwain, M. A., Drmanac, R., Beaudet, A. L., Caskey, C. T., & Yang, Y. (2013). Truncating mutations of MAGEL2 cause Prader-Willi phenotypes and autism. *Nature Genetics*, 45, 1405–1408. https://doi.org/10.1038/ng.2776

Schildknecht, S., Karreman, C., Pöltl, D., Efrémova, L., Kullmann, C., Gutbier, S., … Leist, M. (2013). Generation of genetically-modified human differentiated cells for toxicological tests and the study of neurodegenerative diseases. *Altex*, *30*(4), 427–444. https://doi.org/10.14573/altex.2013.4.427

Scholz, D., Pöltl, D., Genewsky, A., Weng, M., Waldmann, T., Schildknecht, S. and Leist, M. (2011), Rapid, complete and large-scale generation of post-mitotic neurons from the human LUHMES cell line. *Journal of Neurochemistry*, 119: 957-971. https://doi.org/10.1111/j.1471-4159.2011.07255.x

Sirois, C. L., Bloom, J. E., Fink, J. J., Gorka, D., Keller, S., Germain, N. D., Levine, E. S., & Chamberlain, S. J. (2020). Abundance and localization of human UBE3A protein isoforms. *Human molecular genetics*, *29*(18), 3021–3031. https://doi.org/10.1093/hmg/ddaa191

Smyth, G. K. (2004). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Statistical Applications in Genetics and Molecular Biology, 3(1), Article 3. doi: 10.2202/1544-6115.1027

Soejima, H., & Wagstaff, J. (2005). Imprinting Centers, Chromatin Structure, and Disease. Journal of Cellular Biochemistry, 95, 226–233.

Sourya Bhattacharyya, Vivek Chandra, Pandurangan Vijayanand, and Ferhat Ay (2019), Identification of significant chromatin contacts from HiChIP data by FitHiChIP, *Nature Communications*, Vol 10, No 4221, https://doi.org/10.1038/s41467-019-11950-y

Splinter, E., Heath, H., Kooren, J., Palstra, R.-J., Klous, P., Grosveld, F., Galjart, N., and de Laat, W. (2006). CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev.* 20, 2349–2354.

Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., Stein, T. I., Nudel, R., Lieder, I., Mazor, Y., Kaplan, S., Dahary, D., Warshawsky, D., Guan-Golan, Y., Kohn, A., Rappaport, N., Safran, M., & Lancet, D. (2016). The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Current protocols in bioinformatics*, *54*, 1.30.1–1.30.33.

Syding, L. A., Kubik-Zahorodna, A., Nickl, P., Novosadova, V., Kopkanova, J., Kasparek, P., Prochazka,

J., & Sedlacek, R. (2022). Generation and Characterization of a Novel Angelman Syndrome Mouse Model with a Full Deletion of the Ube3a Gene. Cells, 11(18), 2815. https://doi.org/10.3390/cells11182815

Taniura, H., Taniguchi, N., Hara, M., & Yoshikawa, K. (1998). Necdin, a postmitotic neuron-specific growth suppressor, interacts with viral transforming proteins and cellular transcription factor E2F1. The Journal of Biological Chemistry, 273, 720–728.

Tarjan, D. R., Flavahan, W. A., & Bernstein, B. E. (2019). Epigenome editing strategies for the functional annotation of CTCF insulators. *Nature Communications*, *10*(4258), 1–8. https://doi.org/10.1038/s41467-019-12166-w

UC Davis bioinformatics core RNA-Seq Analysis. (2021). Retrieved from https://ucdavis-bioinformatics-training.github.io/2022-June-RNA-Seq-Analysis/

Wang, J., Lou, S. S., Wang, T., Wu, R. J., Li, G., Zhao, M., Lu, B., Li, Y. Y., Zhang, J., Cheng, X., Shen, Y., Wang, X., Zhu, Z. C., Li, M. J., Takumi, T., Yang, H., Yu, X., Liao, L., & Xiong, Z. Q. (2019). UBE3A-mediated PTPA ubiquitination and degradation regulate PP2A activity and dendritic spine morphology. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(25), 12500–12505. https://doi.org/10.1073/pnas.1820131116

Washington University epigenome browser (2022). Retrieved from https://epigenomegateway.wustl.edu/browser/

Wiehle, L., Thorn, G. J., Raddatz, G., Clarkson, C. T., Rippe, K., Lyko, F., Breiling, A., & Teif, V. B. (2019). DNA (de)methylation in embryonic stem cells controls CTCF-dependent chromatin boundaries. *Genome research*, 29(5), 750–761. https://doi.org/10.1101/gr.239707.118

Wijesuriya, T. M., De Ceuninck, L., Masschaele, D., Sanderson, M. R., Carias, K. V., Tavernier, J., & Wevrick, R. (2017). The Prader-Willi syndrome proteins MAGEL2 and necdin regulate leptin receptor cell surface abundance through ubiquitination pathways. Human Molecular Genetics, 26(21), 4215–4230. https://doi.org/10.1093/hmg/ddx311

Williamson, C. M., Turner, M. D., Ball, S. T., Nottingham, W. T., Glenister, P., Fray, M., ... & Peters, J. (2011). Uncoupling Antisense-Mediated Silencing and DNA Methylation in the Imprinted Gnas Cluster. PLoS Genetics, 7(3), e1001347

Wu, H., & Feng, H. (2020). DSS: An R package for differential splicing analysis of RNA-Seq data. Retrieved from https://github.com/haowulab/DSS/tree/master

Yashiro, K., Riday, T., Condon, K., Roberts, A., Bernardo, D., Prakash, R., … Philpot, B. (2010). Ube3a is required for experience-dependent maturation of the neocortex. *Nat Neurosci*, *23*(1), 1–7. https://doi.org/10.1038/jid.2014.371

Yusufzai, T.M., Tagami, H., Nakatani, Y., and Felsenfeld, G. (2004). CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. Mol. Cell 13, 291–298.

Xicoy, H., Wieringa, B., & Martens, G. J. (2017). The SH-SY5Y cell line in Parkinson's disease research: a systematic review. Molecular neurodegeneration, 12(1), 10. https://doi.org/10.1186/s13024-017-0149-0

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., & Liu, X. S. (2008). Model-based Analysis of ChIP-Seq (MACS). *Genome Biology, 9*(9), R137.

Zhao, H., Sun, Z., Wang, J., Huang, H., Kocher, J.-P., & Wang, L. (2013). CrossMap: a versatile tool for coordinate conversion between genome assemblies. *Bioinformatics* (Oxford, England), btt730

Zheng, Z., Li, S., Su, J., Leung, A. W.-S., Lam, T.-W., & Luo, R. (2022). Symphonizing pileup and full-alignment for deep learning-based long-read variant calling. Nature Computational Science, 2, 797-803, https://doi.org/10.1038/s43588-022-00387-x