Table of Contents

Abstract……. 2

Introduction…….. 3-9

Methods…… 10-12

Chapter 1: ASH1L Gene Expression Patterns During Human Brain Development…… 13-22

Chapter 2: ASH1L Co-Expression Patterns in the Prefrontal Cortex During Prenatal Development… 23-28

Chapter 3: ASH1L Co-Expression and Protein-Interaction Network During 8-16pcw…… 29-37

Conclusion ……. 38-39

Bibliography…….. 40-42

Appendices …… 42-51

Acknowledgments …… 52

ABSTRACT

Autism spectrum disorders (ASD) are associated with defects in neuronal connectivity and are highly heritable. A significant proportion of ASD cases are of complex genetic etiology; complexity which might reflect the impact of gene-environment interactions. However, there is a gap in our understanding of the mechanisms that underlie the gene-environment interaction in autism complex etiology. Genome wide association studies in large ASD cohorts identified high risk variants associated with autism in genes that regulate histone modifications and remodel chromatin. These findings highlight the relevance of chromatin regulatory mechanisms in the pathology of ASD. Changes in Histone H3 methylation have been identified in a subset of neuronal genes in postmortem cerebral cortex of autism patients. ASH1L is a histone H3-methyltransferase that was previously identified in whole exome sequencing studies, as a gene strongly enriched for variants likely to increase ASD risk. However, the role of ASH1L during human neurodevelopment is not well understood on a cellular or molecular basis. To investigate ASH1L during human brain development, I analyzed developmental transcriptome data collected from donated post-mortem human brain tissue; tissues that was processed in the Allen Brain Atlas. My analysis suggests that ASH1L is active during early prenatal development, along with other genes important to chromatin modification. Furthermore, I find that co-expression network analysis implicates ASH1L in a cluster of genes important to the development of neuronal projections, protein ubiquitination, and neurotrophin signaling pathways. This analysis supports cellular and molecular phenotypes seen in ASH1L knockdown studies performed in human-induced neurons. Through these analyses, ASH1L is shown to be important both in early neurodevelopment, and to be strongly associated with ASD pathology.

INTRODUCTION

ASD is a common neurodevelopmental disorder and growing public health concern. On average the cost of taking care of an individual with ASD is about $1.4 million dollars per individual during their lifetime (Baio, 2018). A recent study by the CDC reports that ASD affects 1 in 59 children aged 8 years in 2014. The latest findings provide evidence that the prevalence of ASD is higher than previously reported estimates. An increased prevalence means there is an increasing number of families in the United States in need of services, and in search for viable bio-therapeutic options to improve the lives of their children with ASD (Baio, 2018). To develop therapeutic options, the mechanisms underlying autism-associated phenotypes must be better understood through research.

ASD are defined diagnostically by impaired social interactions, language deficits, and restricted or repetitive patterns of thought and behavior, present early in a child's development (Reynolds and Kamphaus, 2013). These phenotypes are thought to result from disordered neurodevelopment, although the precise etiology is not known. The challenge of understanding ASD's etiology stems, to an extent, from the pathology's strong genetic heterogeneity. Hundreds of genes have been linked to ASD, but only a few reproducible genetic variants have recurrent and statistically significant evidence linking them to ASD (Loke et al., 2015).

The Simons Foundation Autism Research Initiative (SFARI) gene scoring panel annually reviews published literature on the genetics of autism to assess the strength of the evidence linking

candidate genes to ASD. Genes in category 1, the high confidence category, are genes with the strongest evidence to be linked with ASD. To be placed in the category independent studies must have rigorous statistical comparison between cases and controls yielding genome-wide statistical significance for a specific genetic mutation and ASD (Abrahams et al., 2013). SFARI ASD gene database has identified 25 genes with de-novo rare mutations to be category 1 genes, genes with the highest association with ASD. Eight of these category 1 genes are chromatin regulatory genes, genes important to the regulation epigenetic landscape of neurodevelopment. Epigenetic programs are important to embryonic and subsequently neuronal development as they orchestrate series of events that require spatiotemporal control (Habibi and Stunnenberg, 2017). Specifically, epigenetic proteins are important in mediating the response to environmental cues that initiate processes such as neurogenesis, neuronal migration, gliogenesis, and synaptic formation.

One chromatin-modifying and category 1 gene, ASH1L, has been associated with Autism. Both frameshift and missense de-novo loss-of-function (LoF) ASH1L variants have been reported in individuals with ASD (De Rubeis et al., 2014; Willsey et al., 2013, Tammimies et.al 2015). In clinical studies of small cohorts of children LoF mutations in ASH1L, phenotypes such as intellectual disability, speech difficulties, mild thickening of the corpus callosum, microcephaly, as well as facial and skeletal abnormalities have been reported (Okamoto et al., 2017; Shen et al., 2019; Stessman et al., 2017). However, the impact of these ASH1L variants on protein function and cellular processes have yet to be determined.

ASH1L Molecular Function:

ASH1L is a histone methyl-transferase known to di-methylate H3K36 and H3K4 in humans. The protein is a member of the Trithorax group of proteins (TrxG), a group of well known chromatin modifying factors associated with gene activation during development (Miyazaki et al., 2013). ASH1L, as a TrxG member, plays a vital role in maintaining Hox gene expression and correct skeletal formation during embryonic development. Best understood in *Drosophila,* the methylation of H3K36 by ASH1L promotes the binding of the H3K4 methyltransferase complex, COMPASS, to the H3K4 mark, and subsequent methylation of the mark by MLL3 (Schuettengruber et al., 2017). The methylation of H3K36 and H3K4 histone marks modify chromatin structure allowing for the activation of transcription of target genes. Counteracting these TrxG methyltransferases, are demethylases such as KDM2B. KDM2B specifically demethylates H3K36me2 marks deposited by ASH1L, while also recruiting the non-canonical Polycomb repressive complex 1 (PRC1) which leads to gene repression or inactivation (Corley and Kroll, 2015). The PRC1 complex is a member of the Polycomb Group (PcG), which acts antagonistically to TrxG protein complexes, to orchestrate the expression and repression of key genes during embryonic development. In addition to the PRC1 complex, the PcG also contains the Polycomb repressive complex 2 (PRC2) complex, made of the catalytic subunits EZH2, EED, and SUZ1/2 (Schuettengruber et al., 2017). The methylation of H3K36 by ASH1L has also been reported to inhibit the catalytic activity of PRC2 (Huang and Zhu, 2018). Together TrxG and PcG proteins work in opposition to coordinate developmental processes across species through epigenetic modification.

Given the strong implication of ASH1L in neurodevelopmental disease pathogenesis, it is important to understand the mechanisms by which ASH1L operates specifically during brain development. Despite all that is known about PcG protein complexes, less is understood about the spatiotemporal function of ASH1L during neurodevelopment.

The Role of PRC2 in Neuronal Development

Prenatally, the PRC2 complex, is known to be highly expressed in neural progenitor cells in the ventricular zone, the proliferative region in the developing brain helping to maintain pluripotency and prevent differentiation during the brain's neurogenic phase (Fig.1). The PRC2 complex prevents neuronal differentiation through the tri-methylation of H3K27 mark near neural fate genes like *NGN1* and *GFAP*. Deletion of *EZH2*, the PRC2 catalytic subunit, during embryonic development leads to precocious neuronal differentiation and astrogliogenesis (Corley and Kroll, 2015). PRC2 proteins also promote proliferation during cerebellum formation, gliogenesis, and adult neurogenesis. Expression of *EZH2* is high in oligodendrocyte and cerebellar progenitor cells (Liu et al., 2018). In postnatal brains, the expression of *EZH2* promotes proliferation through the inhibition of PTEN in the hippocampus, and the expression of another PRC2 catalytic subunit, *EED*, maintains neuroblast pluripotency in the olfactory bulb (Liu et al., 2018; Sun et al., 2018). Not only is PRC2 important to proliferation, it is important in preventing neuronal cell death. As a person ages, PRC2 activity leads to increased levels of H3K27me3 marks on genes important to cell death and transcriptional regulators (*PMAIP, BOL, CDKN2A, GATA6*) in human neurons (von Schimmelmann et al., 2016). PRC2's role in the repression of apoptotic genes and activation of proliferation genes in the brain, has also been implicated in glioblastoma and myeloid malignancies (Ronan et al., 2013)

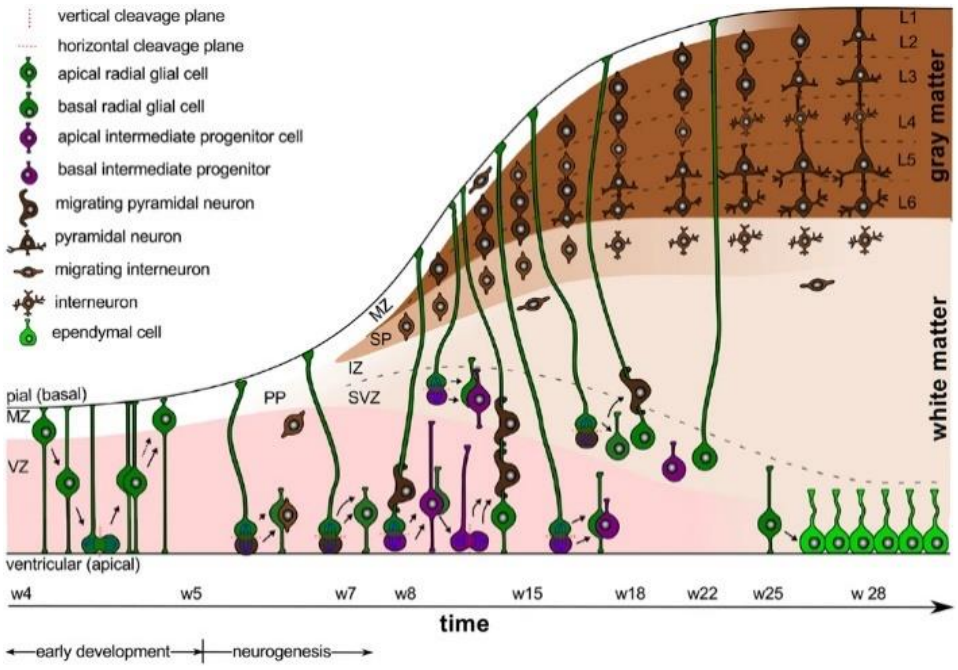


Figure 1. Human Cortical Development. Budday et.al. Front.Cell Neurosci. 2015; 9:257

Finally, recent studies have found PRC2 to play an important role in establishing activity dependent neuronal morphology. The conditional knockout of a H3K27 demethylase gene, *KDM6A*, in mice showed that the knockout led to abnormal neuronal morphology, cognitive deficits, and decreased expression of genes critical for synaptic plasticity (*ACTN2, EGR3, TGFB2, NOS1).* Abnormal neuron morphology is an attribute often associated with ASD, a disease PcG proteins have been associated with. Deletion mutations in the H3K27me2 demethylase, JARID2, has been highly associated with ASD (Liu et al., 2018). Furthermore, a study of chromatin modifier, CHD8, in neural progenitor cells found that CHD8 binds to the promoters of EZH2 and Suz12, promoting PRC2 repression of neural genes. When functional, CHD8 ensure the generation of appropriate number of neurons during cortical development. When mutated, CHD8 upregulates PRC2 target genes and leads to the outgrowth of the cortex (Durak et al., 2016). Interestingly, studies have shown CHD8 to also be a highly associated-ASD risk gene with gene targets strongly enriched for other ASD-risk genes, including *ASH1L* (Cotney et al., 2015)

ASH1L in Neuronal Development

Unlike PcG proteins, ASH1L's function and interaction in epigenetic networks during development is not a well understood. In mice, *ASH1L* knock-out was shown to be embryonically lethal, while heterozygous mutants had severe skeletal abnormalities. Outside of the nervous system, ASH1L targets genes important for muscle cell differentiation, many of which are also Polycomb target genes, in human myoblasts (Castiglioni et al., 2018). Within the nervous system, ASH1L has been shown to regulate the transcription of *NRXN1α* in response to neuronal activity (Zhu et al., 2016b). The effect of ASH1L prior to synaptogenesis is under investigation in the laboratory of Dr. Sofia Lizarraga. *ASH1L* shRNA knockdown (KD) experiments performed in human embryonic stem cell-derived neurons, show a decrease in overall neurite length and neural arborization. These phenotypes are correlated with the downregulation of the gene encoding for the neurotrophin receptor TrkB (*NTRK2*) in *ASH1L* KD neurons. However, the mechanism behind this observed phenotype has yet to be discovered.

The primary goal of my thesis work is to gain a better understanding on the molecular pathways that might be associated with ASH1L and that might be relevant to its neuronal function. Many fundamental questions surrounding ASH1L's role in brain development currently exist. For example, where and when is *ASH1L* expressed in the developing human brain? Does ASH1L regulate neuronal function by counteracting the activity of PcG complexes? Does ASH1L regulate pathways associate with neuronal morphogenesis or cell fate? My thesis aims to investigate these questions by: (i) investigating where in the brain and at what levels is *ASH1L* expressed during human development; (ii) comparing the expression of *ASH1L* in the human prefrontal cortex to the

expression of other chromatin modifiers in the TrxG and PcG complexes; (iii) exploring the potential co-expression networks within the dorsolateral prefrontal cortex during early prenatal brain development. In this manner, my thesis work, will help to better understand the neuronal morphogenesis phenotypes seen in the *ASH1L* KD experiments, as well as elucidate associations between ASH1L-related co-expression networks and ASD-risk co-expression networks.

MATERIALS AND METHODS:

Developmental Transcriptome Data Analysis

Gene expression data was obtained from the Allen Brain Atlas Developmental Transcriptome Database (Miller et al., 2014). Transcriptome data was obtained from RNA-sequencing and exon microarray hybridization that was generated across 13 developmental stages and 8-16 brain structures. RNA-sequencing was performed in human brain specimens collected by the Department of Neurobiology at Yale School of Medicine and the National Institute of Mental Health. Postmortem brains were collected from individuals ranging from 8 post-conceptional weeks (pcw) to over 40 years of age and were selected according to criteria that would ensure consistency between samples; i.e. brains with chromosomal abnormalities, malformations, neuronal loss, or associations with alcohol abuse and neurological disease were excluded from analysis. Eligible tissues were screened for high RNA quality and cytoarchitectural integrity. Tissue and RNA sequencing preparation, and processing were performed as described by the Allen Brain Reference Atlas. In total, expression data was collected from 37 different donors; 21 males and 16 females, across the human lifespan.

Expression heat maps were generated using RNA-sequencing expression values downloaded from the Developmental Transcriptome database for genes of interest. Expression values, age of donor, and brain structure were further analyzed using Broad Institute's Morpheus software (https://software.broadinstitute.org/morpheus). The colors on the heat map range from dark blue, representing low expression, to bright red representing high expression, as ranked based on expression values contained in each row. Each column represents brain tissue from one donor;

apart from a few columns, where there were multiple donors for one age. In these cases, the expression was averaged from donors of the same age (12pcw, 13pcw, 16pcw,4 months, 3 years) in order to improve the visualization of age-related expression trends. Donor information can be found in Appendix A. Expression data was not available for every brain structure for each donor, as denoted by the gray boxes in the heat map. Average expression, hierarchical clustering, and box and whisker plots were calculated and created using the Morpheus software.

Gene Correlation Study

To elucidate genes with similar expression patterns to *ASH1L* and other genes of interest in the study, the correlative search option within the Developmental Transcriptome Database was used. This function finds genes with similar expression patterns over the entire dataset or over a subset of structures and developmental stages (Allen Brain Atlas, 2017). For this study correlative genes were found over a time frame of 8-37pcw, the time frame which corresponds to prenatal development. Furthermore, the correlative search focused on structures of the prefrontal cortex (PFC): dorsolateral prefrontal cortex (DFC), ventrolateral prefrontal cortex (VFC), anterior (rostral) medial prefrontal cortex (MFC). The top 500 genes with the highest positive correlation (r) to *ASH1L* gene expression in the DFC, VFC, and MFC during 8-37pcw were analyzed ($0.741 < r < 0.959$).

Functional Enrichment Analysis of Correlation Data

To analyze the function of top correlated genes Toppgene Suite ToppFun software (https://toppgene.cchmc.org/enrichment.jsp) was used to detect functional enrichment of these

top genes based on Transcriptome, Proteome, Ontologies (GO, Pathway), Phenotype (human disease and mouse phenotype), and literature co-citation (Chen et al., 2009). Genes were imported into the portal, correction was set and resulting functional enrichments were given in tabular form and ranked by P-value.

Co-expression Network Construction and Data Visualization

ASH1L co-expression networks were created using the correlative search method, and modelled after the ASD co-expression network construction method performed by Willsey et.al (Willsey et al., 2013). First, the top 20 positively correlated and top 15 anti-correlated genes with *ASH1L* in the DFC during 8-16pcw timeframe were found. The expression data in this timeframe came from ten different donors. Next a correlative search for each of the 35 genes, identified above, was performed. The correlated genes with R-values above 0.97 or below -0.97 were recorded for each gene. In total 1270 gene pairs and corresponding R-values were used to construct the *ASH1L* co-expression networks.

To visualize the co-expression data, the data was uploaded to STRINGdb (https://string-db.org/) . Protein interaction networks were further elucidated by modifying the minimum interaction score to high confidence, this modifies the network to only show gene interactions that exhibited a score of 0.7 and above. STRINGdb predicts the strength of gene interactions based on a combined score computed from probabilities from different evidence channels and correcting for probability of randomly observing the interaction (Szklarczyk et al., 2015). The correlative and anti-correlative co-expression networks for ASH1L were computed separately.

CHAPTER 1

ASH1L GENE EXPRESSION PATTERNS DURING HUMAN BRAIN DEVELOPMENT

ASH1L is Highly Expressed in Early-Mid Fetal and Late Adolescent Brains

Developmental Transcriptome data collected from human brain tissue specimens in different areas of the brain and at different stages of development is useful in helping to understand where and when a gene may be important to human neurodevelopment. Non-intronic ASH1L expression data from the Developing Human Brain Atlas is visualized in a heatmap (Fig. 1A). As indicated by the red coloration in the heat map, high expression values are seen predominantly during mid fetal development and adolescence for most structures. There is a notable region of low expression for most brain structures from 2-4 years of age. Exceptions to this trend are seen in the cerebellar cortex (CBC) and the mediodorsal nucleus of the thalamus (MD), where higher expression is seen later from 4-15 years and 18-40 years respectively. Interestingly, in Willsey's study of ASD co-expression networks, a MD-CBC network was found to be highly enriched for ASD-associated genes from the period of birth to six years of age (Willsey et al., 2013). The MD-CBC co-expression network within this time period and region included ASH1L, among other ASD-associated genes, such as NRXN1 which also has been shown to be regulated by ASH1L (Zhu et al., 2016b). Therefore, analysis of *ASH1L* expression correlates with previously published data sets

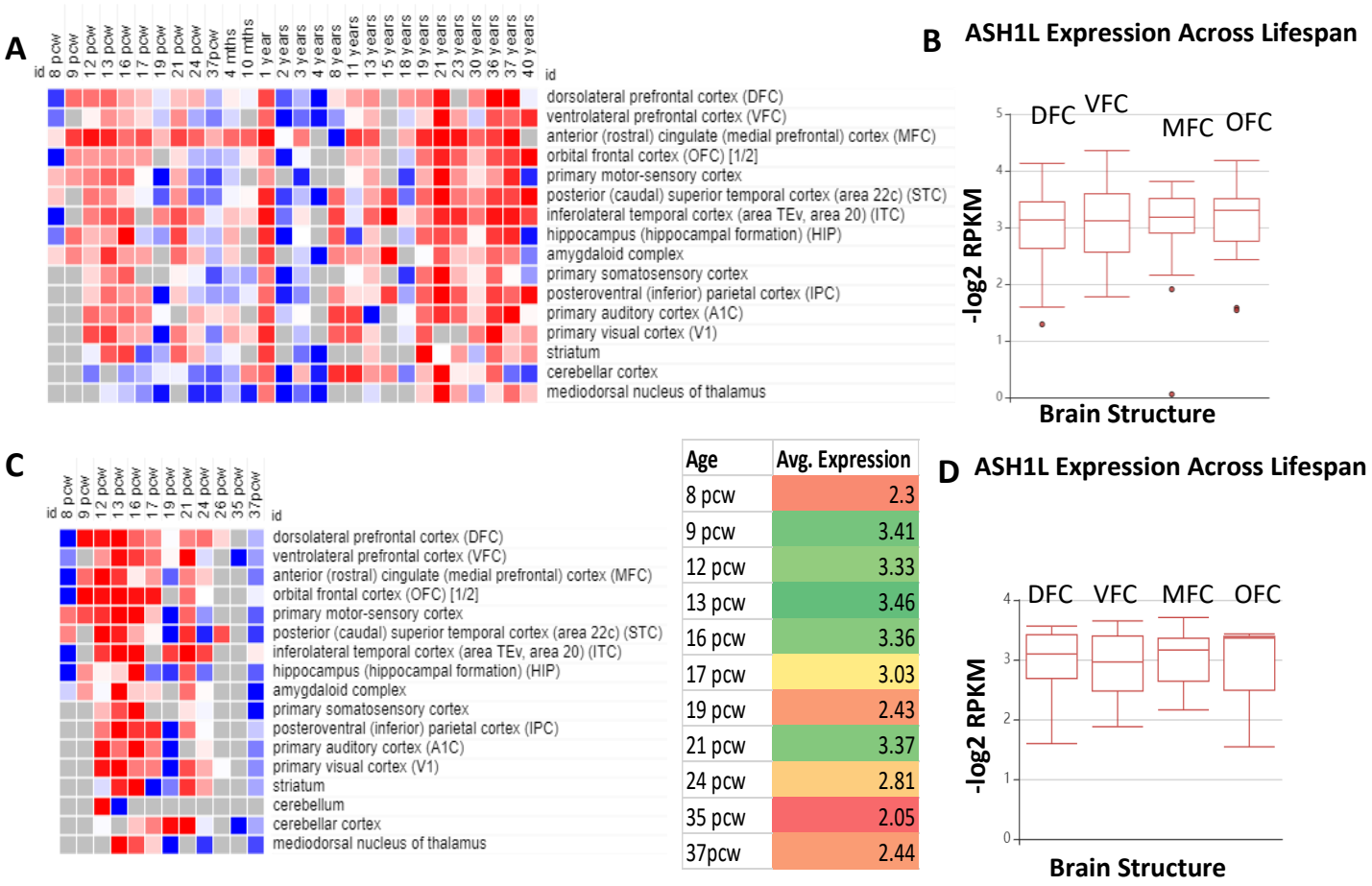
Analysis of the heat map showed that the highest average expression of *ASH1L* during the human lifespan was in the CBC. The lowest average expression of *ASH1L* in the MD (Fig.2A). The MD-CBC structures and association with ASH1L is interesting. However, I decided to focus on the structures of the prefrontal cortex (PFC) because there is evidence suggesting a role of the PFC in autism pathogenesis, and because our in-vitro neuronal data on ASH1L was obtained in human cortical neurons (Donovan and Basson, 2017). The average expression of *ASH1L* in each of the structures of

the prefrontal cortex (DFC, VFC, MFC, OFC) was roughly the same, 3.05 ± 0.05 (log2) reads per kilobase of transcript (RPKM), (Fig.2B and Fig. 2D). Temporally, *ASH1L* was most highly expressed during prenatal development in the brains of donors with ages of 9pcw-16pcw, a timeframe that was confirmed to be significant after performing hierarchical clustering of the average expression values across the columns of the heat map (Fig.2C).

The Brain Atlas is useful in understanding gene expression on a larger scale, as it focuses on total expression in one brain structure. The atlas, however, does not provide information on what type of cells within the brain structure are expressing *ASH1L*, and what proportion of the ASH1L positive cells make up the entire structure. A cellular-based study would be of interest, especially during the 5pcw-15pcw when the cerebral cortex is undergoing neurogenesis, radial migration, and neuronal differentiation. During this period the brain is made of radial glial cells (RGCs), apical and basal intermediate progenitor cells, and early born post-mitotic neurons (Budday et al., 2015). A study of the epigenetic factor landscape in the developing mouse brain found that ASH1L was specifically enriched in apical and basal intermediate progenitor cells , while members of the MLL complex (H3K4me3-depositing complex) and demethylases were enriched in post-mitotic neurons, and members of the PRC2 complex were enriched in RGCs (Elsen et al., 2018).

Intermediate progenitor cells (IPC) are unpolarized cells in the ventricular and subventricular zone that divide symmetrically to generate upper neocortical layers, and influences cortical thickness (Pontious et al., 2008). IPCs also found in the intermediate zone, an area of white matter where neurons made in the ventricular zone migrate through in order to reach the cortical plate during

corticogenesis. This area eventually transforms into adult white matter (24pcws onwards) (Budday et al., 2015). Delayed myelination and microcephaly are phenotypes associated with de-novo mutations in ASH1L, and perhaps tied to disturbances in ASH1L expression in IPCs during critical points in the development of the neocortex. ASH1L may be important to the formation of the white matter of the intermediate zone and the migration of neurons across this white matter to the cortical plate. In the future I would analyze the expression of ASH1L using Allen Brain Atlas microarray data collected from the cortical plate and ventricular zone during human brain development the cortical plate and ventricular zone. This would allow ASH1L expression to be studied more on a cortical layer-based level, rather than brain structure-based level.



Age	Avg. Expression
8 pcw	2.3
9 pcw	3.41
12 pcw	3.33
13 pcw	3.46
16 pcw	3.36
17 pcw	3.03
19 pcw	2.43
21 pcw	3.37
24 pcw	2.81
35 pcw	2.05
37pcw	2.44

Figure 2. ASH1L expression analysis at pre and post-natal stages (A) *ASH1L* Expression Heat Map where each box represents the *ASH1L* expression level at one specific age in one structure of the brain. Red represents high levels of expression, while blue represents low levels of expression. The gradient is calculated over the entire period of development (all the columns). Grey boxes indicate no expression data was available for analysis. **(B)** Box-and-whisker plots shows median values of expression of *ASH1L* across the lifespan of the individual starting at fetal stages in the four structures of the PFC. Points outside of the box represent outlier data values. **(C)** *ASH1L* Expression Heat Map over prenatal brain developmental period. The table corresponds to the average expression of *ASH1L* at each age during prenatal development; green coloration represents higher expression and red represents low expression. **(D)** Box-and-whisker plots of gene expression in the PFC structures during prenatal development.

ASH1L DFC Prenatal Expression is similar to that of MLL and Demethylases of PcG Complexes

As a methyltransferase and TrxG member, ASH1L is also a member of the epigenetic factor landscape of development. However, ASH1L's modulation of the epigenetic landscape during neurodevelopment is not fully understood, but expression data can be used to compare *ASH1L* to other TrxG and PcG genes (Fig.3A-B). ASH1L expression data from one brain structure, the DFC, was used for the comparative analysis between *ASH1L* and TrxG or PcG genes. The DFC was chosen because all donors with ages in prenatal development had expression values for this structure, also the DFC is of interest due to its relation to ASD phenotypes. By comparing *ASH1L*'s expression to the expression of the genes of these complexes, we can discover the genes expressed most similarly to *ASH1L* which can implicate potential molecular interaction of these genes or suggest epigenetic mechanisms at a specific time point in brain development.

The MLL complex is a TrxG complex, known to interact with ASH1L. The complex contains *MLL3* or *KMT2C*, an ASD-associated gene, which similar to ASH1L, is a histone methyltransferase (H3K4me3) important to gene activation and implicated in neuronal differentiation (De Rubeis et al., 2014; Elsen et al., 2018; Schuettengruber et al., 2017). The histone mark, H3K4, can be demethylated by three proteins: KDM1A, KDM5B, KDM2B. In contrast, ASH1L's histone mark, H3K36, has one demethylated by KDM2B. My analysis shows that *ASH1L* is expressed in similar levels to *KDM2B*, and *MLL3*. Similarly, expression of *KDM1A* and *KDM5B* are higher than the other genes, suggesting that there might be greater demethylase activity rather than methylase activity in the DFC during this time period (Fig.3C). Moreover, target genes of MLL3 during this time period may be more likely to be repressed given higher levels of H3K4 demethylase expression. Hierarchical clustering

further validates the similarities observed between the methyltransferases and demethylases (Fig.3C).

The PRC1 complex is responsible for the mono-ubiquitination of lysine 119 on histone H2A, (H2AK119), and canonical recruitment of the PRC2 complex (Schuettengruber et al., 2017). The complex is made up of different components depending on its function. Canonically, PRC1 leads to repression of genes through its recruitment of the PRC2 complex; the non-canonical PRC1 complex is different. The H3K36 demethylase, *KDM2B*, recruits the PRC1 non-canonical complex (RYBP, PCGF3/5, RBBP7, RING1), and then recruits the PRC2 complex and promote gene repression. (Corley and Kroll, 2015). I find that most of the PRC1 components are expressed at higher levels than *ASH1L* during prenatal development, except for *PCGF5*. ASH1L is linked through hierarchical clustering to *PCGF3* both in prenatal development (Fig.3D) and first-early second trimester development (Fig.3B). The first-early second trimester development association is also notable as *PCGF5* is most closely clustered with *KDM2B*, then *ASH1L* (Fig. 3B). Further analysis of the PRC1 complex, KDM2B, and ASH1L, could help to elucidate the mechanistic link between the two epigenetic antagonistic complexes during neuronal development.

In addition to the PRC1 and MLL complexes, the PRC2 complex is also active during prenatal brain development, especially during early fetal development. I find that members of the catalytic component of the PRC2 complex (EED, RBBP7 SUZ12, and EZH2) have a high early expression trend that drops off significantly around 13pcw (Fig.3E). The timing of expression in the DFC is supported by studies that have shown the PRC2 complex to be most active during neurogenesis and cell

proliferation (Corley and Kroll, 2015). EZH1 has the opposite trend, increasing significantly in expression at the end of prenatal development. This observation in consistent with findings that suggest EZH1 contributes to the maturation of post-mitotic neurons (Liu et al., 2018). Our analysis also indicates that EZH2 is notable for its range in expression during prenatal DFC development. Being the catalytic driver of the PRC2 complex, responsible for the orchestration of proliferation and differentiation developmental switches, it is unsurprising that *EZH2* has such a large range of expression.

The clustering analysis of the PRC2 complex expression during development implicates a relationship between ASH1L and the H3K27 demethylases: KDM6A, and KDM6B (Fig.2E and Fig.2b). One study found that the two demethylases were upregulated in IPCs and early born neurons of the ventricular zone, and led to the de-repression of neuronal differentiation genes (Elsen et al., 2018). Our analysis, which shows similar expression levels of H3K27 demethylases and ASH1L, also support previous findings that report high expression of ASH1L in IPCs. Having similar expression levels to genes expressed in the progenitor regions and important to neurogenesis, suggests that ASH1L may have role in regulating proliferation/differentiation in neuronal stem cells; a role of ASH1L reported in in mesenchymal and embryonic stem cells (Kanellopoulou et al., 2015; Yin et al., 2019). However, the extent to which ASH1L regulates neuronal differentiation is largely unknown.

In conclusion, the comparative analysis performed here elucidates trends in expression that are similar across different TrxG and PcG complexes. Clustering analysis during early prenatal development, links methyltransferases of the PRC2 complex to subunits of the PRC1 complex.

Similarly, members of the TrxG complex (*ASH1L* and *MLL3*) cluster with PRC2 demethylases and *PCGF3* of the PRC1 complex (Fig.3B and Appendix B). Furthermore, looking at the three heat maps (Fig 2C-E), *ASH1L* exhibits most similar expression patterns to that of the MLL complex (Fig 2A). This finding correlates with the strong functional interaction of ASH1L and MLL observed in leukemia cells (Zhu et al., 2016a). When studying ASH1L and its targets, it may also be important to consider the MLL complex and its target genes.

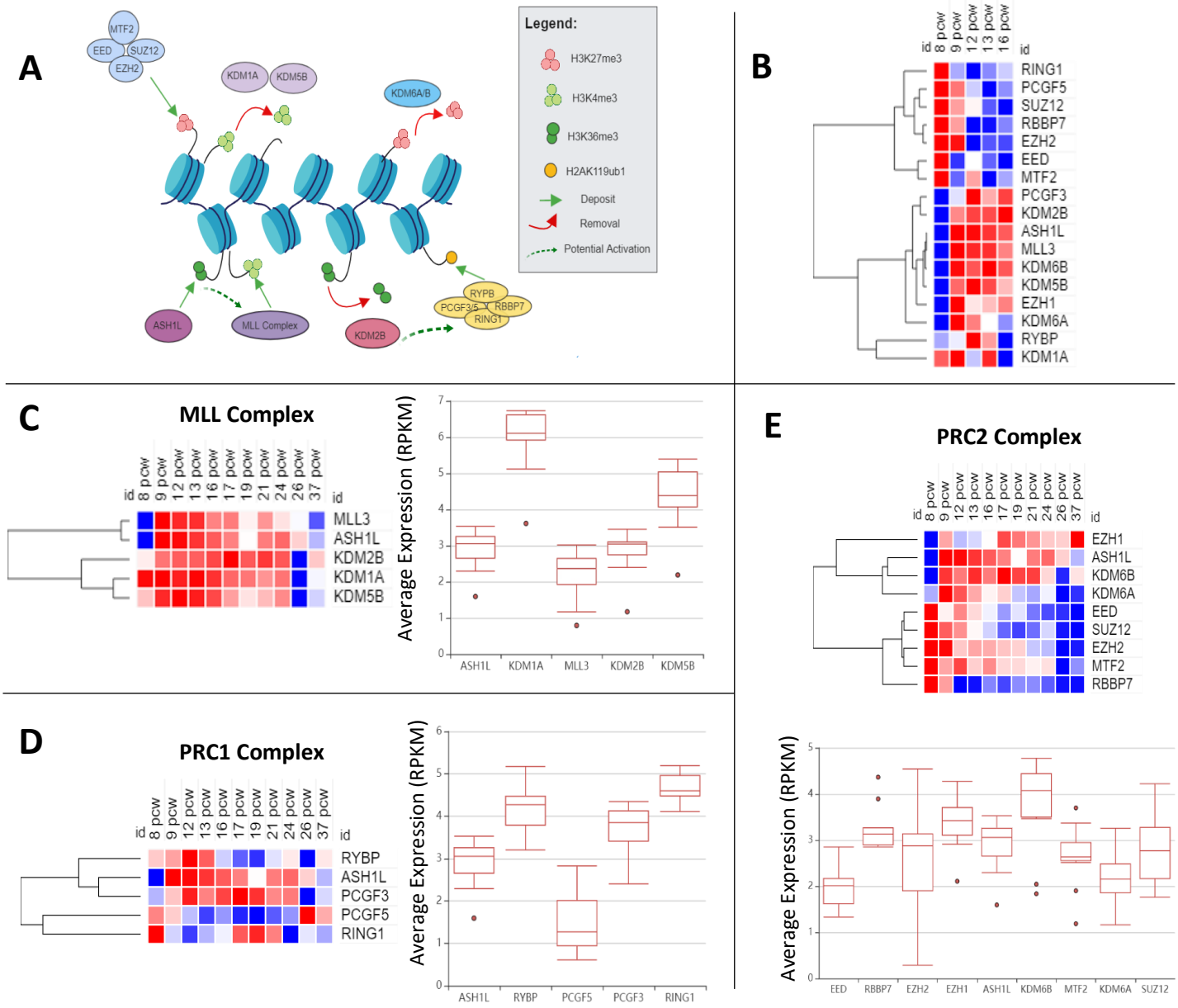


Figure 3. Comparison of MLL, PRC1, and PRC2 Complexes in the Development of the DFC. **(A)** Hypothesized schematic of the TrxG and PcG interactions during chromatin remodeling. (**B)** Heat map and hierarchical clustering of the expression of all TrxG and PcG relevant genes during 8-16pcw of development **(C-E)** Heat map analysis, hierarchical clustering, and box-and-whisker plots of gene expression for the relevant genes of the MLL, PRC1, and PRC2 complexes and ASH1L, shown respectively in that order.

CHAPTER 2

ASH1L CO-EXPRESSION PATTERNS IN THE PFC CORTEX DURING PRENATAL DEVELOPMENT

Transcriptome data is not only useful in understanding spatiotemporal patterns of gene expression, it also can be used to generate gene co-expression networks. Co-expression network analysis is one method used to infer potential gene function and gene-disease associations. This approach constructs networks of genes with a tendency to show a coordinated expression patterns across a group of samples, and subsequently analyzes these networks to elucidate functions of correlated gene pairs and clustered modules. In this chapter, I present my studies in neuronal co-expression networks associated with *ASH1L* expression during prenatal development.

ASH1L Positively Correlated Genes in the PFC Have Functional Enrichments Important to Neuronal Projection Development

To further understand the function of ASH1L during neurodevelopment, a co-expression network was created using the top 500 positively correlated genes to *ASH1L*, genes with r values above 0.74 in the PFC during prenatal development. The genes with expression most strongly correlated *ASH1L* expression in the PFC over the prenatal period were *KIF1B, BIRC6, EP300,* and *MLL5* (0.947<r <0.959); *KIF1B* is a motor protein, *BIRC6* is a E3 ubiquitin-protein ligase (anti-apoptotic protein), *EP300* is histone acetyltransferase (H3K27ac), *MLL5* is a histone methyltransferase (H3K4me3). Functional enrichment for the nodes (top correlated genes) of the network was performed using ToppFun. Cellularly, the correlated genes were enriched in the nucleoplasm part of the cell in addition to different parts of neuron (synapse, axon, dendrite). Genes in this network were also highly significantly enriched in neuron projection, microtubule end, and growth cone of the cell (Fig.4A). Gene Ontology for Biological Processes indicated significant enrichment in genes important to epigenetic processes, such as histone modification, chromatin organization, and

histone methylation, which fits into the canonical function of ASH1L. Enrichment analysis also revealed that co-expressed genes were enriched in neuron development, neuron differentiation, and interestingly neuron projection development and morphogenesis. Genes were also enriched in the biological process of regulation of neuron projection development and cell projection organization (Fig.4B).

The enrichments associated with cytoskeleton and neuronal morphogenesis are interesting in the context of in-vitro ASH1L shRNA KD experiments performed by Seonhye Chen, of the Lizarraga lab, in hESC-derived cortical-like neurons. I conducted morphometric analysis of ASH1L KD neurons using Neurolucida software and found that shRNA transfected neurons exhibited abnormal cytoskeletal features shorter neurite length and decreased branching in comparison to GFP controls (Fig.4C-E). The underlying mechanism is not well understood, but the overrepresentation of genes important to cell projection morphogenesis in the ASH1L co-expression network suggest that ASH1L plays a morphometric role in neurodevelopment.

The hESC-derived neurons resemble early development (8-16pcw) dorsolateral prefrontal cortex neurons (Mariani et al., 2015). Given this resemblance, a second ASH1L correlative search was performed to find the top 500 genes in the DFC during 8-16pcw of development. Functional enrichment yielded similar GO biological function and GO molecular function results, but these genes were slightly more enriched for neuron projection development in the DFC (p-value= 9.41E-14) and in the PFC (p-value=1.50E-13) (Appendix D).

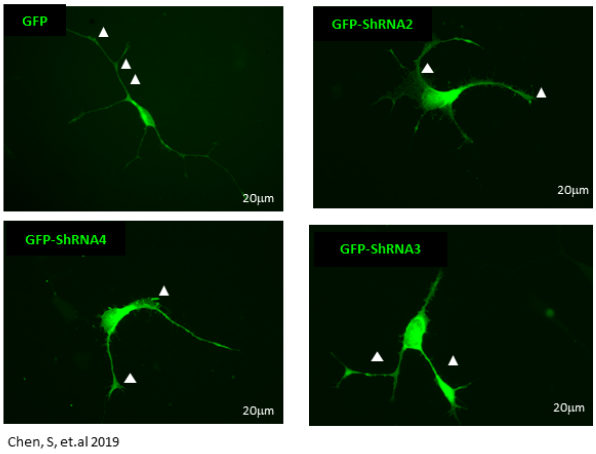
A

Cellular Component	P-value
nucleoplasm part	2.31E-12
somatodendritic compartment	6.92E-09
neuron part	1.90E-08
synapse	2.19E-08
transferase complex	2.23E-08
neuron projection	1.76E-07
postsynapse	3.06E-07
histone methyltransferase complex	3.28E-07
synapse part	3.32E-07
histone acetyltransferase complex	3.33E-07
axon	4.07E-07
dendrite	9.69E-07
acetyltransferase complex	1.20E-06
catalytic complex	1.51E-06
cell projection part	3.05E-06

B

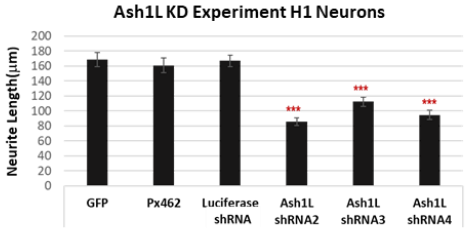
Biological Process	P-value
histone modification	1.86E-14
chromatin organization	4.69E-14
covalent chromatin modification	5.53E-14
neuron projection development	1.50E-13
neuron projection morphogenesis	1.29E-12
neuron development	3.54E-12
generation of neurons	6.72E-12
neurogenesis	7.72E-12
peptidyl-lysine modification	1.01E-11
neuron differentiation	4.34E-11
cell projection morphogenesis	7.48E-11
cell morphogenesis involved in neuron differentiation	1.22E-10
chromosome organization	1.53E-10
histone lysine methylation	1.56E-10
cell part morphogenesis	1.71E-10
regulation of neuron projection development	5.26E-10
cell morphogenesis involved in differentiation	5.40E-10

C



Chen, S, et.al 2019

D



E

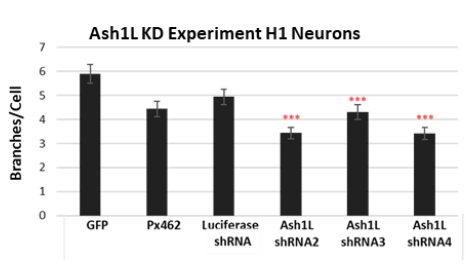


Figure 4. ASH1L might regulate neuronal morphogenesis (A) Stem cell derived H1 cortical neurons at day 35 of neuronal induction were nucleofected with ASH1L-shRNAII+GFP constructs (top right panel), ASH1L-shRNAIII+GFP (bottom right panel), ASH1L-shRNAIV+GFP (bottom left panel) GFP construct alone (left panel). White arrows indicate longest neurite**. (B)** Histogram representation of neurite length analysis shows a reduction in neurite length in ASH1L-shRNA-II, ASH1L-shRNA-III and ASH1L-shRNA-IV compared to control GFP; ***p<3.00E-05 **(C)** Histogram representation of branches per cell analysis shows a reduction in branches in ASH1L-shRNA transfected neurons compared to GFP control; ***p<0.001 **(D)** GO Cellular Component Analysis of top 500 ASH1L correlated genes in the PFC during prenatal development. **(E)** GO Biological Function Analysis of top 500 ASH1L correlated genes in the PFC during prenatal development

ASH1L Top Correlated Genes in Fetal PFC Overlap with many ASD-risk Genes

The ASH1L co-expression networks were analyzed to find enrichment of disease-related genes. In the PFC and DFC co-expression analysis, genes in both sets were enriched highly for intellectual disability and Autism, and in the PFC for speech delay (Fig.5A-B). These associated disease phenotypes are have been reported numerous times in clinical observations seen in patients with de-novo ASH1L mutations (Okamoto et al., 2017; Shen et al., 2019). Genes in both networks also showed to overlap with ASD-risk genes found in the SFARI database (Fig.5C-D). Forty-nine genes were shared between the two co-expression networks and SFARI database genes (Appendix D). Recent studies have found that ASD risk genes are co-expressed in the mid-fetal human cortex, suggesting ASD-risk genes converge in a regulatory network during neurodevelopment (Cotney et al., 2015). My co-expression analysis implicates ASH1L as potential member of this ASD-related regulatory network.

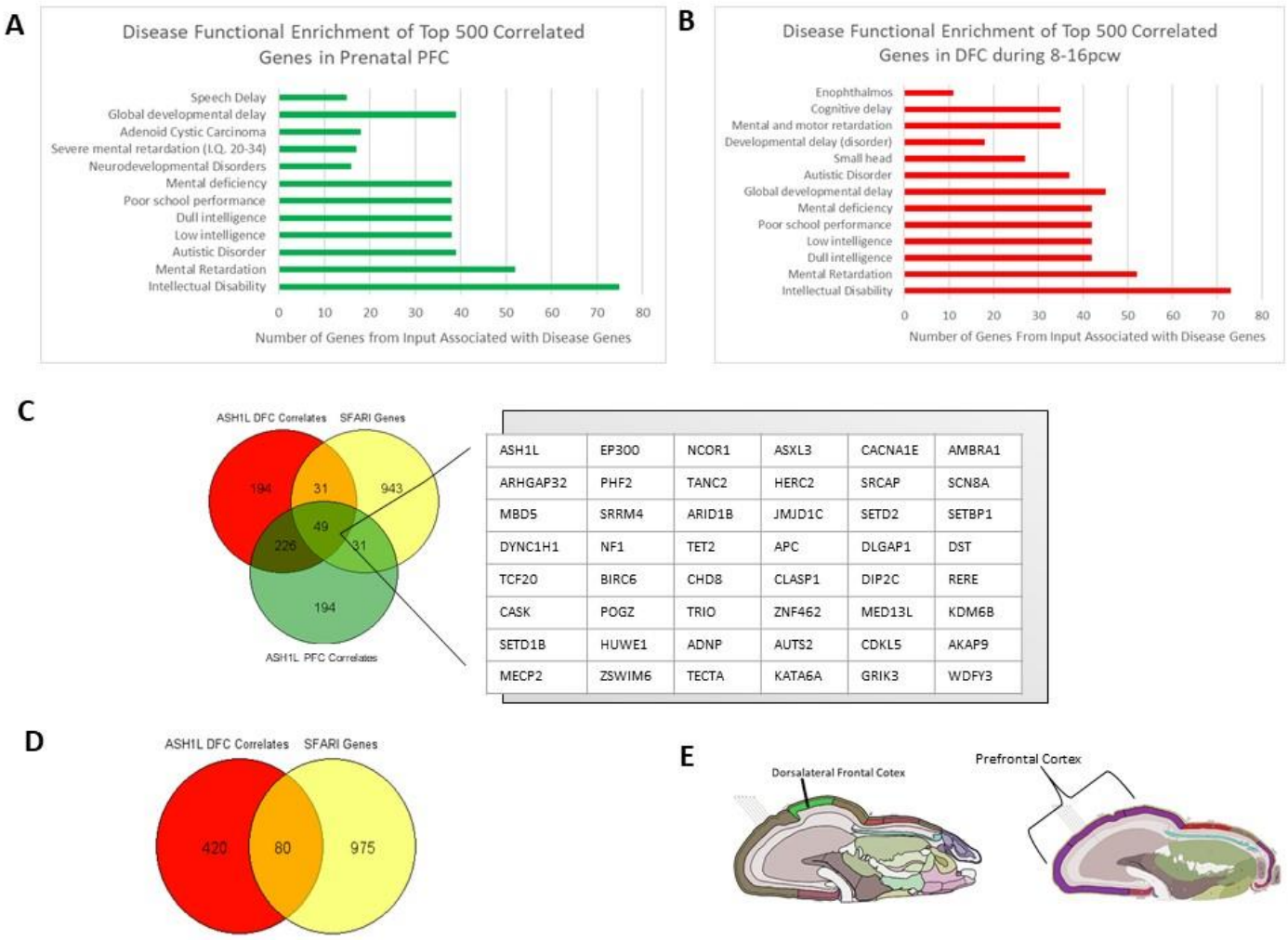


Figure 5. ASH1L Correlated Genes are enriched for disease phenotypes seen in ASD individuals with de-novo mutations in ASH1L (**A)** Disease Functional Enrichment for top correlated genes in the prenatal PFC; p-values associated with disease enrichment are found in Appendix C. **(B)** Disease Functional Enrichment for top correlated genes in 8-16pcw **(C)** Overlap of the genes in both correlative studies and genes from the SFARI database. The 49 genes at the intersection of the datasets are represented in the table. **(D)** Eighty genes are represented at the intersection of DFC correlated genes and SFARI genes **(E)** Diagrams of the DFC and PFC human brain tissue at 16pcw used to perform the developmental transcriptome study in the Allen Brain Atlas

CHAPTER 3

ASH1L DFC CO-EXPRESSION AND PROTEIN-INTERACTION NETWORK DURING 8-16PCW OF DEVELOPMENT

Co-expression networks are limited in what they can communicate about gene interactions. Networks are effectively only able to identify correlations an do not confer information on causality or distinguish between regulatory and regulated genes. A method which may confer regulatory information from co-expression networks, is the increasingly used method of differential co-expression analysis. This approach identifies genes with varying co-expression partners during different conditions. Using a method like that of Willsey et.al, a differential co-expression network was created manually using correlation data from the Allen Brain Atlas (Fig.6). Four co-expression networks were generated using this method described in Figure 6, and then analyzed for significant protein interactions within the network using the STRING database (Fig.7, Fig.8, Appendix E).

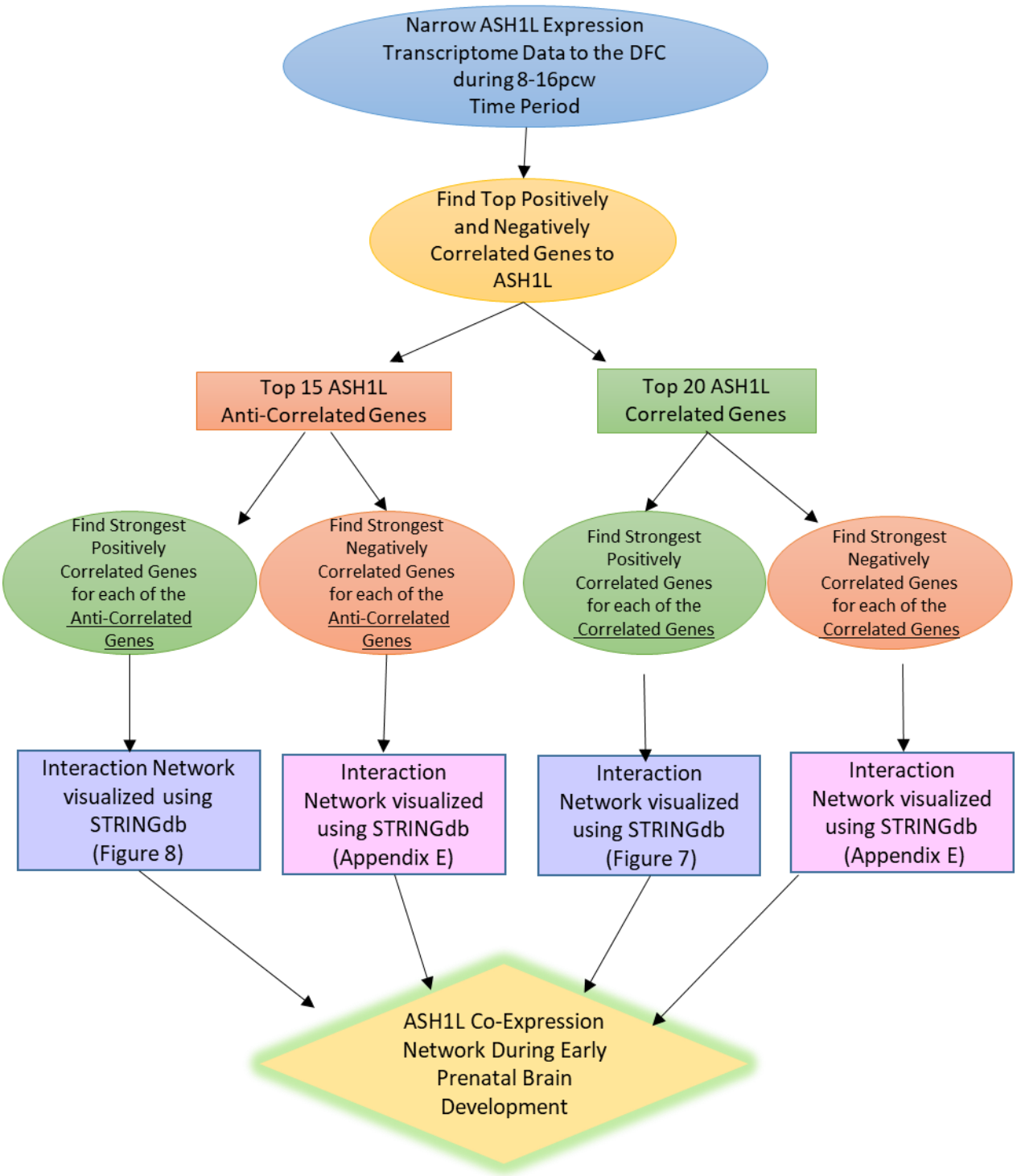


Figure 6. Differential Co-Expression Network Creation Schematic

ASH1L Protein Interaction Network is associated chromatin Modification, protein ubiquitination, ASD pathogenesis, and neurotrophin signaling pathways

The network of positive correlates of the positive ASH1L correlated genes, which will be called the positive network, is useful in understanding the interaction of groups of genes that co-activate with ASH1L in early DFC development. STRING database was used to predict significant protein interactions within the gene correlation data set (Fig.7). The interaction network was made from 141 of the 433 genes and had significant Protein-Protein Interaction (PPI) p-value of 0.000637, meaning the proteins in this network have more interactions among themselves than what would be expected for a random set of proteins of similar size, drawn from the genome. Such an enrichment indicates that the proteins are at least partially biologically connected, as a group.

Functional enrichment in this network shows an overrepresentation of proteins important to ubiquitination, cytoskeleton development, chromatin modification, and the neurotrophic signaling pathways. The ubiquitination-related proteins are represented by the pink nodes in Fig.7. Studies have shown targeted protein degraded through ubiquitination to be important during early neuronal differentiation (Saritas-Yildirim and Silva, 2014). In the context of epigenetics, ubiquitination performed by the PRC1 complex on H2AK119 has also been reported to be important to neuronal differentiation (Yao et al., 2018). This functional enrichment is very interesting because, as elaborated on in Chapter 1, ASH1L was clustered with the expression of the PCGF3 component of the PRC1 complex. Significant enrichment for cytoskeleton-related proteins and biological functions (microtubule binding, adherens junctions, tubulin binding) is also of interest given findings in Chapter 2 that suggest ASH1L correlates are important to cytoskeleton development.

Additionally, of interest, many of the nodes important to the cytoskeleton are also genes associated with ASD, as indicated by the dark blue and maroon coloration.

Chromatin modifying enzymes are also overrepresented in the functional enrichment of this network (yellow nodes in Fig. 7). The False Discovery Rate (FDR) was lowest for this Reactome pathway (1.51 E-05) for all enriched pathways. ASH1L's interaction with EP300 has an interaction score of 0.709; the interaction scores in on a scale of 0-1 and is an approximate probability that a predicted link exists between two proteins. The interaction between EP300 and ASH1L relates ASH1L to the module of genes most associated to chromatin modification. EP300 has very high interaction scores with KMT2C (MLL3) and CREBBP. MLL3, as mentioned in Chapter 1, is a TrxG protein important to gene activation during neuronal differentiation. CREBBP is a histone acetyltransferase which acetylates H3K27, antagonizing Polycomb silencing (Tie et al., 2014). The relationship and co-activation of these genes may elucidate a network of chromatin modifier important to turning on and off genes important in the development of the DFC. Future CHIP studies could be performed to investigate ASH1L intersection with other proteins pertinent to chromatin modification.

Finally, analysis of enrichment of Reactome pathways in this gene set, links the co-expression network to downstream signaling pathways of the human neurotrophin signaling pathway (Appendix H). This pathway involves the binding of neurotrophic factors such as BDNF, NT3, and NGF to Trk receptors which initiates signaling cascades such as the MAPK signaling pathway (dark green node in Fig.5) important to cellular differentiation and axonal outgrowth, Ubiquitin mediated

proteolysis (pink nodes) important to retrograde transport, and P13K-AKT pathway whose downstream genes are important in the regulation of actin cytoskeleton and cell survival (Kanehisa Laboratories, 2018; Khotskaya et al., 2017). *NTRK2*, the gene encoding for the BDNF receptor, Trkb, is an upstream regulator of the P13K-AKT pathways. The Lizarraga lab has performed QPCR analysis on *ASH1L* KD neurons and found a downregulation of NTRK2 (Appendix G). Other Trk receptor encoding genes such as NTRK1, showed no downregulation in KD neurons. The presence of Trkb-pathway related genes within this co-expression network suggests that ASH1L may regulate genes important to neurotrophic signaling. Furthermore, the neurotrophic signaling pathway's downstream effect on the regulation of actin cytoskeleton is relevant to understanding the morphology seen in *ASH1L* KD neurons (Fig. 4A, Appendix F). In summary, the positive co-expression network helps elucidate potential protein interactions relevant to ASH1L, as well as the functional importance of these interactions. Future studies can further investigate the presence of these interactions under different conditions, such as different developmental stage, different brain structure, or in a disease-state (*ASH1L* mutant or knock down).

ASH1L Positive Correlates' Co-Expression Network

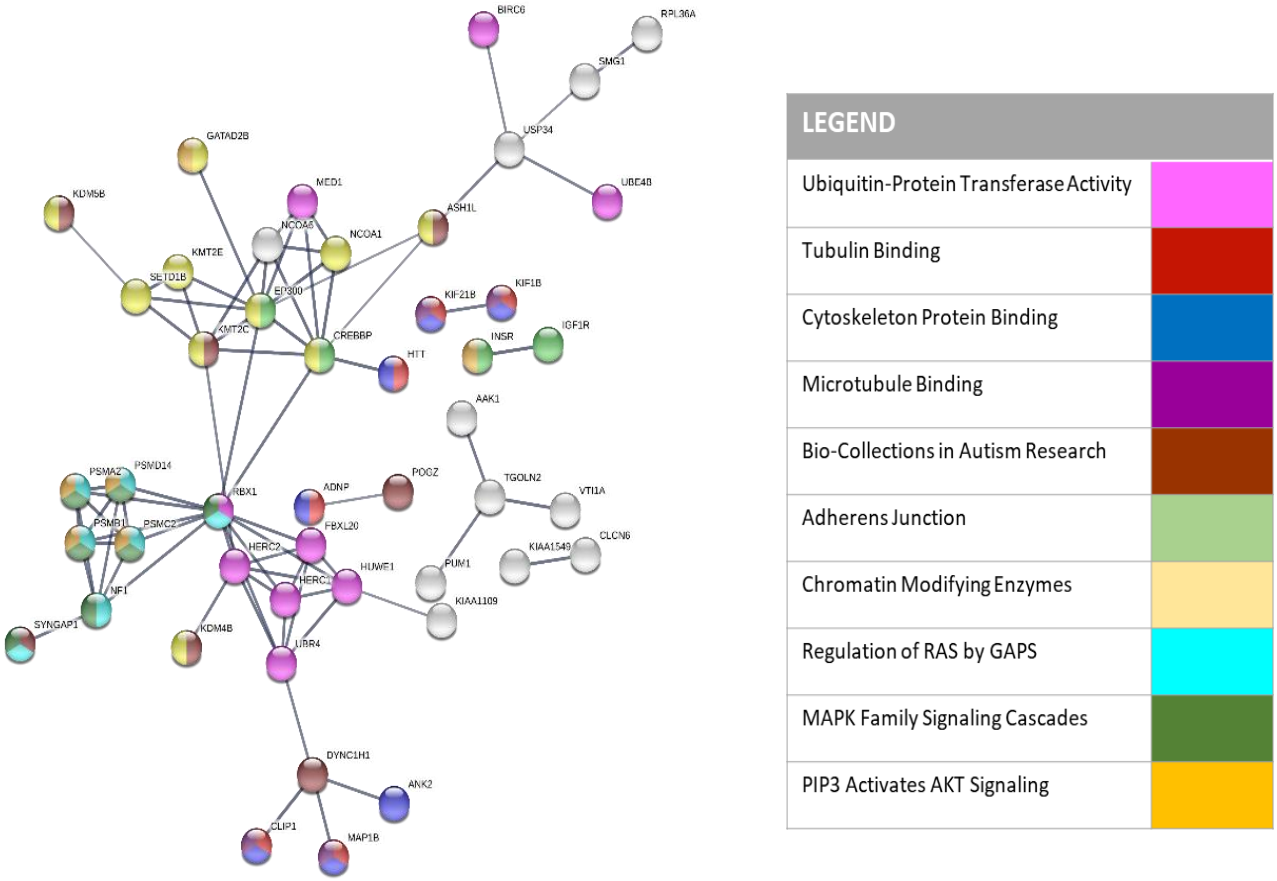
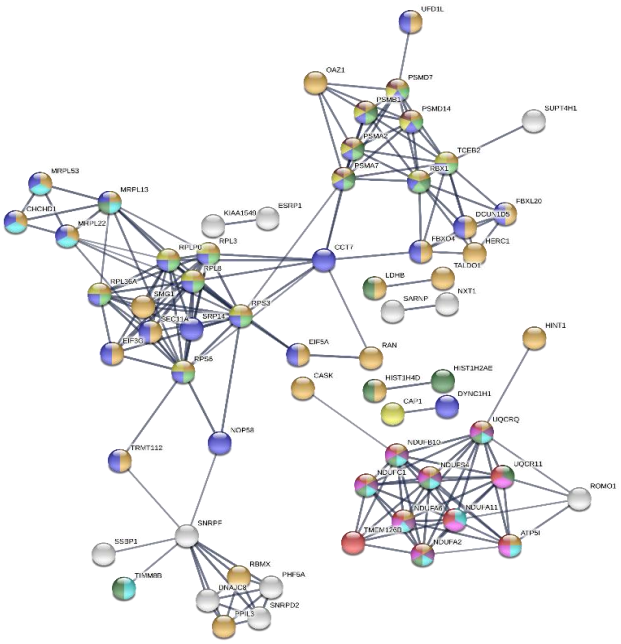


Figure 7. Co-Expression Protein Interaction Network for the positively correlated genes of the positive correlated genes of ASH1L in the DFC during 8-16pcw of development. Nodes represent genes and edges represent interaction between two nodes. The thickness of the edges indicates the strength of data support. The node colors and significance of the colors are summarized in the adjacent legend. Grey nodes are nodes that are not enriched for any of the attributes described in the legend. This network is like the Negative Correlates of the Negative Correlates of ASH1L co-expression network (Appendix E Fig.7). Both networks contain clusters of genes around *EP300, HERC1, PSMC2,* and *USP34*.

ASH1L Anti-Correlation Network

Interactions between anti-correlates of *ASH1L* provide valuable information about genes that are inactive in the same pattern that *ASH1L* is activated. The anti-correlate co-expression network was found by finding the positively correlated genes for each of the 15 *ASH1L* negatively correlated genes. The co-expression network had a PPI p-value of 1.33E-15, meaning that the network interactions are highly significant. Genes in this network are highly enriched for mitochondria-related biological functions (green, cyan, purple, and pink nodes). There are also genes enriched in processes important to neuronal migration such as axon guidance and regulation of SLITs and ROBOs. The functional enrichment of these genes and their association with current literature on *ASH1L* is not especially clear. Further research into the expression of these genes during *ASH1L* KD experiments would be insightful in understanding ASH1L's regulatory in activation and repression of specific modules of genes

ASH1L Anti-Correlates' Co-Expression Network



LEGEND	
Oxidative Phosphorylation	
Respiratory Electron Transport, ATP Synthesis	
Metabolism of Proteins	
Mitochondrial ATP Synthesis Coupled Electron Transport	
MAPK6/MAPk4 Signaling	
Regulation of Expression of SLITs and ROBOs	
Axon Guidance	
Mitochondrial Protein Complex	
Genes Related to Mitochondrial Function	
Organonitrogen Compound Metabolic Process	

Figure 8. Co-Expression Protein Interaction Network for the positively correlated genes of the negatively correlated genes of ASH1L in the DFC during 8-16pcw of development. Clusters of genes enriched for functions related to the mitochondria are seen in the lower left corner of the network. Clusters of genes associated with axonal guidance and metabolism are seen to the upper portion of the network. This network is like the Negative Correlates of the Positive Correlates of ASH1L co-expression network (Appendix E Fig.8). Both networks contain clusters of genes around *NDUF34, PSMA2, PHF5A,* and *MRPL13*.

CONCLUSION

The transcriptomic analysis of the developing human brain provides new information into the orchestration and activation of different networks of genes during different time points in development. Here, we describe ASH1L's predicted role in early neurodevelopment based on developmental transcriptome data. ASH1L is highly expressed in most structures of the brain during the early-mid fetal stage and again starting in adolescence. Furthermore, in the structures of the PFC, ASH1L is seen to be expressed at similar levels and in pattern with the TrxG gene *MLL*3 and demethylases on the PRC2 complex. The timing and similar expression of these genes may elucidate the molecular mechanism of TrxG activation and PcG silencing, seen in other species. Genes with expression patterns to ASH1L in specific contexts, like the co-expressed genes reported in Chapter 2 and 3, are proteins that could be interacting partners with ASH1L or genes that ASH1L helps to regulate through its function a chromatin modifier. Our analyses of these co-expression network in the PFC and DFC, implicated gene clusters associated with neuronal cytoskeleton development, chromatin modification, and neurotrophin signaling pathways.

Through transcriptome analysis, this study also finds that ASH1L is highly associated with ASD and Intellectual Disability; a finding well supported by previously published GWAS ASD studies, clinical case reports, ASD co-expression network studies. ASH1L, as an epigenetic modifier, can affect the activation and repression of many genes in response to the environment and activity of other chromatin modifying complexes. A de-novo LoF mutation in this gene can therefore affect a wide range of genes and downstream processes; processes that can lead to the neurodevelopmental

pathogenesis in individuals who acquire the mutation during early prenatal development. A better understanding of ASH1L's target genes and effect on cellular pathways is therefore important in understanding the pathogenesis of LoF-related ASD and ID. Furthermore, ASH1L is reported here to be associated with many other significant ASD-risk genes, thereby making the understanding of ASH1L useful in understanding ASD pathogenesis of related and co-expressed ASD-risk genes.

Future studies should be performed to validate these transcriptome results in disease-state models, like in ASH1L KD human neurons or ASH1L mutant mice. In these experiments, the perturbation of ASH1L could lead to disturbances in genes most strongly correlated to ASH1L, or biological processes enriched in these co-expression networks. Additional bioinformatic studies of ASH1L in different brain regions, especially the MD-CBC would be an interesting direction to study. Changes in co-expression patterns of ASH1L across different developmental time periods and/or brain structures would further explain ASH1L's role in human brain development. Finally, investigations into ASH1L on the chromatin level, using CHIPseq and ATACseq will be useful in understanding ASH1L target genes, some of which may be reported here, as well as understanding ASH1L's interaction with other histone marks and chromatin modifiers. Ultimately, my findings here provide a glimpse into the complex yet important role ASH1L plays in human neurodevelopment

BIBLIOGRAPHY

Abrahams, B.S., Arking, D.E., Campbell, D.B., Mefford, H.C., Morrow, E.M., Weiss, L.A., Menashe, I., Wadkins, T., Banerjee-Basu, S., and Packer, A. (2013). SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). Mol. Autism *4*, 36.

Baio, J. (2018). Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2014. MMWR Surveill. Summ. *67*.

Budday, S., Steinmann, P., and Kuhl, E. (2015). Physical biology of human brain development. Front. Cell. Neurosci. *9*.

Castiglioni, I., Caccia, R., Garcia-Manteiga, J.M., Ferri, G., Caretti, G., Molineris, I., Nishioka, K., and Gabellini, D. (2018). The Trithorax protein Ash1L promotes myoblast fusion by activating Cdon expression. Nat. Commun. *9*, 5026.

Chen, J., Aronow, B.J., and Jegga, A.G. (2009). Disease candidate gene identification and prioritization using protein interaction networks. BMC Bioinformatics *10*, 73.

Corley, M., and Kroll, K.L. (2015). The Roles and Regulation of Polycomb Complexes in Neural Development. Cell Tissue Res. *359*, 65–85.

Cotney, J., Muhle, R.A., Sanders, S.J., Liu, L., Willsey, A.J., Niu, W., Liu, W., Klei, L., Lei, J., Yin, J., et al. (2015). The autism-associated chromatin modifier CHD8 regulates other autism risk genes during human neurodevelopment. Nat. Commun. *6*.

De Rubeis, S., He, X., Goldberg, A.P., Poultney, C.S., Samocha, K., Cicek, A.E., Kou, Y., Liu, L., Fromer, M., Walker, S., et al. (2014). Synaptic, transcriptional, and chromatin genes disrupted in autism. Nature *515*, 209–215.

Donovan, A.P.A., and Basson, M.A. (2017). The neuroanatomy of autism – a developmental perspective. J. Anat. *230*, 4–15.

Durak, O., Gao, F., Kaeser-Woo, Y.J., Rueda, R., Martorell, A.J., Nott, A., Liu, C.Y., Watson, L.A., and Tsai, L.-H. (2016). Chd8 mediates cortical neurogenesis via transcriptional regulation of cell cycle and Wnt signaling. Nat. Neurosci. *19*, 1477–1488.

Elsen, G.E., Bedogni, F., Hodge, R.D., Bammler, T.K., MacDonald, J.W., Lindtner, S., Rubenstein, J.L.R., and Hevner, R.F. (2018). The Epigenetic Factor Landscape of Developing Neocortex Is Regulated by Transcription Factors Pax6→ Tbr2→ Tbr1. Front. Neurosci. *12*.

Habibi, E., and Stunnenberg, H.G. (2017). Transcriptional and epigenetic control in mouse pluripotency: lessons from in vivo and in vitro studies. Curr. Opin. Genet. Dev. *46*, 114–122.

Huang, C., and Zhu, B. (2018). Roles of H3K36-specific histone methyltransferases in transcription: antagonizing silencing and safeguarding transcription fidelity. Biophys. Rep. *4*, 170–177.

Kanehisa Laboratories (2018). KEGG PATHWAY: Neurotrophin signaling pathway - Homo sapiens (human).

Kanellopoulou, C., Gilpatrick, T., Kilaru, G., Burr, P., Nguyen, C.K., Morawski, A., Lenardo, M.J., and Muljo, S.A. (2015). Reprogramming of Polycomb-Mediated Gene Silencing in Embryonic Stem Cells by the miR-290 Family and the Methyltransferase Ash1l. Stem Cell Rep. *5*, 971–978.

Khotskaya, Y.B., Holla, V.R., Farago, A.F., Mills Shaw, K.R., Meric-Bernstam, F., and Hong, D.S. (2017). Targeting TRK family proteins in cancer. Pharmacol. Ther. *173*, 58–66.

Liu, P.-P., Xu, Y.-J., Teng, Z.-Q., and Liu, C.-M. (2018). Polycomb Repressive Complex 2: Emerging Roles in the Central Nervous System. The Neuroscientist *24*, 208–220.

Loke, Y.J., Hannan, A.J., and Craig, J.M. (2015). The Role of Epigenetic Change in Autism Spectrum Disorders. Front. Neurol. *6*.

Mariani, J., Coppola, G., Zhang, P., Abyzov, A., Provini, L., Tomasini, L., Amenduni, M., Szekely, A., Palejev, D., Wilson, M., et al. (2015). FOXG1-dependent dysregulation of GABA/glutamate neuron differentiation in autism spectrum disorders. Cell *162*, 375–390.

Miller, J.A., Ding, S.-L., Sunkin, S.M., Smith, K.A., Ng, L., Szafer, A., Ebbert, A., Riley, Z.L., Royall, J.J., Aiona, K., et al. (2014). Transcriptional landscape of the prenatal human brain. Nature *508*, 199–206.

Miyazaki, H., Higashimoto, K., Yada, Y., Endo, T.A., Sharif, J., Komori, T., Matsuda, M., Koseki, Y., Nakayama, M., Soejima, H., et al. (2013). Ash1l Methylates Lys36 of Histone H3 Independently of Transcriptional Elongation to Counteract Polycomb Silencing. PLoS Genet. *9*.

Okamoto, N., Miya, F., Tsunoda, T., Kato, M., Saitoh, S., Yamasaki, M., Kanemura, Y., and Kosaki, K. (2017). Novel MCA/ID syndrome with ASH1L mutation. Am. J. Med. Genet. A. *173*, 1644–1648.

Pontious, A., Kowalczyk, T., Englund, C., and Hevner, R.F. (2008). Role of Intermediate Progenitor Cells in Cerebral Cortex Development. Dev. Neurosci. *30*, 24–32.

Reynolds, C., and Kamphaus, R. (2013). DSM5 Diagnostic Criteria Autism Spectrum Disorder.

Ronan, J.L., Wu, W., and Crabtree, G.R. (2013). From neural development to cognition: unexpected roles for chromatin. Nat. Rev. Genet. *14*, 347–359.

von Schimmelmann, M., Feinberg, P.A., Sullivan, J.M., Ku, S.M., Badimon, A., Duff, M.K., Wang, Z., Lachmann, A., Dewell, S., Ma'ayan, A., et al. (2016). Polycomb repressive complex 2 (PRC2) silences genes responsible for neurodegeneration. Nat. Neurosci. *19*, 1321–1330.

Schuettengruber, B., Bourbon, H.-M., Di Croce, L., and Cavalli, G. (2017). Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. Cell *171*, 34–57.

Shen, W., Krautscheid, P., Rutz, A.M., Bayrak-Toydemir, P., and Dugan, S.L. (2019). De novo loss-of-function variants of ASH1L are associated with an emergent neurodevelopmental disorder. Eur. J. Med. Genet. *62*, 55–60.

Stessman, H.A.F., Xiong, B., Coe, B.P., Wang, T., Hoekzema, K., Fenckova, M., Kvarnung, M., Gerdts, J., Trinh, S., Cosemans, N., et al. (2017). Targeted sequencing identifies 91 neurodevelopmental disorder risk genes with autism and developmental disability biases. Nat. Genet. *49*, 515–526.

Sun, B., Chang, E., Gerhartl, A., and Szele, F.G. (2018). Polycomb Protein Eed is Required for Neurogenesis and Cortical Injury Activation in the Subventricular Zone. Cereb. Cortex N. Y. NY *28*, 1369–1382.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., et al. (2015). STRING v10: protein–protein interaction networks, integrated over the tree of life. Nucleic Acids Res. *43*, D447–D452.

Tie, F., Banerjee, R., Saiakhova, A.R., Howard, B., Monteith, K.E., Scacheri, P.C., Cosgrove, M.S., and Harte, P.J. (2014). Trithorax monomethylates histone H3K4 and interacts directly with CBP to promote H3K27 acetylation and antagonize Polycomb silencing. Dev. Camb. Engl. *141*, 1129–1139.

Willsey, A.J., Sanders, S.J., Li, M., Dong, S., Tebbenkamp, A.T., Muhle, R.A., Reilly, S.K., Lin, L., Fertuzinhos, S., Miller, J.A., et al. (2013). Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. Cell *155*, 997–1007.

Yao, M., Zhou, X., Zhou, J., Gong, S., Hu, G., Li, J., Huang, K., Lai, P., Shi, G., Hutchins, A.P., et al. (2018). PCGF5 is required for neural differentiation of embryonic stem cells. Nat. Commun. *9*, 1463.

Yin, B., Yu, F., Wang, C., Li, B., Liu, M., and Ye, L. (2019). Epigenetic Control of Mesenchymal Stem Cell Fate Decision via Histone Methyltransferase Ash1l. STEM CELLS *37*, 115–127.

Zhu, L., Li, Q., Wong, S.H.K., Huang, M., Klein, B.J., Shen, J., Ikenouye, L., Onishi, M., Schneidawind, D., Buechele, C., et al. (2016a). ASH1L Links Histone H3 Lysine 36 di-methylation to MLL Leukemia. Cancer Discov. *6*, 770–783.

Zhu, T., Liang, C., Li, D., Tian, M., Liu, S., Gao, G., and Guan, J.-S. (2016b). Histone methyltransferase Ash1L mediates activity-dependent repression of neurexin-1α. Sci. Rep. *6*.

APPENDICES:

APPENDIX A: Table 1. Developmental Transcriptome Donor Information

Developmental Period	Donor ID	Age	Sex	Race
1st Trimester (1-13 pcw)	H376.IIA.51	8 pcw	M	European
	H376.IIA.50	9 pcw	M	European
	H376.IIB.50	12 pcw	F	Asian
	H376.IIB.51	12 pcw	F	African American
	H376.IIB.52	12 pcw	F	Two or more races
	H376.IIIA.50	13 pcw	M	European
	H376.IIIA.51	13 pcw	F	European
2nd Trimester (14-26pcw)	H376.IIIB.50	16 pcw	M	Hispanic
	H376.IIIB.51	16 pcw	M	European
	H376.IIIA.51	16 pcw	M	Two
	H376.IIIB.53	17 pcw	F	European
	H376.IVA.51	19 pcw	F	Hispanic
	H376.IV.54	21 pcw	M	African American
	H376.IV.50	24 pcw	M	White
	H376.IIA.51	26 pcw	F	African American
3rd Trimester	H376.V.50	35 pcw	F	White
	H376.V.53	37pcw	M	European
Infantood (0-1 year)	H376.VI.50	4 months	M	White
	H376.VI.51	4 months	M	European
	H376.VI.52	4 months	M	African American
	H376.VII.51	10 months	M	African American
	H376.VII.51	1 year	F	European
Early Childhood (2-4years)	H376.VIII.53	2 years	F	European
	H376.VIII.52	3 years	F	European
	H376.VIII.53	3 years	M	Hispanic
	H376.VIII.50	4 years	M	African American
Childhood	H376.IX.51	8 years	M	African American
	H376.IX.50	11 years	F	African American
Teenage Years	H376.X.51	13 years	F	African American
	H376.X.50	15 years	M	African American
	H376.X.53	18 years	M	European
	H376.X.52	19 years	F	European
Adulthood	H376.XI.60	21 years	F	European
	H376.XI.50	23 years	M	African American
	H376.XI.52	30 years	F	White
	H376.XI.53	36 years	M	White
	H376.XI.54	37 years	M	African American
	H376.XI.56	40 years	F	African American

APPENDIX B:

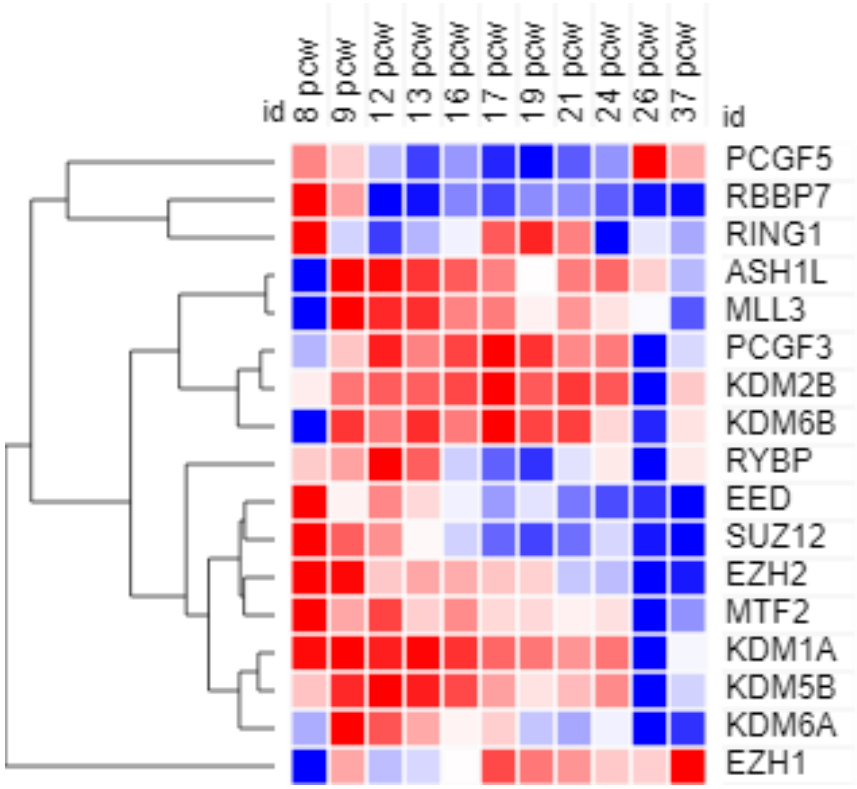


Figure 9. Clustering Analysis on the Expression Heat Map Constructed for genes of the PcG and TrxG complexes during prenatal development of the DFC.

APPENDIX C. ToppFun Results Top 500 Correlated Genes in the DFC during 8-16pcw period

GO: Molecular Functional Enrichment	P-VALUE
protein-containing complex binding	2.18E-12
GTPase activator activity	6.48E-10
GTPase regulator activity	1.03E-09
nucleoside-triphosphatase regulator activity	8.03E-09
microtubule binding	1.33E-08
transcription coregulator activity	4.37E-08
enzyme activator activity	6.05E-08
cytoskeletal protein binding	8.44E-08
tubulin binding	1.30E-07
guanyl-nucleotide exchange factor activity	2.77E-07
chromatin binding	6.04E-07
microtubule plus-end binding	1.34E-06

Disease	P-VALUE
Intellectual Disability	3.75E-21
Mental Retardation	2.04E-13
Low intelligence	2.57E-12
Poor school performance	2.57E-12
Mental deficiency	2.84E-12
Global developmental delay	7.18E-12
Autistic Disorder	1.63E-09
Small head	2.79E-09
Developmental delay (disorder)	4.05E-09
Mental and motor retardation	5.73E-09
Cognitive delay	5.73E-09

GO BIOLOGICAL FUNCTION	P-VALUE
regulation of small GTPase mediated signal transduction	4.40E-14
neuron projection development	9.41E-14
generation of neurons	1.69E-12
neuron development	1.77E-12
neuron projection morphogenesis	1.96E-12
positive regulation of GTPase activity	2.18E-12
neurogenesis	1.27E-11
neuron differentiation	1.47E-11
regulation of GTPase activity	3.48E-11
cell morphogenesis involved in neuron differentiation	5.92E-11
cytoskeleton organization	9.70E-11
regulation of nervous system development	2.63E-10
cell morphogenesis	3.68E-10

APPENDIX D:

Table 2. SFARI Gene Overlap

Shared Gene Categories	**Genes**	**Number of Genes**
ASH1L DFC 8-16pcw and SFARI Genes	ADARB1, ADCY5, AHDC1, ANK2, ARHGAP33, CDC42BPB, CHD3, DNAH10, DYRK1A, EXOC6B ,FGD1, GRID1, KDM4B, KDM5B, KIF21B, MNT, NCKAP5L, NFIX, NLGN2, NLGN3, NTRK3, OPHN1, PACS1, PER1, PRR12, SHANK2, ST8SIA2, SYNE1, SYNGAP1, SYNJ1, TSC1	31
ASH1L Prenatal Cortex (DFC, VFC, MFC 8-35pcw) and SFARI Genes	AGAP1 ANK3 ATRX CACNA1A CLSTN2 CNTNAP5 CTNND2 ELAVL3 EP400 GIGYF2 GPR139 GRIN2B GSK3B IL16 KCNB1 MTOR MYO5A NAV2 NBEA NCKAP5 NIPBL NSD1 SMARCC2 SYN3 TAF1L UNC79 VPS13B YEATS2 ZNF407 ZNF8 ZNF827	31
ASH1L Prenatal Cortex (DFC, VFC, MFC 8-35pcw), ASH1L DFC 8-16pcw, and SFARI Genes	ASH1L ARHGAP32 MBD5 DYNC1H1 TCF20 CASK SETD1B MECP2 EP300 PHF2 SRRM4 NF1 BIRC6 POGZ HUWE1 ZSWIM6 TECTA ADNP TRIO CHD8 CREBBP TET2 ARID1B TANC2 NCOR1 ASXL3 HERC2 JMJD1C APC CLASP1 ZNF462 AUTS2 KAT6A GRIK3 CDKL5 MED13L DIP2C DLGAP1 SETD2 SRCAP CACNA1E AMBRA1 SCN8A SETBP1 DST RERE KDM6B AKAP9 WDFY3	49

APPENDIX E:

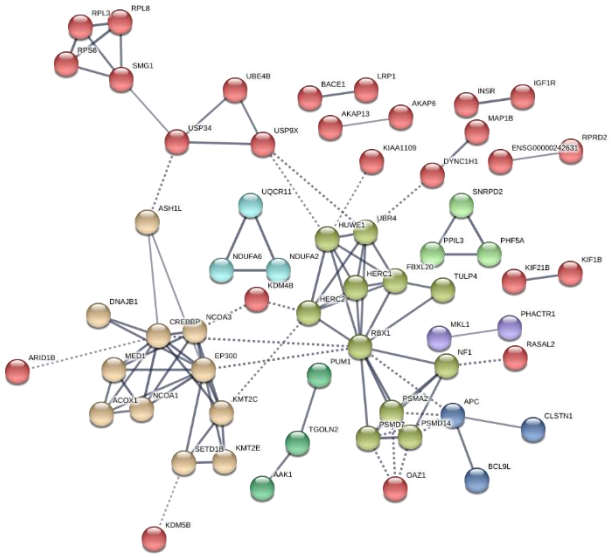


Figure 10. Co-Expression Protein Interaction Network for the negatively correlated genes of the negatively correlated genes of ASH1L in the DFC during 8-16pcw of development. Clusters of genes are colored based on k-means clustering.

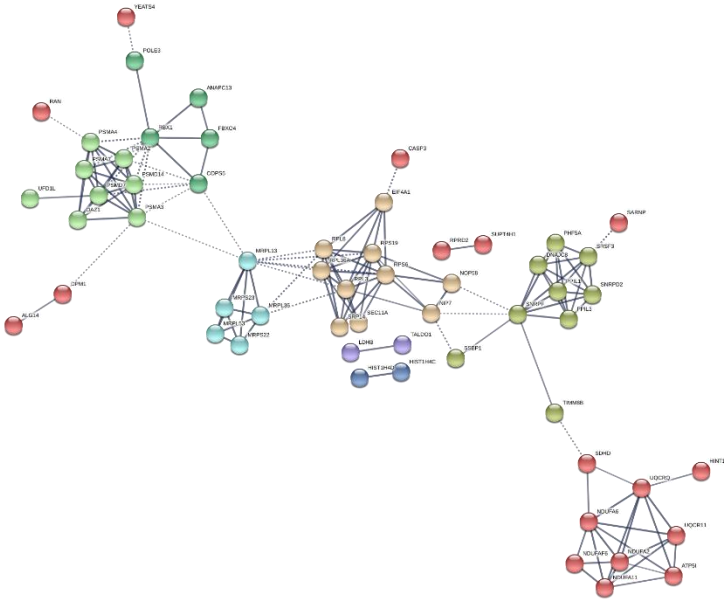


Figure 11. Co-Expression Protein Interaction Network for the positively correlated genes of the negatively correlated genes of ASH1L in the DFC during 8-16pcw of development. Clusters of genes are colored based on k-means clustering.

APPENDIX F:

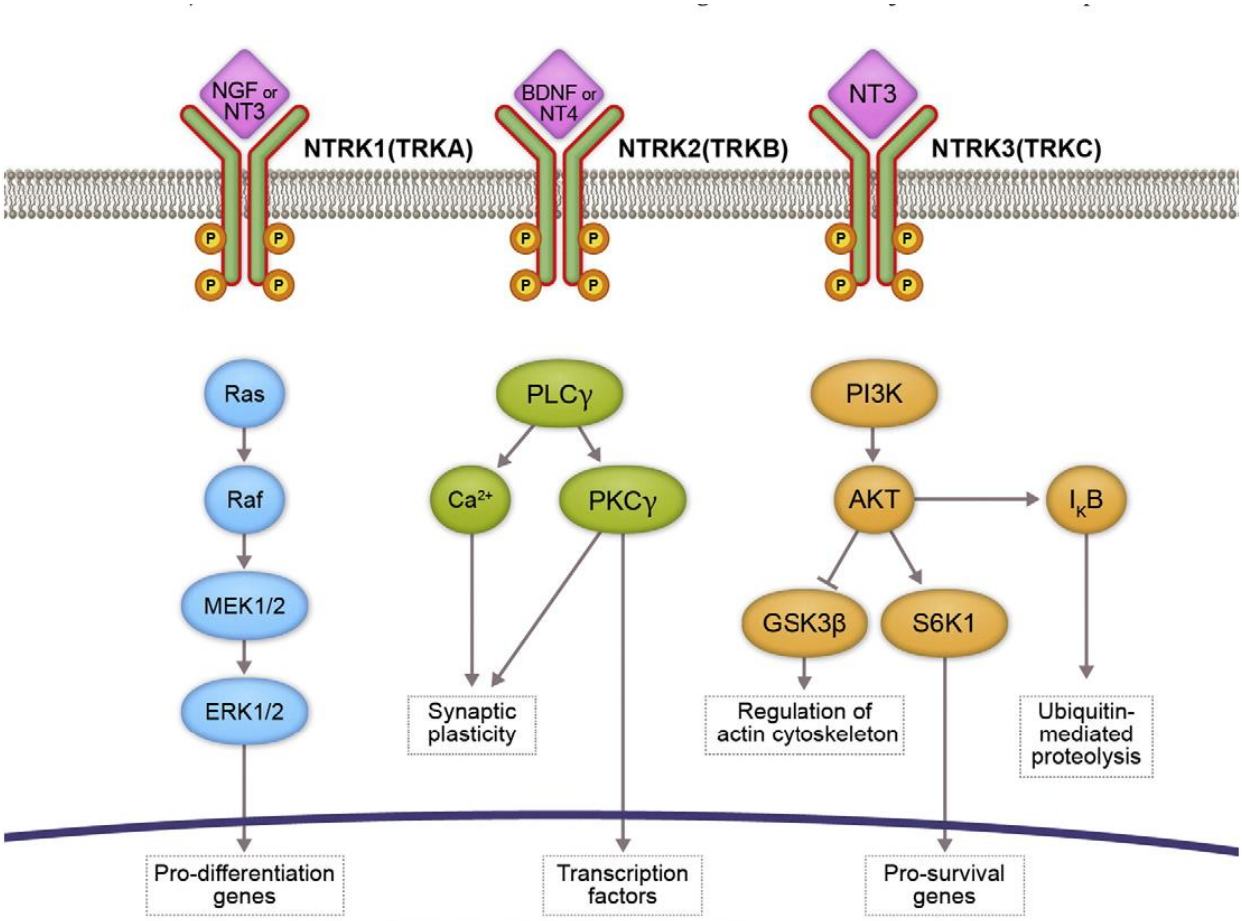


Figure 12. Neutrophin Signaling Pathway. Neurotrophin Signaling Pathway has three main receptors which activated different pathways which lead to different effect on the function of a cell. Trkb (encoded by *NTRK2*) is important to promoting synaptic plasticity, while Trkc triggers signaling cascades important for regulation of actin cytoskeleton and ubiquitin mediated proteolysis. This image was found in (Khotskaya et al., 2017)

APPENDIX G:

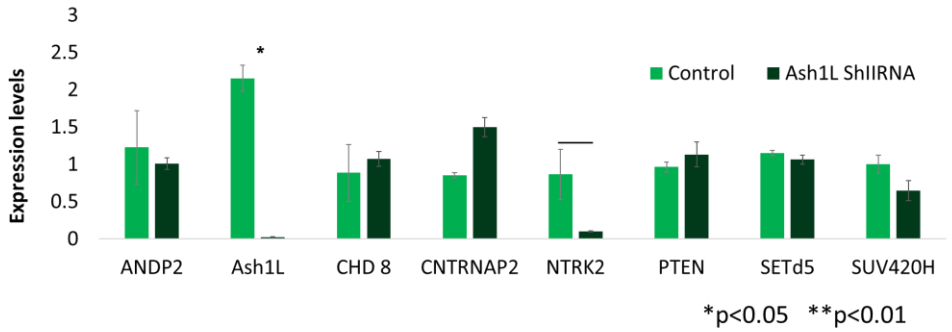


Figure 13. Knockdown of ASH1L in stem cell derived human neurons significantly reduces the expression of NTRK2 (gene encoding TrkB receptor). Human cortical neurons at day 35 of neuronal induction were nucleofected with ASH1L-shRNA-II. Cells were subjected to puromycin selection for 3 days after nucleofection. Histogram shows gene expression analyzed by QPCR for several chromatin regulatory factors previously associated with autism. We found the most significant change in gene in expression for ASH1L and for NTRK2 compared to untransfected human neurons (control). The average of three independent experiments is shown. These experiments were performed by Seonhye Chen.

APPENDIX H:

Functional enrichment tables for the positive co-expression network as calculated using STRINGdb. This data was used to created Figure 7 and its figure legend.

Cellular Component	false discovery rate
nucleus	0.0052
nucleoplasm	0.0052
postsynaptic density	0.0052
axon	0.0052
dendrite	0.0052
nuclear lumen	0.0052
somatodendritic compartment	0.0052
cell projection	0.0052
neuron projection	0.0052

Function	false discovery rate
chromatin binding	0.0193
histone acetyltransferase activity	0.0193
catalytic activity, acting on a protein	0.0193
histone-lysine N-methyltransferase activity	0.0196
ubiquitin-protein transferase activity	0.0255
steroid hormone receptor binding	0.0288
tubulin binding	0.031
insulin binding	0.031
transferase activity	0.0338
nuclear receptor binding	0.0338
histone demethylase activity	0.0338
phosphatidylinositol binding	0.0338
ion binding	0.0338
phosphatidylinositol 3-kinase binding	0.0338
transcription coregulator activity	0.0339
transcription coactivator activity	0.0339
cytoskeletal protein binding	0.0339
histone methyltransferase activity (H3-K4 specific)	0.0339
microtubule binding	0.0397
insulin-like growth factor I binding	0.0397

Cellular Process	false discovery rate
cellular protein modification process	0.0049
learning	0.0049
cellular component organization	0.0049
histone modification	0.0049
macromolecule modification	0.0049
cellular protein metabolic process	0.0118
nervous system development	0.0184
learning or memory	0.0184
regulation of cell differentiation	0.0184
regulation of multicellular organismal development	0.0184
peptidyl-lysine modification	0.0196
cellular macromolecule metabolic process	0.0196
regulation of macromolecule metabolic process	0.0203
chromatin organization	0.0212
microtubule cytoskeleton organization	0.0218
organelle organization	0.0218
regulation of metabolic process	0.0218
cytoskeleton organization	0.0226
regulation of developmental process	0.0226
regulation of cellular metabolic process	0.0265

REACTOME	false discovery rate
Chromatin modifying enzymes	1.51E-05
RORA activates gene expression	8.42E-05
Regulation of RAS by GAPs	9.85E-05
Regulation of Hypoxia-inducible Factor (HIF) by oxygen	0.00014
BMAL1:CLOCK,NPAS2 activates circadian gene expression	0.00017
Transcriptional regulation by RUNX1	0.00017
Signaling by NOTCH4	0.00017
Circadian Clock	0.00066
TCF dependent signaling in response to WNT	0.0008
PKMTs methylate histone lysines	0.0014
MAPK family signaling cascades	0.0014
Vif-mediated degradation of APOBEC3G	0.0016
RAF/MAP kinase cascade	0.0016
Transcriptional regulation by RUNX3	0.0016

ACKNOWLEDGMENTS

I would like to thank various people who contributed to this project. I would have not been able to do this research without the hard work of Seonhye Chen, and her in-vitro ASH1L studies. Furthermore, I appreciate all the help and support I received from members of the Lizarraga lab these past two years; Mara Cowen, Mary-Kate Lawlor, Dr. Evelyn Chukwurah, Janay Clytus, Pankaj Ghate, and Gustavo Martinez. Special thanks should be given to Dr. Lizarraga, my thesis director and research mentor, for her professional guidance and valuable support through this entire process. Her feedback and research insight helped me to conduct these analyses. Thank you also to my second reader, Dr. Katie Kathrein, who provided me feedback and assistance on my thesis. This project was also conducted with the assistance of the USC South Carolina Honors College Scientific Undergraduate Funding Grant and USC Office of Undergraduate Research Magellan Scholar Grant. Finally, I wanted to extend my gratitude to my parents, Laura and David Bagnell, for allowing me the opportunity to attend the University of South Carolina and be a member of the South Carolina Honors College.