Functional Genomics Studies of Human Brain Development and Implications for Autism Spectrum Disorder

Mark Nicholas Ziats

Robinson College

University of Cambridge

This dissertation is submitted for the degree of

Doctor of Philosophy

November 2013

For Mom, Dad, Ann, and Catherine

Declaration

This dissertation is the result of my own work and contains nothing that is the outcome of work done in collaboration, except as specifically described in the Appendix.

The length of this thesis does not exceed the 60,000 word limit and it has been typeset using the specifications set by the Biology Degree Committee.

This dissertation is not similar to any other that I submitted for a degree, diploma, or other qualification at any other University. Furthermore, I state that no part of this dissertation has been, or is concurrently being, submitted for any degree, diploma, or other qualification.

Mark N. Ziats

November 2013

Summary

Human neurodevelopment requires the coordinated expression of thousands of genes, exquisitely regulated in both spatial and temporal dimensions, to achieve the proper specialization and inter-connectivity of brain regions. Consequently, the dysregulation of complex gene networks in the developing brain is believed to underlie many neurodevelopmental disorders, such as autism spectrum disorders (ASD). Autism has a significant genetic etiology, but there are hundreds of genes implicated, and their functions are heterogeneous and complex. Therefore, an understanding of shared molecular and cellular pathways underlying the development ASD has remained elusive, hampering attempts to develop common diagnostic biomarkers or treatments for this disorder.

I hypothesized that analyzing functional genomics relationships among ASD candidate genes during *normal* human brain development would provide insight into common cellular and molecular pathways that are affected in autistic individuals, and may help elucidate how hundreds of diverse genes can all be linked to a single clinical phenotype. This thesis describes a coordinated set of bioinformatics experiments that first (i) assessed for gene expression and co-expression properties among ASD candidates and other non-coding RNAs during normal human brain development to discover potential shared mechanisms; and then (ii) directly assessed for changes in these pathways in autistic post-mortem brain tissue.

The results demonstrated that when examined in the context of normal human brain gene expression during early development, autism candidate genes appear to be strongly related to the neurodevelopmental pathways of synaptogenesis, mitochondrial function, glial cytokine signaling, and transcription/translation regulation. Furthermore, the known sex bias in ASD prevalence appeared to relate to differences in gene expression between the developing brains of males and females. Follow up studies in autistic brain tissue confirmed that changes in mitochondrial gene expression networks, glial pathways, and gene expression regulatory mechanisms are all altered in the brains of autistic individuals. Together, these results show that the heterogeneous set of autism candidate genes are related to each other through shared transcriptional networks that funnel into common molecular mechanisms, and that these mechanisms are aberrant in autistic brains.

Acknowledgements

This work would not have been possible without the tremendous amount of support I received from many individuals and organizations.

First, I would like to thank my graduate program and my two thesis advisors. The National Institutes of Health-University of Cambridge Biomedical Scholars Program has been a tremendous training opportunity for me. I have had great support from both NIH and the University of Cambridge throughout my studies, and I am incredibly grateful to have been given this opportunity. I am especially indebted to Dr. Azim Surani at the University of Cambridge for his guidance and flexibility with me as this work developed away from our original plans, and for his continued mentoring and guidance. To Dr. Owen Rennert at the National Institute of Child Health and Human Development, I will be forever grateful for all that he has done for me.

Throughout my training I have been funded by or received awards from a number of organizations that have allowed me to pursue this work. The National Institute of Child Health and Human Development at NIH, the NIH-Oxford/Cambridge Biomedical Scholars Program, the NIH MD/PhD Global Doctoral Partnership program, Baylor College of Medicine Medical Scientist Training Program, the Allen Institute for Brain Science, the American Academy of Neurology, Robinson College, and the University of Cambridge have all provided me with generous support.

I would like to thank others with whom I have worked, learned from, and have been given help during my studies, including members of Dr. Surani's and Dr. Rennert's laboratories, the administrators at NICHD, the NIH-Oxford/Cambridge Program, Robinson College, and Baylor College of Medicine, and our collaborators at TU Delft.

Most importantly, I would like to thank my parents, Nicholas and Lucille Ziats, for their unwavering support, guidance, and inspiration.

Table of Contents

Publications Resulting From This Work

Mahfouz A,[*] Ziats MN,[*] Rennert OM, Lelieveldt BP, Reinders MJ. Co-expression Network Analysis of the Developing Human Brain Transcriptome Reveals Shared Pathways among Autism Candidate Genes. *Revision Submitted*. [*]equal contribution

Edmonson C[*], Ziats MN[*], Rennert OM. Altered glial marker expression in autistic post-mortem pre-frontal cortex and cerebellum *Mol Autism. 2014;5(1):3.* [*]equal contribution

Ziats MN, Rennert OM. The cerebellum in autism: pathogenic or an anatomical beacon? *Cerebellum*. 2013 Oct;12(5):776-7.

Ziats MN, Rennert OM. Identification of differentially expressed microRNAs across the developing human brain. *Mol Psychiatry*. 2013 Aug 6. [Epub ahead of print]

Ziats MN, Rennert OM. Sex-biased gene expression in the developing brain: implications for autism spectrum disorders. *Mol Autism*. 2013 May 7;4(1):10.

Ziats MN, Rennert OM. Aberrant expression of long noncoding RNAs in autistic brain. *J Mol Neurosci*. 2013 Mar;49(3):589-93.

Ziats MN, Rennert OM. Expression profiling of autism candidate genes during human brain development implicates central immune signaling pathways. *PLoS One*. 2011;6(9):e24691.

List of Abbreviations

Amyg, Amy	amygdala
ASD	autism spectrum disorder
ATP	adenosine triphosphate
BA	Broadman's area
CDC	Center for Disease Control and Prevention
Cere	cerebellum
CGH	comparative genomic hybridization
CNV	copy number variation
DSM	Diagnostic and Statistical Manual of Mental Disorders
DLPC, DFC	dorsolateral prefrontal cortex
ECM	extra cellular matrix
endo-siRNA	endogenous small interfering RNA
ETC	electron transport chain
FAD	falvin adenonucleotide
FC	fold change
FDR	false discovery rate
GABA	gamma aminobutyric acid
GO	gene ontology
Hipp, Hip	hippocampus
IHC	immunohistochemistry
ILTC	inferior lateral temporal cortex
IPA	ingeunuity pathway analysis
lncRNA	long non-coding RNA
miRNA	micro RNA
MPC, MFC	medial prefrontal cortex
mRNA	messenger RNA
mtDNA	mitochondrial DNA
ncRNA	non-coding RNA
NCX	neocortex
OPC, OFC	orbital prefrontal cortex
PCW	post-conception weeks
PET	positron emission tomography
piRNA	Piwi-interacting RNA
PMC	primary motor cortex
PMI	post mortem interval
PSTC	posterior superior temporal cortex
qRT-PCR	quantitative, real time, polymerase chain reaction
RISC	RNA-induced silencing complex
RNA-seq	RNA sequencing
RNAi	RNA interference
RPKM	reads per kilobase of exon model per million mapped reads
rRNA	ribosomal RNA
Stri, Stry	striatum
tRNA	transfer RNA
VLPC, VFC	ventrolateral prefrontal cortex

List of Figures and Tables

Chapter 1. Introduction

The human brain is exceedingly complex, and the mechanisms underlying its development and functioning are only beginning to be understood. As a consequence, the etiology underlying disorders of neurodevelopment, such as autism spectrum disorder (ASD), remains unclear. However, recent efforts have demonstrated that a significant component of the etiology of ASD is genetic, but the exact genetic and molecular mechanisms underlying the disorder have proven exceedingly difficult to define. This is largely because the genetics guiding normal human brain development are still not clear, and studies of genes implicated in ASD mostly have not considered the unique function of these genes in the specific context of human neurodevelopment.

Therefore, in order to more comprehensively understand the genetic etiology of autism spectrum disorders, it is critical to understand the function and regulation of autism candidate genes during normal human brain development. To do so requires integrating what has previously been discovered about the genetics of autism with what is known about normal human neurodevelopmental genomics, as is reviewed in this chapter. This information can then be applied to analyze new datasets of human gene expression during neurodevelopment. Furthermore, it is critical that parallel efforts are made to determine what molecular genetic mechanisms are aberrant in autistic brain tissue directly, as the overlap of these two lines of evidence may help focus on the main inherited genetic etiologies of autism.

Finally, as is shown in this chapter, there is clearly a need to more comprehensively understand the expression and regulation of ASD candidate genes during normal human brain development, and to determine if the molecular mechanisms they implicate are abnormal in autistic brains. This chapter concludes by describing these major unanswered questions in the field of ASD genomics, which are then addressed in the studies described in Chapter 2 and Chapter 3 of this thesis.

1.1 Autism Spectrum Disorders

The autism spectrum disorders are a heterogeneous set of neurodevelopmental syndromes defined by impairments in verbal and non-verbal communication, restricted social interaction, and the presence of stereotyped patterns of behavior. The prevalence of ASD is rising, and the diagnostic criteria and clinical perspectives on the disorder continue to evolve in parallel. Although the majority of individuals with ASD will not have an identifiable cause, almost 25% of cases have genetic lesions (Huguet et al. 2013). The rapidly improving ability to identify genetic mutations because of advances in next generation DNA sequencing, coupled with previous epidemiological studies demonstrating high heritability of ASD, have led to many recent attempts to identify causative genetic mutations underlying the ASD phenotype. However, although hundreds of mutations have been identified to date, they either are rare variants affecting only a handful of ASD patients, or are common variants in the general population with only a small effect size on the risk for ASD (Devlin and Scherer 2012). Furthermore, the genes implicated thus far are heterogeneous in their structure and function, hampering attempts to understand shared molecular mechanisms among all ASD patients; an understanding that is crucial for the development of targeted diagnostics and therapies. Therefore, a major unmet need in the field of ASD research—and the main goal of this work—is to integrate the heterogeneous genetic findings in ASD in order to begin to understand common molecular and cellular pathways that are perturbed in patients with the disorder.

Clinical Phenotype and Incidence

Autism was first described seventy years ago by the American child psychiatrist Leo Kanner (Kanner 1943). While originally reported by Kanner as an isolated syndrome with the core components being 'obsessive insistence on the preservation of sameness' and 'autistic aloneness,' autism was considered mainly as a childhood form of schizophrenia for more than thirty years (Eisenberg and Kanner 1955). Autism was first formally recognized as its own clinical diagnostic entity in 1980 (American Psychiatric Association, DSM-III, 1980), defined as encompassing three essential features: impairment in communication, lack of interest in other people, and 'bizarre' behaviors. Since that time, the criteria required to obtain a diagnosis of ASD, and its relation to other similar disorders such as Asperger's and Rett syndrome, have changed multiple times—reflecting both the clinical heterogeneity of the disorder and the poor understanding of its underlying pathophysiology.

The most recent definition of ASD recognizes abnormalities in two clinical domains: 'social and communication defects' and 'fixed interests and repetitive behaviors' (American Psychiatric Association, DSM-V, 2013). All of the following three symptoms describing persistent deficits in social interaction and communication must be present for a diagnosis of ASD to be made: (i) problems reciprocating social or emotional interaction, inability to initiate an interaction, and problems with shared attention or sharing of emotions and interests with others; (ii) problems maintaining relationships and problems adjusting to different social expectations; and (iii) nonverbal communication problems such as abnormal eye contact, posture, facial expressions, tone of voice and gestures, as well as an inability to understand these. Additionally, these interaction/ communication deficits cannot be better accounted for by general developmental delay. Two of the four following symptoms related to restricted and repetitive behavior must also be present: (i) stereotyped or repetitive speech, motor movements, or use of objects; (ii) excessive adherence to routines, ritualized patters of verbal or nonverbal behavior, or excessive resistance to change; (iii) highly restricted interests that are abnormal in intensity or focus; and (iv) hyper- or hypo-reactivity to sensory input or unusual interest in sensory aspects of the environment.

Furthermore, the severity of each symptom must be defined based on the level of support required for that symptom, in an attempt to more thoroughly capture the 'spectrum' nature of the disease. In all cases, symptoms must have been present in early childhood (even if initially unrecognized); although they may not become fully manifest until later in life when social demands exceed capacities. The symptoms must impair everyday functioning, and cannot be better described by another Diagnostic and Statistical Manual of Mental Disorders-5th Edition (DSM-5) diagnosis.

Autism spectrum disorders are one of the most common neurodevelopmental problems affecting children in the Western world. The most recent estimates have shown that ASD affects between 1 in 88 children (Centers for Disease Control and Prevention (CDC) 2012) and perhaps as many as 1 in 50 (Blumberg et al. 2013) depending on the methodology employed. This represents a staggering 1.17%-2% of all children. Boys are at least four times more likely to receive a diagnosis of ASD as compared to girls (CDC 2012), and this ratio increases significantly when only mildly affected children are considered (Gilberg et al. 2006). Furthermore, prevalence estimates have been increasing substantially in recent

years—form 1 in 150 children in the year 2000—although it is unclear to what extent this represents a true biological increase or is a result of expanding diagnostic criteria and better clinical recognition of the disorder (Fombonne 2009).

The costs associated with autism are similarly great. Economically, direct and indirect medical costs are estimated to be nearly £2 million pounds person over his or her lifetime, or more than £21 billion pounds per year for all people with ASD (Moldin and Rubenstein 2006). Perhaps more importantly, the emotional toll placed on parents and caregivers of children with autism is immense, unrelenting, and has a serious impact on family relationships (Rao and Beidel 2009), marriages (Benson and Kersh 2011), and couples' future reproductive decisions (Selkirk et al. 2009).

Consequently, it is of upmost urgency to patients with ASD, their caregivers, and society at large that the underlying cause(s) of the disorder are understood. Doing so will enable the development of better, more specific diagnostic tests that can recognize ASD earlier in life, which has been shown to be important to improve long-term outcomes (Howlin et al. 2009), provide parents with an explanation for their child's symptoms, and may eventually enable the development of targeted therapeutics. Moreover, by understanding the mechanisms that lead to the altered higher cognitive functioning seen in patients with ASD, the field of human neuroscience as a whole can be advanced, as it will provide insights into the genetic and molecular basis of higher cognition.

However, the underlying pathophysiology of autism spectrum disorders has long been a mystery. Various hypothesis ranging from psychosocial abnormalities to environmental insults have been purported, yet it was not until twin and sibling epidemiological studies were undertaken in the 1980s that the strong heritability of ASD began to be realized. Subsequently, a large amount of work has firmly established a significant genetic component to ASD's etiology.

Genetic Etiology

Evidence for a strong heritable risk of ASD was initially described in twin and sibling epidemiological studies of autism (Folstein and Rutter 1977), and has since been firmly established through multiple genetic approaches (Berg and Geschwind 2012; Geschwind

2011). It was first recognized that the risk of having a second child with autism was higher in families that already had one child with ASD than was the risk of having a child with ASD in the general population. Originally this recurrence risk was estimated to be 5% (compared with approximately 1% in the general population), although more recent estimates suggest it may be as high as 20% (Ozonoff et al. 2011). Following these initial observations, the first twin studies in ASD demonstrated a concordance rate approaching 90% in monozygotic twins and 10% in dizygotic twins (Bailey et al. 1995; Steffenburg et al. 1989; Smalley et al. 1988; Ritvo et al. 1989). Subsequently, larger studies have shown the dizygotic concordance rate to be greater than 20% (Hallmayer et al. 2011).

These observations, coupled with the identification of causative genetic mutations in monogenic disorders with autism as a component, such as Fragile X and Rett syndromes (Amir et al. 1999; Pieretti et al. 1991), led to an ongoing effort to identify genetic causes of 'idiopathic' ASD using a number of genomic approaches. As the technology behind these approaches has improved, the ability to identify mutations with incredibly sensitivity and genomic resolution has resulted in over 200 genetic loci implicated in ASD to date (Freitag 2007; Anney et al. 2010; Holt et al. 2010). However, as more genes and loci are identified, it is becoming increasingly clear that the genomic architecture of ASD is incredibly heterogeneous and complex, necessitating a functional integration in order to decipher common molecular mechanisms underlying ASD.

Genomic Architecture of ASD

The identification of genomic loci and individual genes disrupted in patients with ASD has progressed in tandem with the rapid development of sensitive genomic tools. Initially, microscopically-visible chromosomal aberrations were observed in patients with ASD who received karyotyping analysis. These case reports were variable, but a number of loci were repeatedly implicated, including 7q11, 15q11-13, and 22q11.2 (Vorstman et al. 2006)—regions already associated with syndromes that had autistic symptoms as a component, and known to contain a number of critical neurodevelopmental genes and some of the first identified functional non-coding RNAs (Mabb et al. 2011; Szafranski et al. 2010).

Subsequently, the development of microarray technology such as comparative genomic hybridization (CGH, Alkan et al. 2011), allowed the unbiased assessment of copy-number

variation (CNV) across the whole genome at a resolution of as low as 100 kilobases. The first of these analysis indicated that individuals with ASD had 10-20 times the number of CNVs as controls (Jacquemont et al. 2006; Sebat et al. 2007). Numerous studies have since used CGH or similar approaches to follow up and improve upon these initial reports with larger and more homogenous patient populations, with thousands of individuals with ASD having been analyzed to date (Christian et al. 2008; Cooper et al. 2011; Gilman et al. 2011; Glessner et al. 2009; Itsara et al. 2010; Marshall et al. 2008; Pinto et al. 2010; Sanders et al. 2011; Szatmari et al. 2007; Huguet et al. 2013). These studies have consistently demonstrated that individuals with ASD have more CNVs than non-related controls. Furthermore, studies employing a family cohort model have been able to compare individuals with ASD to their parents and unaffected siblings, which has revealed that *de novo* mutations in particular are more frequent in children with ASD.

Functionally, it was also shown that larger CNVs (i.e. affecting more genes) are associated with decreased cognition (Girirajan et al. 2012), and that females with ASD tend to have larger CNVs than males with ASD (Itsara et al. 2010; Sanders et al. 2011), suggesting they are somehow more 'genetically tolerant' of these disruptions. Moreover, some of the identified loci result in nearly opposite phenotypes depending on whether they are duplicated or deleted (Jacquemont et al. 2011). Taken together, these functional CNV findings suggest that identification of the genes in these regions is not sufficient to understand the mechanisms underlying autism, as it appears that a finely-regulated dosage of each gene is necessary to avoid neurodevelopmental problems such as ASD.

Despite the progress made with CGH arrays, the findings from these studies only identified CNVs in 5-15% of individuals with ASD, suggesting that other types of mutations must be operant in ASD as well. However, investigations at higher genomic resolution were traditionally limited to specific candidate genes until the recent advent of next-generation sequencing technologies. Since then, seven exome sequencing studies have been completed in ASD, encompassing more than 1,000 affected individuals (Klei et al. 2012; Kong et al. 2012; Neale et al. 2012; O'Roak et al. 2011; O'Roak et al. 2012a; O'Roak et al. 2012b; Sanders et al. 2012). In addition to identifying a number of high-confidence ASD candidate genes (likely representing 5-10% of ASD cases), these studies provided two other more broad insights into the functional genomics of ASD that are particularly motivational to the work

describe in this thesis. First, with the exception of a few identified genes, there was very little replication of ASD candidates among the studies. This suggest that common variants (i.e. those accounting for greater than 1% of cases) are unlikely to play a major role in ASD pathogenesis, confirming similar prior findings from genome-wide association studies and linkage analysis that failed to identified many replicable loci (Szatmari et al. 2007; Wang et al. 2009; Weiss et al. 2009). Consequently, it has been predicted that up to 1,000 genes may be found to be associated with ASD based on statistical modeling (Sanders et al. 2012; Iossifov et al. 2011). Therefore, understanding how such a large and varied number of genes can all be associated with one common clinical phenotype is a major challenge in the field, and one of the focuses of this work. Secondly, a meta-analysis of these studies at the group level showed that the average rate of mutations in individuals with ASD was not significantly different than controls—or even unaffected siblings—unless the analysis was restricted to genes that are known to be expressed during human brain development (Sanders et al. 2012). This highlights the tissue- and human-specific nature of gene function, which underscores the importance of understanding the function of ASD candidate genes in the context of human brain development specifically.

Lastly, there is a growing appreciation that the presence of multiple mutations and/or inherited protective or risk alleles—each at different loci within one individual—may interact with each other to result in the emergent ASD phenotype, and that this may help explain the complex and heterogeneous nature of ASD genomics. For instance, a number of studies have described individuals with ASD who have more than one deleterious mutation (Girirajan et al. 2010; Girirajan et al. 2012; Leblond et al. 2012), and the presence of more than one mutation correlates with an increased risk of developmental delay (Girirajan et al. 2012). Other studies have suggested certain inherited variants may be protective against other ASD-causing mutations, especially in females (Robinson et al. 2013). While the identification of multiple mutations within individuals is becoming a relatively straightforward task, the challenge of understanding how combinations of susceptibility genes interact during human brain development to cause disease (epistasis) has only begun to be explored, and is a major theme of this thesis.

In summary, much work has attempted to elucidate the molecular genetics underlying autism, with many linkage, genome-wide association, copy number variation, and whole-exome

sequencing projects having implicated hundreds of genes in ASD. Yet understanding how this diverse set of genes relates to the underlying molecular mechanisms and subsequent neuropathology of ASD is still unclear. The genetic etiology of ASD is variable, complex, and likely involves gene-gene, gene-environment, and epigenetic interactions, as is evidenced by the incomplete concordance among monozygotic twins, and the considerable variability within pedigrees (Piven et al. 1997; Ronald et al. 2006). This genetic heterogeneity reflects the overlying broad clinical presentation of ASD, and is captured by the 'spectrum' designation of the disorder. Furthermore, ASD shares considerable clinical and genetic overlap with other neuropsychiatric disorders such as schizophrenia and mental retardation (Mitchell 2011), and ASD patients have significantly increased neurologic co-morbidities like hypotonia, tics, and epilepsy (Levy 2009). In fact, many of the same gene mutations have been found to predispose to more than one of these neurodevelopmental disorders (Ching et al. 2010; Guilmatre et al. 2009). This body of evidence suggests that while identification of candidate genes in ASD is a critical first step toward understanding the genetic etiology of this disorder, a comprehensive, disorder-specific understanding of the molecular mechanisms cannot be realized until the functional genomics of ASD candidate genes are properly understood in the context of human brain development.

Cellular Etiology of ASD

Although autism currently lacks any unifying principles at the genetic and molecular levels, both human and animal studies have begun to demonstrate that disruption of synaptogenesis and improper connectivity of local and distant brain networks likely underlie the cellular pathophysiology responsible for the broad ASD phenotype (Geschwind and Levitt 2007; Zoghbi 2003). Multiple different brain regions have been implicated in both post-mortem and neuroimaging studies, notably the prefrontal and temporal cortices, and the cerebellum (Abrahams and Geschwind 2010). Histological analysis has revealed increased cell densities, changes in synaptic spine morphology, mini-columnar disorganization, and glial activation (Pickett and London 2005). Intriguingly, many of the genes known to be integral to these processes have been independently linked to autism in genetics studies. For instance, the Shank family of proteins, which interact with themselves and other transmembrane proteins at the post-synaptic density, are one of the main regulators of synaptic spine morphology (Sala et al. 2001). Multiple Shank family genes, notably Shank 1 and Shank 3, have been repeatedly implicated in ASD (Bourgeron 2009). Similarly, genes involved both in formation

and maintained of cortical mini-columns, such as the cadhedrin family of proteins (Redeis et al. 2012), and genes involved in glial activation, such as members of the Wnt/B-catenin pathway like DOCK1 and WNT2 (Yang 2012), have been independently implicated in ASD through genetic studies (Wang et al. 2010; Kalkman 2012).

Despite these observations, the underlying mechanism(s) responsible for this "disconnection" phenotype remains obscure, as a complex interplay between diverse cell types and functions modulate the developing network architecture in both a temporally and spatially regulated manner (Levitt 2003; Vogel et al. 2010; Bolton et al. 2009).

In particular, studies have shown that long-distance communication between disparate neocortical areas may be disrupted in ASD, causing delays in information processing within the brain that manifest as the communication, language, and social development problems seen in children with autism (Just et al. 2007). Additionally, parallel research has shown that neuronal micro-circuitry within brain areas may also be disrupted in ASD, and that this may result in local processing deficits within brain regions related to higher functioning, such as the prefrontal cortex (Rubenstein and Merzenich 2003). Underlying these circuit disruptions is a large body of evidence that has demonstrated decreased numbers of neurons (and their various subtypes) throughout the autistic brain by early childhood in post-mortem studies (Courchesne et al. 2007).

In addition to the body of evidence implicating aberrant local and long-distance synaptic dysfunction in ASD, many studies have demonstrated microglial and astrocyte dysfunction in ASD brains. For instance, post-mortem pathological studies of autistic brain using immunocytochemistry (IHC) and/or stereology have identified microglial activation patterns (Vargas et al. 2005; Morgan et al. 2010; Morgan et al. 2012), and have demonstrated increased microglial cell density in multiple brain regions (Morgan et al. 2010; Tetreault et al. 2012). Furthermore, positron emission tomography (PET) using a microglial-specific radiotracer also demonstrated microglial activation in multiple brain regions of autistic cases (Suzuki et al. 2013). Additionally, studies in a Rett syndrome mouse model, a single-gene deletion disorder with autism as a component, have also demonstrated cellular microglial abnormalities (Maezawa and Jin 2010), and a remarkable study demonstrated that autistic-like phenotypes can be partially reversed by replacing mutant microglia with their respective wild-type cells (Derecki et al. 2012).

Increased numbers of astrocytes, with altered cell size and branching patterns, have also been demonstrated in post-mortem autistic brains (Cao et al. 2012). Additionally, astrocyte-specific cell marker proteins are increased in multiple autistic brain regions (Laurence and Fatemi 2005; Fatemi et al. 2008). Similar to microglial studies in ASD mouse models, astrocytes have been shown to be abnormal in number of single-gene ASD models, including Rett (Maezawa et al. 2009; Yasui et al. 2013), Fragile X (Yaskaitis et al. 2010), and Tuberous Sclerosis (Uhlmann et al. 2002). In parallel to the aforementioned microglial study, it was also shown that replacing mutant astrocytes in Rett syndrome mice could correct some aspects of the phenotype (Lioy et al. 2011).

Overall, the cellular pathology in the brains of individuals with ASD is equally as complicated as the underlying genetics. While there is strong evidence to suggest that the autistic phenotype ultimately results from aberrant local and long-distance synaptic wiring, it remains unclear if the repeated observation of altered microglia and astrocytes are contributory to the phenotype or represent a reaction to synaptic pathology. However, previous functional genomics studies of ASD brain tissue (discussed in Chapter 1.3) have demonstrated altered immune and glial gene expression in autistic brains, suggesting that glial cell abnormalities may contribute to defects in synaptic wiring. The complex interplay between ASD genetics and glial cell abnormalities is explored further throughout this work.

Conclusion

In summary, autism spectrum disorders are common, and have considerable consequences for individuals with ASD, their families, and society at large. Because the underlying causes of ASD are not understood, specific diagnostic tests and therapeutic strategies are unavailable. The evolution of ASD's clinical definition is indicative of the heterogeneous and complex nature of the disorder. While ASD has recently been shown to have a significant genetic etiological component, the genes implicated are equally heterogeneous, hampering attempts to define common molecular mechanisms. In parallel, cellular studies have revealed that ASD likely ultimately results from disrupted synaptic function, but a large body of evidence has also implicated immune and glial abnormalities in autistic individuals. Therefore, studies that attempt to reconcile the heterogeneous and varied nature of ASD genomics, and the interplay between neurons and glial, are necessary to move the field forward toward a common understanding of the mechanisms underlying the development of ASD.

1.2 Functional Genomics of Human Brain Development

Transcription of the inherited DNA sequence into copies of messenger RNA (mRNA) is the most fundamental process by which the genome functions to guide development. Furthermore, encoded sequence information, inherited epigenetic marks, and environmental influences all converge at the level of mRNA gene expression to allow for cell type-specific, tissue-specific, spatial, and temporal patterns of expression. Thus, the transcriptome represents a complex interplay between inherited genomic structure, dynamic experiential demands, and external signals. This property makes transcriptome studies uniquely positioned to provide insight into complex genetic-epigenetic-environmental processes such as human brain development, and disorders with non-Mendelian genetic etiologies such as autism spectrum disorders.

As humans develop, an individual gene can be expressed in multiple ways depending on the particular developmental context; that is, the tissue, stage of development, and local or long-distance signaling mechanisms being received. Therefore, in order to understand how a gene may contribute to a developmental disorder, it is critical to assess its expression and function in the appropriate tissue and developmental time window. Human brain gene expression has been demonstrated to be particularly unique evolutionarily, compared to other human tissues, and in its complex regulatory processes, underscoring the need to understanding the functional genomics of genes implicated in autism spectrum disorders during human brain development.

A Brief Overview of Human Brain Development at the Cellular Level

The complex processes that lead to the fully formed human brain encompass a spectrum of mechanisms spanning genetic determinates to environmental and experimental influences. While the functional genomic mechanisms underlying human brain development remain poorly understood—motivating much of the work in this thesis—over the past several decades significant advances have been made to document the cellular and anatomical events that occur as the human brain develops and matures. It is therefore important to consider studies of gene expression in this context of cellular/anatomic brain developmental patterns.

Cellular human brain development is a protracted process that begins around the third post-conception week (pcw) and arguably extends nearly into adulthood (Stiles and Jernigan, 2010). Conventionally, human brain development is considered in gross stages within which major cellular and anatomic transitions occur (**Figure 1.2.1**, Insel 2010); namely the embryonic, fetal, early and late postnatal, adolescent, and adult periods.

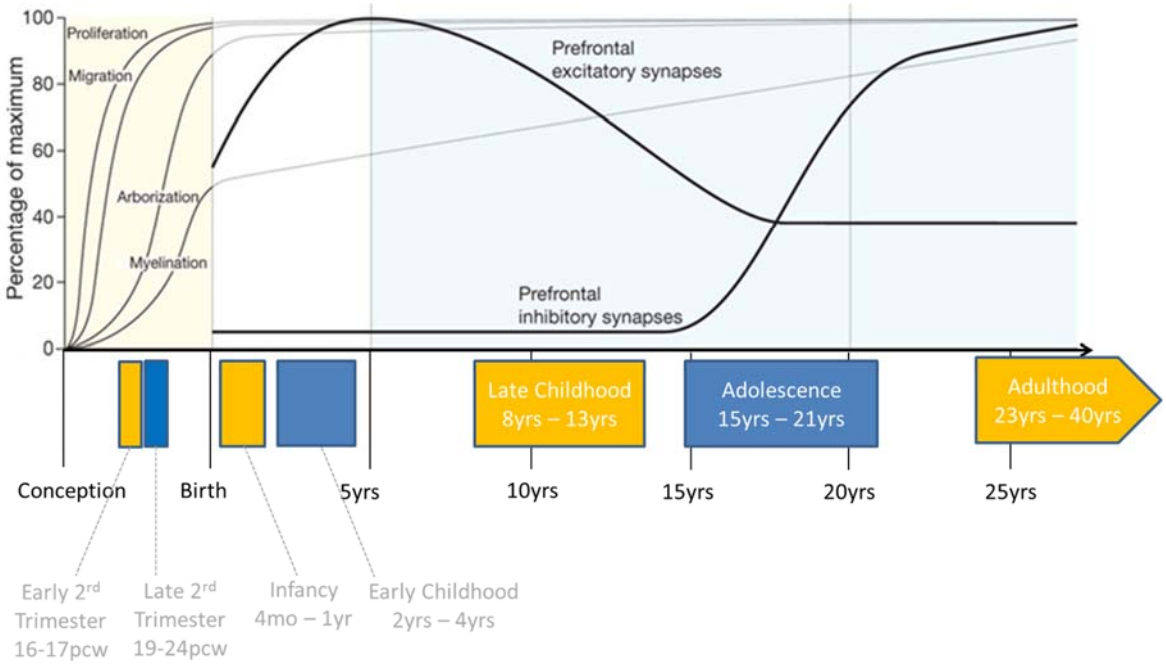


Figure 1.2.1. Trajectory of major brain developmental processes and their relationship to work performed in this thesis. Top: the trajectory of the major cellular phases of brain development are depicted as a percentage of their maximum abundance across human development. Bottom: Stages of brain development that are assessed in Chapter 2 of this thesis are depicted in colored blocks. Names and age ranges of the stages are given. Pcw; post-conceptional weeks. Figure adapted from: Insel TR. *Nature*. 2010;468(7321):187-93.

Beginning early in the embryonic period (defined as conception to eight pcw), the basic structures of the brain, spinal cord, and peripheral nervous system are established. The first major differentiating event of the embryonic period is gastrulation, during which the single-layered blastula forms a trilaminar structure containing the ectoderm, mesoderm, and endoderm. Gastrulation is completed by the third pcw, at which time some cells of the ectodermal layer differentiate into neural progenitors (Ozair et al. 2013). The first well-defined neural structure, the neural tube, begins forming during the third pcw and serves as the basis of the early developing central nervous system, within which reside populations of neural stem cells. From this basic tubular structure, more specific neural patterning of what

will become the major brain structures and compartments occurs through the creation and migration of neural cells from the stem cell proliferative zones. Through graded patterns of molecular signaling, neural progenitors migrate outward from proliferative zones and begin differentiation such that a primitive map of the brain is established by the end of the embryonic period. For instance, through comparative studies of other mammals it has been projected that the sensimotor regions of the neocortex (Bishop et al. 2002), the major compartments of the diencephalon and midbrain (Nakamura et al. 2005; Kiecker and Lumsden 2004), and the organization of the hindbrain and spinal column are all well established by the end of the embryonic period in humans (Lumsden and Keynes 1989; Gavalas et al. 2003).

Around the ninth pcw, the fetal period of development ensues and extends until birth, during which time there is rapid growth of the structures established during the embryonic period. Grossly, the brain develops its characteristic gyri and sulci during the fetal period (Chi et al. 1977), reflecting the underlying dramatic cellular changes occurring during this period. The majority of neuronal and glial proliferation occurs between the 9^{th} and 16^{th} pcw, with the peak period of migration of these cells to their appropriate region following closely thereafter (Volpe 2000). In fact, production of new neurons is largely finished by midgestation, except for the ongoing production of neurons in a few specialized areas (Bystron et al. 2008).

After their production in the proliferative regions, neurons migrate in an orderly manner to their final position in the developing brain. In the neocortex, the arriving cells establish a 6-layered structure, with the earlier migrating neurons forming the deeper layers and the later migrating neurons forming the more superficial layers (Cooper 2008). Their migration from the proliferative zone to their final position in the neocortex is helped by the guidance of radial glial cells, a population of stem cells that serve as a scaffold in the developing brain of all vertebrates (Borrell and Götz, 2014). Different layers of the neocortex contain different types of neurons as a result of both cell-intrinsic mechanisms operant in the progenitor cells from which they derive (Leone et al. 2008), and through soluble signaling cascades that direct progenitors toward a restricted mature neuronal type (Desai and McConnell 2000).

Of particular note in this migration process are a set of structures that appear only transiently during the fetal period to help guide the migration of progenitors to the developing

neocortical layers. The very first neurons to populate the developing neocortex form a primitive and transient structure termed the preplate, which is then split into two separate structures by arriving neurons—the marginal zone and the subplate (Molnár et al. 2006). The region between the marginal zone and subplate serves as a hub for new arriving neurons, and will eventually become layer 6 (the deepest) of the developing neocortex. Subsequently, all newly arriving cells will form progressively more superficial layers of the neocortex from this base structure. Intriguingly, both the marginal zone and subplate disappear by the end of the fetal period, yet they have been shown to highly express some of the genes most significantly linked to neurodevelopmental disorders such as autism and schizophrenia, such as Reelin (Bielle et al. 2005; Hoerder-Suabedissen et al. 2013). Consequently, an important caveat to post-mortem tissue research, both cellular and genetic, is the possibility of omitting the contribution of these transient structures to the proper formation and potential abnormalities in neocortical patterning.

Once the migrating neural cells have reached their destination, they begin to be incorporated into newly developing neural networks through a dynamic process of synaptogenesis and pruning that continues late into adolescence. Young neurons initially develop processes (dendrites and axons) that allow them to form synapses with other neurons both locally and long-distance. The growth cone of an axon is able to sample the neuron's environment for both chemical and electrical signals that guide its wiring to other neurons to create a new synapse (Brown et al. 2001). Initial patterns of connectivity in the fetal and early postnatal brain are characterized by exuberant synaptic connections that will later be pruned away to leave only the connections indicated through postnatal experience (Stiles and Jernigan, 2010). This process of network refinement occurs through both synaptic rewiring and neuronal apoptosis, with rates of apoptosis as high as 70% of cells in some regions of the cortex (Rabinoqicz et al. 1996). Physiological neuronal apoptosis in development occurs both as the result of intrinsic neuronal cell death mechanisms mainly responding to the absence of local neurotrophic factors (Huang and Reichardt, 2001), and also through glial-initiated mechanisms which have recently become more widely recognized (Kettenmann et al. 2013), which are of particular relevance to much of the work presented in Chapters 2 and 3 of this thesis. This synaptic and network refinement continues through early adulthood, largely in response to interaction with one's environment (Huttenlocher, 1987, Paus et al. 2001).

By the end of fetal development, all major adult brain structures are present, major connections between them are established, and the brain is poised for the rapid and dynamic growth that occurs in the first few years of life. The brain develops rapidly in the first few years after birth, reaching almost adult volume by age six (Lenroot and Giedd, 2006). While the production and migration of neurons are mainly prenatal events (with the notable exception of subventricular zone), glial progenitors have been shown to proliferate and differentiate throughout childhood (Cayre et al. 2009), like helping to sculpt the developing synaptic networks.

One main function of these proliferating glial cells during the early and late postnatal periods is to accomplish the extensive amount of axon myelination that occurs during this time. Increased myelination of axons allows for increased growth of axon diameter, and ultimately enables faster and long-distance neuronal connections (Zalc et al. 2008). Robust increases in myelination have been reported across the brain from ages 5 – 12 years, with a varying rate of fiber tract myelination in various brain regions (Lebel et al. 2008; Lebel and Beaulieu, 2009).

While the early postnatal period is characterized anatomically by an over-abundance of synaptic connections ("overconnectivity") between neurons, these connections are gradually pruned back over the course of development by competitive experiential processes. In fact, this particular pruning mechanism is hypothesized to be one of the main altered mechanisms in neurodevelopmental disorders like autism (Just et al. 2004). Modern neuroimaging techniques such as diffusor tensor imaging (DTI) and functional magnetic resonance imaging (fMRI) have made significant recent advances in linking brain structural changes to functional and behavioral development, with longitudinal neuroimaging studies having demonstrated changes in grey matter density throughout the neocortex into the mid-twenties, with the prefrontal cortex being the last to mature (Paus et al. 2008). While much of the organization of the postnatal brain is genetically determined (Stiles and Jernigan, 2010), it has been clearly demonstrated that this intrinsic development remains extremely malleable to experience-dependent processes (Hubel and Wiesel, 1977; Markham and Greenough, 2004).

Moreover, epigenetic mechanisms that ultimately converge to influence gene expression have been shown to be one of the main mediators between environmental experiences and developmental synaptic plasticity (Fagiolini et al. 2009). For instance, studies in mice have

shown that environmental enrichment results in increased chromatin remodeling that modifies gene expression patterns in the hippocampus, resulting in improved spatial memory (Fischer 2007). Alternatively, an increase in methylation of the *BDNF* promoter and consequent decrease in BDNF mRNA in the prefrontal cortex was found in association with exposure to periods of abusive maternal care, and these effects are perpetuated to the F1 generation suggesting a role for transgenerational effects (Champagne 2008). Yet while studies of model organisms are beginning to demonstrate that gene expression represents a critical nexus of experience dependent plasticity, human studies of neurodevelopmental disorders in which this process may go awry are limited, and the general landscape of gene expression in the developing human brain as relates to neurodevelopmental disorders like autism is largely unexplored.

In summary, great progress in understanding the anatomical and cellular trends underlying human brain development have been made over the past few decades. We have come to appreciate though various approaches that human neurodevelopment is a dynamic and protracted process, characterized by an initial period of neurogenesis leading to the formation of the basic CNS framework in early embryonic development. This is following by substantial cellular proliferation, migration, and differentiation in the fetal period that establishes the main areas and pathways of the brain by birth. The early postnatal period is a time of rapid growth through glial proliferation, myelination, and organization of developing neural networks. Importantly, this process is very malleable particularly with regard to environmental and experiential events. Precise refinement of these developing neural networks occurs throughout adolescence and into early adulthood.

While the cellular events that lead to the initial formation and subsequent refinement of human neuroanatomy are fairly well defined, the underlying molecular and genetic determinates that in part encode for these events are much well-less understood. Recent efforts to profile genome-wide expression patterns in post mortem human brains across development have begun to expose the uniqueness of human brain functional genomics, as is discussed in the next section.

Human Brain Gene Expression

Compared to other species, human brains express mRNA transcripts at much higher levels and with much greater complexity. For instance, comparisons of human brain gene expression with both mouse (Enard et al. 2002; Lockhart and Barlow 2001) and primates (Caceres et al. 2003; Khaitovich et al. 2004) has demonstrated that most of the differentially expressed genes between the species are up-regulated in humans, but this phenomena is not apparent in other tissues. Additionally, the human brain expresses ~86% of all genes encoded in the human genome at some point during development (Kang et al. 2011), which is greater than any other individual tissue type. It is hypothesized that this increased level of gene expression in the human brain is at least partially responsible for the higher level of neuronal activity and overall cognitive function in humans.

Within humans specifically, the brain also displays a distinct gene expression profile from other tissues. Using both array (de la Grange et al. 2010) and sequencing-based techniques (Ramskold et al. 2009), the brain has been shown to have higher expression levels and greater transcriptome complexity than other human tissue and cell types. In particular, human brain gene expression displays a high level of alternative splicing and a unique diversity of non-coding RNA types expressed. For example, studies have demonstrated that the human brain transcriptome has an unusually high level of alternatively spliced transcripts compared to other tissues (Yeo et al. 2004; Wang et al. 2008; Mortazavi et al. 2008), and the set of isoforms produced in brain differs considerably from other tissue types (Yeo et al. 2004; de la Grange et al 2010). In addition to increased numbers and types of spliced mRNAs, the human brain transcriptome also displays a uniquely high abundance of transcribed non-coding RNAs (ncRNAs). In fact, the brain displays the greatest abundance of transcribed ncRNAs among all tissues studied thus far (Qureshi and Mehler 2012). Both short ncRNAs, such as microRNAs (miRNAs) and piwi-interacting RNAs (piRNA), and long non-coding RNAs (lncRNAs) are highly enriched in the brain (Chodroff et al. 2010; Kuss and Chen 2008; Ponjavic et al. 2009; Schonrock et al 2010; St. Laurent et al. 2009). As ncRNAs are becoming increasingly recognized as important regulatory elements in genome processing during neurodevelopment and in the pathogenesis of neurodevelopmental disorders (Qureshi and Mehler 2011), their abundance in the brain further highlights the uniqueness of neurodevelopmental functional genomics (discussed further below).

While within a given brain region the human transcriptome has been shown to be incredibly complex, it is also of importance to consider the relationship among different anatomical regions of the brain, as 'disconnectivity' between disparate brain regions is thought to underlie a number of neurodevelopmental syndromes including ASD (Geschwind and Levitt 2007). Perhaps unsurprisingly, there is strong evidence that distinct regions of the human brain have distinct gene expression profiles, and animal studies have suggested that this variation is related to both structural and functional differences (Nadler et al. 2006). For instance, a microarray study of twenty distinct brain and spinal cord sites showed that expression profiles can cluster samples from different donors by anatomical origin, and that some anatomical regions have up to 2,000 region-specific genes (Roth et al. 2006). Multiple studies have shown that the cerebellum contains the most unique gene expression pattern compared to other brain structures (Lockhart an Barlow 2001; Roth et al. 2006; Strand et al. 2007), which is of consequence to autism in particular, as this region has been consistently implicated in the pathogenesis of the disorder (Fatemi et al. 2012). Even just within the neocortex, different cortical layers each express a detectably distinct profile of mRNA transcripts (Molveneaux et al. 2007). Underscoring the importance of region-specific expression are results that have shown gene expression differences between any two brain areas with one individual are more pronounced than are gene expression differences between two different individuals within the same brain region (Strand et al. 2007; Khaitovich et al. 2004; Naumova et al. 2008).

In summary, the human brain has been demonstrated to have a unique pattern and complexity of gene expression both compared to other species and compared to other human tissues, including region specific gene expression patterns, and pervasive transcription of ncRNAs. This highlights the importance of understanding human neuropsychiatric disorders, such as ASD, in the context of human brain gene expression specifically, as it is likely that animal, cellular, and other models do not recapitulate the uniqueness of human brain functional genomics with the appropriate level of fidelity. Moreover, recent evidence is accumulating that suggests gene expression patterns within the human brain vary considerably across developmental time, and therefore temporal patterns of gene expression are also an important consideration.

Changes in Gene Expression During Human Neurodevelopment

The developing human brain grows remarkably fast—the weight of a newborn's brain is approximately 25% of its adult weight, but within two years, it nearly reaches its adult size (Dekaban and Sadowsky 1978). During this time, the brain grows mainly through glial multiplication, myelination, formation of new synaptic connections, and pruning of unused synaptic connections. While the human brain continues to mature up to the age of 25 years (Sowell et al. 2004), the greatest changes occur in the periods of infancy and early childhood. Coincidentally, most neurodevelopmental disorders, including autism spectrum disorders, become clinically recognizable around this age.

Underlying these dramatic early changes in brain development are complex and dynamic broad patterns of gene expression, which have only recently begun to be understood. The most comprehensive study to date of the developing human brain transcriptome (published after the onset of this work; Kang et al. 2011) documented that genome-wide patterns of gene expression correspond closely to the major stages of clinical development (namely prenatal, early infancy, childhood, adolescence, and adulthood), and that the molecular profile of these stages are distinct from each other. The most striking observation was that the greatest shifts in gene expression occur around the period of birth, where the authors found almost 60% of genes change their expression patterns in the neocortex (Kang et al. 2011). Other studies have demonstrated similar changes, and have showed that many of the genes identified during this shift are known to be involved in cortical development and higher order cognitive functioning (Johnson et al. 2009; Lambert et al. 2011).

The Kang et al. study, which profiled RNA expression using whole-genome microarrays on tissue derived from neurologically normal donor brains spanning the 2nd trimester through adulthood, also demonstrated that after infancy the number of genes whose expression profile changes in the neocortex decreases dramatically to approximately 9% of expressed genes between infancy and adolescence, and less than 1% of genes between adolescence and adulthood. Functional annotation of these gene sets further revealed that genes expressed very early in prenatal development are highly related to the process of cell differentiation, proliferation, and migration, while genes expressed later in gestation are more related to synaptogenesis, suggesting that time-period specific gene expression patterns drive cell-level developmental programs. Again, these findings highlight the importance of assessing autism

candidate gene expression and function during the appropriate developmental time window, in order to gain the most relevant insight into this disorder.

In addition to the greatest number of genes shifting their expression trajectory shortly after birth, the changes in gene expression in early post-natal life have also been shown to have greater amplitude of change (Colantuoni et al. 2011; Somel et al. 2009; Somel et al. 2010). In fact, it was shown that many genes actually reverse their expression trajectory in early life (Calantuoni et al. 2011), mostly shifting from a pattern of increasing expression in fetal life and infancy to a decrease in expression beginning in childhood. Moreover, as the brain begins to mature, the gene expression profile within each anatomical region becomes more similar to other regions, with the exception of the cerebellum, suggesting that most of the region-specific development is completed early in life. Interestingly, these broad gene expression patterns appear to reverse themselves in older age, at least in the prefrontal cortex (Somel et al. 2010).

Gene expression dynamics in early human brain development are clearly both spatially and temporally specific. This suggests not only that they are highly regulated, but that different genes and gene networks will have dynamic expression throughout space and time. Despite this increasingly recognized property of human neurodevelopmental genomics, few studies of autism candidate genes have considered their expression and function in early human brain development. Furthermore, the molecular regulators of brain mRNA expression, such as non-coding RNAs, have not been extensively characterized in the developing human brain, and their potential involvement in ASD has hardly been studied. Accordingly, an understanding of the functional genomics of autism, the genetic interaction networks that ASD candidate genes participate in, and the potential ncRNA regulators of these genes, represent important unresolved avenues of research, and are addressed by work in this thesis.

Non-coding RNAs in Human Brain Development

Since the advent of high-throughput, unbiased, genome expression arrays and sequencing platforms, the recognition that the genome is pervasively transcribed at loci that do not encode for protein products is becoming well recognized. The term ‘non-coding RNA’ (ncRNA) is commonly employed for RNA that is transcribed in the cell but does not encode for a corresponding protein product (Mattick and Makunin, 2006). While originally consider

to represent transcriptional 'noise' (Ponting and Belgard 2010), this non-coding RNA component of the transcriptome is increasingly implicated in regulating the genomic landscape though a myriad of mechanisms (**Figure 1.2.2**), and as such are increasingly being recognized as important modulators of gene expression. Moreover, they are also beginning to be implicated in disease.

These ncRNAs can interact with DNA to induce methylation or histone modifications, recruit transcription factors, and modulate the three-dimensional architecture of chromosomes in the nucleus (Ponting et al. 2009). They can bind to other RNA molecules, especially mRNA with complementary sequences, to inhibit translation through RNA degradation, or they can act as 'sponges' and thereby dilute the effect of mRNAs or other ncRNAs (Hansen et al. 2013). Conversely, they can increase the rate of translation by acting as molecular stabilizing scaffolds. They can also interact with proteins and protein complexes to catalyze reactions by acting as linkers among otherwise scarce proteins in the cytosol, and participate in cellular trafficking of RNA binding proteins (Mansfield and Keene 2009). Additionally, ncRNAs have been shown to participate in intracellular communication by helping transport cargo between adjacent cells (Skog et al. 2008; Balaj et al. 2011).

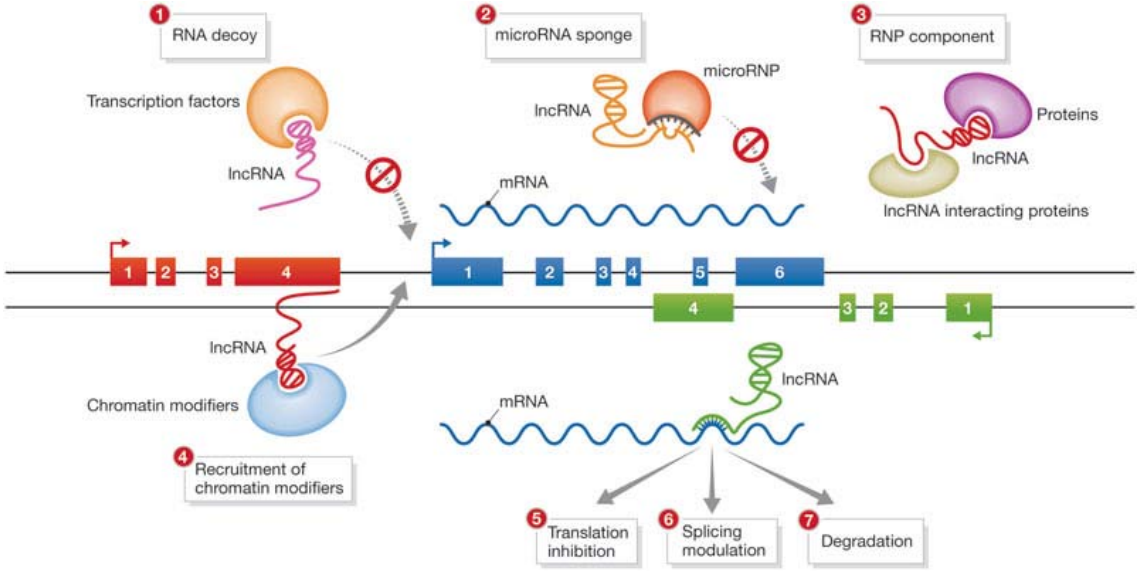


Figure 1.2.2. Schematic of some experimentally validated functions of ncRNAs. Long non-coding RNAs, one class of ncRNAs, have been demonstrated to act through a variety of mechanisms to modulate gene expression. Adapted from: Hu W, et al. *EMBO reports*. 2012; 13:971.

The increasing recognition of this important layer of transcriptome information is perhaps most important in the brain, where it has been shown that the greatest abundance of ncRNAs exist (Mercer et al. 2008). Furthermore, there is strong evidence that ncRNAs played a critical role in the evolution of human brain structure and function. For instance, the fastest evolving regions of the primate genome are sequences that are transcribed to ncRNAs, and it has been shown that these particular ncRNAs are primarily involved in regulating neurodevelopmental genes (Pollard et al. 2006). Similarly, regulation at the level of RNAs through RNA editing mechanisms (which many classes of ncRNAs exhibit high degrees of) has undergone a significant evolutionary expansion in higher primates and humans (Paz-Yaacov et al. 2010).

This expanding inventory of ncRNAs, and their increasing functional and regulatory activities in humans, appear to play an important role in neurodevelopment and neuropsychiatric diseases. In particular, microRNAs (miRNAs) and long non-coding RNAs (lncRNAs) are the best studied of this class of regulatory RNAs, and have recently been implicated in a number of neurodevelopmental disorders, including ASD.

MicroRNAs

The best studied of the ncRNAs are microRNAs, which mainly function to repress translation post-transcriptionally through the RNA interference (RNAi) mechanism. The miRNA family includes a variety of precursor RNA molecules that are classified mainly on their genomic origin, such as endogenous small interfering RNAs (endo-siRNAs) and PIWI-interacting RNAs (piRNAs). However, all classes are quickly processed after transcription into the common mature form of miRNAs that are 20-23 nucleotides in length and single-stranded (Hutvagner and Zamore 2002, **Figure 1.2.3**).

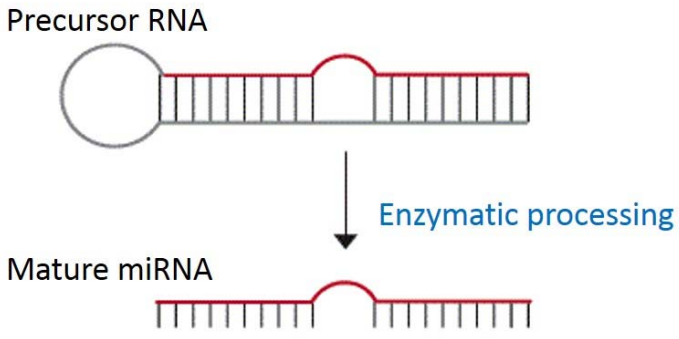


Figure 1.2.3. Representation of precursor and mature miRNAs. A variety of precursor RNAs are processed through a series of steps into their mature, functional forms, which are 20-23nt in length and single-stranded.

In brief, the RNAi mechanism begins when precursor miRNAs are transcribed in their entirety from the genomic DNA, processed by an enzyme complex known as DROSHA, exported to the cytoplasm, and cleaved by the ribonuclease DICER into their mature form. At this point, they bind to a class proteins termed Argonaute proteins, which are then incorporated into a larger multi-protein complex termed the RNA-induced silence complex (RISC). The RISC complex is then guided to mRNAs that are complementary to its associated miRNA, leading to a repression of translation or overt degradation of the mRNA transcript (Liu and Paroo 2010).

Importantly, a single miRNA can target—and therefore regulate—many mRNAs because of their short sequence, their preferential binding to 3'-UTRs, and their imperfect complementary binding to cognate sequences (Hashimoto et al. 2013). Conversely, it has been shown that individual mRNAs are often targeted by multiple miRNAs. Therefore, miRNA-mRNA interactions alone can increase the complexity of gene expression regulation by orders of magnitude. Furthermore, miRNAs are known to target other ncRNAs in addition to protein-coding mRNAs (Hansen et al. 2011). The critical implication of these insights is that a single miRNA has the ability to modulate entire transcriptional networks, and therefore the mis-expression of a single miRNA has the potential to disrupt the proper expression of entire suites of genes.

Recent studies have demonstrated the importance of miRNAs in human brain evolution, cellular development, experience-dependent plasticity, and in neuropsychiatric disorders. A large number of miRNAs exhibit species-specific expression patterns, are conserved only in primates and/or humans, or are exclusively expressed in brain—providing strong evidence for their role in human-specific brain functions and disorders (Somel et al. 2010; Somel et al. 2011). For example, an analysis of human, chimpanzee, and macaque prefrontal cortex and cerebellum showed a substantial degree of divergence in their miRNA expression patterns (Hu et al. 2011).

Additionally, important roles for miRNAs in neural stem cell maintenance and differentiation have been established through a number of studies that have identified and characterized individual miRNAs of interest. For instance, multiple animal studies have shown that

DICER knockout animals display a host of neurodevelopmental defects, including abnormal brain size, structural defects, and improper formation of synapses (Giraldez et al. 2005; Davis et al. 2008; Zhao et al. 2010). One particular miRNA, miR-9, has been studied extensively for its role in developmental patterning and cell migration, where it has been shown to be critical for neural stem cell self-renewal (Zhao et al. 2005), production of some of the earliest neurons in the developing telencephalon, and cortical laminization (Shibata et al. 2011). Cellular studies of pluripotency have demonstrated that introduction of particular miRNAs can reprogram human skin fibroblasts into neuronal-like cells (Yoo et al. 2011), and that this mechanism likely involves the central nature of these miRNAs in canonical transcriptional networks that are known to guide neural cell fate decisions (Wu and Xie 2006). In addition to neurogenesis, miRNAs have also been shown to regulate gliogenesis, in particular the formation of oligodendrocytes and astrocytes (Dugas et al. 2010; Tao et al. 2011).

In the neocortex specifically, miRNA regulation of neural stem cell differentiation and migration has been shown to be critical to normal development in mice, as cortex-specific Dicer mutants have pronounced changes to the proportion of different classes of cortical neurons generated (Saurat et al. 2013). Importantly, the role of miRNAs in neural stem cell differentiation and migration appears to vary over developmental time, as deletion of Dicer before the onset of neurogenesis results in an overall reduction in neuron number (De Pietri et al. 2008), whereas deletion later in development leads to a much milder phenotype (Kawase-Koga et al. 2009). This temporal variability underscores the importance of studying miRNA regulation of gene expression across developmental time.

In addition to being implicated in individual cellular-level functions, miRNAs have been shown to contribute to synaptogenesis and experience-dependent plasticity—both functions thought to underlie complex human behavior, and hypothesized to be disrupted in ASD. For example, specific miRNAs are enriched in the synaptic nerve terminals of axons (Natera-Naranjo et al. 2010), and are up-regulated in expression in the hippocampus following memory tasks in mice (Hansen et al. 2013). In *Drosophila melanogaster*, knockout of subcomponents of the DICER complex results in synaptic transmission defects, but no overt brain structural abnormalities (Smibert et al. 2011), and in mice results in dendritic spine malformations (Davis et al. 2008) and impaired synaptic transmission (Schofield et al. 2011).

Finally, miRNAs are now recognized as contributing to human neurologic disease. Inherited variation in DNA encoding for miRNAs or in miRNA recognition sites has been linked to a number of disorders including schizophrenia (Liu et al. 2012). Dysregulated expression of miRNAs has been demonstrated in brain tumors (Silber et al. 2008), Parkinson's disease (Martins et al. 2011), and Tourette's syndrome (Abelson et al. 2005), and an absence of mature miRNAs due to DICER knockout affects memory and learning in mice (Konopka et al. 2010).

However, despite these substantial observations suggesting that miRNAs are critical regulators of neurodevelopmental transcriptional networks and are often disrupted in neurologic diseases, the miRNA landscape of the developing human brain has not been fully characterized, and only a few small studies have attempted to profile miRNA expression levels in autism spectrum disorder (reviewed below). Thus, there is a need for a comprehensive assessment of miRNAs during human brain development, and a more thorough characterization of miRNA changes in ASD.

Long Non-coding RNAs

In contrast to miRNAs, which are short sequences with well-defined functions in post-transcriptional regulation, long non-coding RNAs (lncRNAs) represent a novel class of transcripts whose function in brain development remains poorly understood. Long non-coding RNAs are defined as RNAs greater than 200 nucleotides in length (as compared to ~21-23 nucleotide length of miRNAs), which do not encode for protein, or lack an appreciable reading frame. LncRNAs undergo post-transcriptional processing similar to other RNAs, providing the first hint at their functional importance. For instance, some lncRNAs are modified to include a 5'-methyl cap and 3' polyadenlyation (Carninci et al. 2005; Diebali et al. 2012). However, unlike miRNAs, long non-coding RNAs are generally poorly conserved evolutionary, and as such, elucidation of their functional roles has relied more upon expression analysis and individual functional studies than on comparative genomic interpretations.

While originally thought to be evolutionary byproducts, long non-coding RNAs have been shown to be involved in major mechanisms of gene expression regulation, such as targeting transcription factors, initiating chromatin remodeling, directing methylation complexes, and

blocking nearby transcription (Ponting et al. 2009). Moreover, pervasive transcription of lncRNAs has been demonstrated to occur in both a temporally and spatially regulated manner during development (Amaral and Mattick 2008), with the central nervous system displaying the greatest abundance of transcribed lncRNAs (Mercer et al. 2008).

Recently, it has been demonstrated that individual lncRNAs play important regulatory roles in the spatial-temporal control of gene expression in the brain specifically. One of the first studies to demonstrate this explored RNA expression from mouse *in situ* hybridization data, and the authors demonstrated that most lncRNAs examined were localized to specific cell types, subcellular compartments, or neuroanatomical regions (Mercer et al. 2008). This work provided some of the first large scale evidence that lncRNAs may have specific functions in their capacity as RNAs alone in the mammalian brain. Subsequently, a number of studies employing both whole-genome and individual candidate lncRNA assessment have begun to expose the importance of lncRNAs to the regulation of the developing brain transcriptome. For example, individual lncRNAs have been shown to be induced in response to neural activity or modulate synaptogenesis (Lipovich et al. 2012; Bernard et al. 2010). They have also been implicated in neuronal differentiation; for instance, the lncRNA *Evf2* recruits a number of important neurodevelopmental transcription factors (such as *Mecp2*, the *DLX* family, and *GAD1*) to their target genes in GABA-ergic interneurons, and *Evf2* knockout mice have reduced interneuron cell numbers (Bond et al. 2009). Transcriptome studies of brain tissue have also begun to characterize the entire landscape lncRNAs during neurodevelopment. It was shown that lncRNAs are differentially expressed across layers of the mouse neocortex (Belgard et al. 2011), and that those expressed in brain are preferentially located in genomic regions containing critical neurodevelopmental genes (Ponjavic et al. 2009).

Long non-coding RNAs have also been found to be abnormally expressed in a number of neurologic disorders. For example, the lncRNA *BACE1-AS* is an antisense transcript of the beta-secretase-1 gene locus, which is implicated in the generation of beta-amyloid plagues in Alzheimer's disease (Faghihi et al. 2008). The *BACE1-AS* lncRNA extensively regulates the level of *BACE1*, and therefore can directly affect the level of beta-amyloid plaque accumulation (Modarresi et al. 2011). Other lncRNAs have been implicated in the

neurodevelopmental disorder Angelman's syndrome (Landers et al. 2005; Meng et al. 2012), which shares many features with ASD.

Therefore, lncRNAs appear to play a critical role in modulating gene expression in the developing brain, and are increasingly implicated in neurological disorders. Their potentially numerous regulatory mechanisms, and largely overlooked sequence variation in disease, makes them an important class of candidate molecules to consider in neurodevelopmental disorders with complex genetics, such as autism spectrum disorders. However, no work has attempted to systematically identify lncRNAs in ASD, despite the increasing recognition of their importance to the complex genomics of brain development.

Gene Networks in the Developing Human Brain

While assessing non-coding regions of the genome is an important approach to comprehensively understand the complex functional genomics of human brain development and neurodevelopmental disorders, it is equally important to consider how disparate genomic elements may work in concert with each other to produce biological effects that are emergent only after their interaction. The study of genetic interactions can be done by modeling large gene sets as networks of interacting nodes and edges, allowing for a statistical assessment of relationships among and between genes, as opposed to the study of individual genes themselves. Such approaches are particularly important in complex genetic syndromes like autism spectrum disorder, as genome wide association studies have consistently demonstrated that most individual variants in ASD have only very small effects by themselves.

The study of gene networks is a subset of the field of network science that applies mathematical and statistical principles to biological systems. A biological network is a system of individual biologic components that interact with each other in a structured, non-random manner, such that properties of the network as a whole emerge that are not apparent by studying the individual components in isolation. Biological networks have been identified at levels spanning molecular (Sharan et al. 2005), cellular (Sanchez-Vives and McCormick 200), organ system (Bullmore and Sporns 2009), and even inter-individual relationships (Croft et al. 2004).

An interaction network consists of nodes and edges. In gene networks, nodes represent discrete genes and edges represent a biological relationship or connection between nodes

(**Figure 1.2.4**). Depending on the interaction being studied, edges can represent many relationships such as known protein-protein binding, correlations of expression levels between genes, or any other metric that can be measured in all nodes and have putative biological relevance to how the system works. As most gene networks are incredibly large (thousands of nodes and millions of possible edges), the study of biological networks relies on a number of mathematical principles adopted from network theory and statistics that allows for the incredible complexity of large networks to be summarized, quantified, and visualized in order to more easily infer biological meaning.

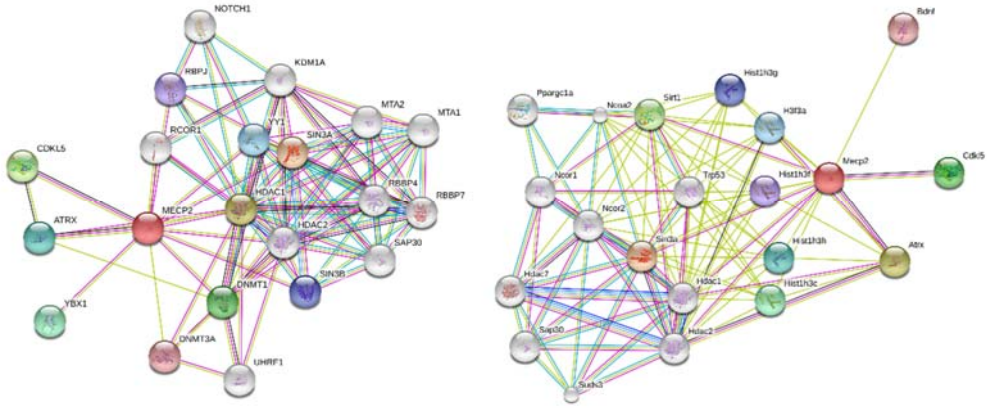


Figure 1.2.4. Examples of gene interaction networks. Properties of networks as a whole can become apparent that would not be appreciated by studying individual genes. For example, the network on the left represents known 2nd degree protein-protein interactions with the gene *Mecp2* in humans, whereas the network on the right represents the known *Mecp2* interactions in mice. As can be seen, the human network is much more densely connected, suggesting *Mecp2* has more known interactions in humans. Networks created with String v9.05 (http://string-db.org/).

A genetic interaction occurs, then, when an unexpected phenotype emerges from the combination of two or more interacting genes. This phenomenon is widely pervasive in genetics and has been recognized for decades. For instance, the phenomenon of synthetic lethality occurs when two genetic mutations that by themselves have no effect, are both present in the same gene and their presence together results in a defective protein product (Hajeri and Amatruda 2012). Furthermore, genetic interactions appear to be ubiquitous throughout the genome (Phillips 2008). However, predicting how independent genes combine with each other to create emergent properties is not straightforward. This is especially true in non-model organism systems, such as the human brain, where it is impossible to experimentally modulate individual genes, and therefore researchers must rely only on observational measures such as gene and protein expression levels.

One validated approach to integrate heterogeneous gene sets, in order to uncover shared molecular mechanisms, is through the analysis of gene co-expression patterns, which invokes the guilt-by-association heuristic that is pervasive in genomics research (Stuart et al. 2003; Wolfe et al. 2005). Several studies have demonstrated that genes with similar brain co-expression patterns are likely to function together in common cellular pathways (Oldham et al. 2008; Winden et al. 2009). These transcriptional co-expression relationships are particularly relevant to neurodevelopment, as the precise regulation of gene expression across brain regions at different ages instructs the exquisite specialization and connectivity within the brain. For instance, if two genes are expressed with similar patterns (i.e. they have a similar magnitude and direction of expression change across developmental time), they would have a higher correlation than two genes whose expression appears to be randomly related to one another. In this network, edges would link genes with similar expression profiles, whereas unrelated genes would not share an edge. Defining edges between genes in this way allows the conclusion that the two nodes share related biological function, and can be used to derive and study large-scale genetic interaction networks.

Another widely used approach to infer interaction networks is to draw upon experimentally determined protein-protein interactions. Protein-protein interaction networks are perhaps the best-characterized networks in all of biology, and many well-curated experimental datasets containing detailed interaction information exist. Studies have demonstrated that protein interaction networks are conserved evolutionarily (Perez-Bercoff et al. 2013), and that proteins in the network with high degrees of connectedness are more important for organismal survival and fitness than those with lesser connectivity (Baryshnikova et al. 2010). This suggests that information on the importance of individual genes/proteins in a network can be inferred by studying the overall structure of the network as a whole. Such an approach could be particularly valuable in disorders like ASD, where there are many implicated genes, but the relative importance of each to the pathophysiology of the disorder is unclear.

Approaches that combine gene co-expression networks and protein interaction networks are also beginning to be developed, and will be an important future approach to understanding the ultimate cellular function of autism candidate genes at the protein level. Recent large-scale proteomics efforts have shown that protein co-expression patterns are slightly better

predictors of protein interactions than are mRNA co-expression patterns (Kim et al. 2014). However, obtaining comprehensive and unbiased datasets of protein co-expression is much more technically challenging than obtaining genome-wide RNA expression levels. Consequently, understanding gene co-expression patterns among autism candidates is an important first step, which can then be used to guide future protein-protein interaction studies of these genes in a more targeted manner, as has been described in other model systems (Tornow and Mewes 2003).

Since neurodevelopmental disorders such as autism are believed to result from functional aberrations within brain regions and/or disruption of inter-regional connectivity between regions (Geschwind and Levitt 2007), investigating the gene expression profiles of autism candidate genes across brain regions and throughout normal human neurodevelopment may provide insight into the complex functional genomics of this neurodevelopmental disorder. Furthermore, as the genetic heterogeneity of ASD continues to increase as more sequencing and association studies are performed, prioritizing candidate genes through their location in interaction networks is an important undertaking.

Conclusion

The human brain transcriptome displays a remarkably complex array of genes and ncRNAs that appear to be highly temporally and spatially regulated. Furthermore, comparative genomics studies have shown that the functional genomics of human brain development diverges significantly from model organisms or even close evolutionary relatives. Non-coding RNAs, miRNAs and lncRNAs in particular, are increasingly recognized as critical regulators of gene expression, and the analysis of gene interaction networks allows for the identification of emergent properties among large sets of genes that are often not apparent when studying individual genes of interest in isolation. Therefore, it is critical that the genes implicated in autism be understood in the specific context of human neurodevelopmental gene expression, that potential critical ncRNA regulators of their expression be identified, and that their network-level properties be explored, as it is increasingly apparent that the appropriate genomic context be applied and understood in complex genetic disorders.

--

1.3 Previous Functional Genomics Studies of ASD

At the onset of this work, very few studies had assessed the expression, regulation, or network properties of autism candidate genes specifically in human neurodevelopment. Similarly, partly due to the scarcity of well-preserved post-mortem tissue, there have been few studies assessing autistic brain tissue directly, and in particular, almost no studies have explored non-coding RNAs and the mitochondrial genome in autistic brain. Recently, newer functional genomics studies in ASD have added critical insight into potentially convergent molecular and regulatory pathways in ASD, many of which are supportive and complementary to the work described in Chapters 2 and 3 of this thesis.

Gene Expression Studies in ASD

The majority of gene expression studies in autism have been performed in peripheral blood lymphocytes (Voineagu 2012); however, as only approximately half of the genes expressed in brain are also expressed in lymphocytes (Cookson 2009), it is critically important that autistic brain tissue be assessed directly. The few genome-wide expression profiling studies in autistic brain tissue have repeatedly identified a number of functions that appear to be disrupted in autistic brain. The first microarray study assessed autistic post-mortem cerebellum and demonstrated dysregulation of AMPA receptor subunits in ASD (Purcell et al. 2001). Subsequently, Garbett et al. analyzed genome-wide microarray expression profiles form six autistic temporal cortices and six controls, and their results suggested an up-regulation of genes involved in immune and inflammatory processes with a concurrent down-regulation of genes involved in neurodevelopment (Garbett et al. 2008). The largest sample size assessed by microarray analysis to date studied three separate brain regions (frontal cortex, temporal cortex, and cerebellum) form 19 autistic cases and 17 controls (Voineagu et al. 2011). This work also demonstrated an up-regulation of genes with known immune function and a concurrent down-regulation of genes involved with the synapse. Importantly, there was a large degree of overlap in the genes identified between the Garbett et al. 2008 and the Voineagu et al. 2011 studies, even though they assessed different cohorts of donor brains.

Interestingly, there is a large body of evidence potentially linking immune-related gene expression changes in autistic brain with prenatal events that may affect the developing brain and result in the autism phenotype. For instance, there is increasing evidence implicating maternal infection during pregnancy and inflammation in the placenta to autism risk. For

example, abnormal inclusions of trophoblasts, the cells that comprise the placenta, are found more frequently in placentas from mothers of children who develop autism than in controls (Anderson et al. 2007). In addition, placental inflammation is associated with impairments in communication and social interaction in the child, as measured by low ratings on the Modified Checklist for Autism in Toddlers (Limperopoulos et al. 2008). It has further been shown that inducing inflammation in the utero-placental compartment in pregnant sheep is sufficient to cause neurological deficits reminiscent of autism in their offspring (Hutton et al. 2007). It is hypothesized that maternal immune activation induces cytokines expression that can activate immune cells within the placenta, eventually leading to abnormal placental signaling and sequent neuroinflammation in the developing fetal brain (Boksa 2010; Hsiao and Patterson 2012). Supporting this theory are studies correlating elevated levels of cytokines in amniotic fluid and maternal blood with increased risk of autism in the child (Abdallah et al. 2012; Goines et al. 2011). Yet it is not understood how inherited genetic risk for autism may relate to these immune findings.

More recently, Chow et al. studied autistic prefrontal cortex across a large age span, and demonstrated dysregulation in pathways governing cell number, cortical patterning, and differentiation in young autistic prefrontal cortex, but in contrast found dysregulation of signaling and repair pathways in adult autistic brain tissue (Chow et al. 2012). Another recent study of autistic cerebellar and occipital brain regions demonstrated no changes in DNA methylation, but significant gene expression abnormalities in mitochondrial oxidative phosphorylation and protein translation pathways (Ginsberg et al. 2012).

Therefore, while the number of studies assessing gene expression in autistic brain are still small, already there appears to be a growing body of evidence implicating disrupted molecular pathways involved in synaptogenesis and immune function, in addition to a number of others. However, how these observations relate to known inherited mutations in autistic candidate genes, or how disruptions of autism candidate gene expression may result in these broad molecular changes, is far from clear.

Non-coding RNA Studies in ASD

Remarkably, the majority of the single nucleotide polymorphisms associated with ASD via GWAS studies have been found in intergenic regions or intronic sequences outside the

protein coding sequences (see Chapter 1.2). Despite the increasing recognition of the importance of non-coding genomic regions in the molecular regulation of human brain development, their relation to the pathogenesis of ASD has hardly been assessed. Moreover, despite the known tissue-specific nature of ncRNA expression, only one study of miRNAs in autism has been done in brain tissue, and until only very recently no studies had assessed long non-coding RNAs in autistic brain.

Four studies have profiled miRNA expression in tissues derived from patients with 'idiopathic' autism. In the only study of autistic post-mortem brain tissue, Abu-Elneel et al. identified 28 differentially expressed miRNAs in ASD cerebellum via qRT-PCR analysis (Abu-Elneel et al. 2008). More recently, a number of microarray studies have assessed for miRNA expression differences in blood lymphocytes of autistic patients as compared to controls, having discovered nearly 100 miRNAs that are abnormally expressed in ASD (Talebizadeh et al. 2008; Sarachana et al. 2010; Ghahramani Seno et al. 2011). Additionally, a recent report described aberrant expression of long non-coding RNAs in autistic frontal cortex and cerebellum that are antisense transcripts to known autism candidate genes (Velmeshev et al. 2013).

Therefore, while it appears that ncRNAs are abnormal in ASD patients, their expression and function during human brain development has not been thoroughly characterized. Furthermore, apart from identification of mis-expressed miRNAs and antisense RNAs, no work has attempted to functionally integrate these findings to determine how they may ultimately result in the ASD phenotype. Consequently, there is a critical need to assess for changes in miRNA and lncRNA expression in autistic brain tissue, and to attempt to determine putative functional consequences of these changes as they relate to the known genetic etiology of ASD.

Network and Pathways Studies in ASD

Finally, several pathway analyses have been performed using either genetic or transcriptome data to gain insight into the biological functions associated with ASD candidate genes. For instance, O'Roak et al. analyzed protein-interaction networks among genes implicated in ASD via whole-exome sequencing studies, and identified that *de novo* mutations in ASD patients are overrepresented among proteins involved in a chromatin remodeling network (O'Roak et al. 2012). Similarly, Gilman et al. demonstrated that CNVs identified in autistic patients are enriched for genes involved in a molecular network related to synaptogenesis,

axon guidance, and neuronal motility (Gilman et al. 2011). Only two studies have attempted to integrate autism candidate genes with known human brain gene expression patterns. Ben-David and Shifman attempted to assess for differences between rare and common ASD candidate genes by studying their co-expression relationships in adult human brain. They discovered these genes were both related to modules involved with synaptogenesis and neuronal plasticity, and that they are expressed in areas associated with learning, memory, and sensory perception (Ben-David and Shifman, 2012). The same authors also recently analyzed the neurodevelopmental expression of ASD candidate genes that had been discovered in cohorts as *de novo* mutations, and demonstrated that these genes appear to relate to networks involved in transcription regulation and chromatin remodeling processes (Ben-David and Shifman 2013).

Summary

In summary, while the number of investigations attempting to integrate genetic findings in ASD lags far behind gene discovery studies, there is evidence that through integrative functional genomics analysis, common pathways and mechanisms underlying ASD may be discovered. This thesis describes some of the first studies that have attempted to integrate large sets of ASD candidate genes to assess for common pathways, and to understand ncRNA regulation of gene expression in autistic brain.

--

1.4 Major Unanswered Questions and Motivations for this Work

The work described in Chapters 1.1 through 1.3 catalogs prior studies that have identified genetic mutations in individuals with ASD, explored the functional genomics of human brain development, and have attempted to integrate what is known about autism genetics with human neurodevelopmental functional genomics. These bodies of work make clear a number of important observations:

(i) Autism spectrum disorders have a significant hereditary/genetic component

(ii) The genetic etiology of ASD is extremely heterogeneous

(iii) Gene expression in the human developing brain is unique from other human tissues and brain gene expression in other model organisms

(iv) Regulation of human neurodevelopmental gene expression is exquisitely regulated through a number of mechanisms including ncRNAs

(v) Autistic brain tissue displays disrupted gene expression patterns

However, there remain a number of fundamental questions about the functional genomics underlying autism spectrum disorders that serve as the main motivations for the studies described in Chapters 2 and 3 of this work:

(i) Are there common developmental gene expression properties/patterns among the genes implicated in autism that may be informative of their role in ASD?

(ii) Do these patterns provide insight into how so many genes with different functions can all relate to the same clinical phenotype?

(iii) Are there inherent gene expression differences between the developing male and female brain that may be informative of the significant bias in ASD seen in males?

(iv) Can studies of non-coding regions of the genome in ASD help explain some of the 'missing heritability' by regulating genes involved in ASD pathogenesis?

These four questions served as the theoretical basis for the work that I performed for this thesis, as is described in the following chapters.

--

Chapter 2. Characterizing ASD Candidate Genes During Human Neurodevelopment

Hundreds of genes have been implicated in autism spectrum disorders through many different approaches. However, most of these studies failed to also assess these genes for expression in human brain tissue, and in early human neurodevelopment in particular. Furthermore, as has already been discussed, it is important to consider the genetic interaction among various autism candidate genes, as their relation to each other throughout human brain development may provide additional layers of information regarding the pathogenesis of ASD.

In this chapter, I describe a set of studies that explored the expression of autism candidate genes throughout human brain development, in order to provide relevant insight into the functional genomics of this complex syndrome. In the first study, I performed the first comprehensive analysis to date of individual ASD candidate gene expression patterns spanning human neurodevelopment. Then, I performed (in collaboration) a study of autism candidate gene co-expression relationships. These two projects represent a comprehensive functional genomics assessment of ASD candidate genes during human brain development, and provide unique insight into common pathways that may underlie ASD pathogenesis.

Next, I performed two analyses on aspects of normal human neurodevelopmental genomics that have important implications for ASD. First, I assessed for sex-specific differences in human brain gene expression during development, as autism spectrum disorders are known to affect a preponderance of males. Then, I describe results of miRNA differential expression analysis across human brain development, and evaluated their relationship to known ASD candidate genes.

The work in this chapter provides critical insight into the functional roles of ASD candidate genes during normal human brain development, and how sex-specific gene expression and miRNA regulation of gene expression may relate to the functional genomics of ASD.

Autism Candidate Genes and Gene List

As detailed in Chapter 1, the evolution of genomic techniques has resulted in a large and increasing set of genes implicated in autism via various approaches. In fact, recent estimates suggest that as many as 1,000 genes or more could contribute to ASD (He et al. 2013, Ronemus 2014), and various reports have already systematically reviewed hundreds of candidate genes with very strong association to ASD (Miles 2011). Given this large and heterogeneous set of putative autism candidate genes, the field has come to rely upon a number of repositories of curated autism candidate genes lists, which themselves continue to grow and evolve over time (Basu et al. 2009; Matuszek et al. 2009; Betancur 2011; Xu et al. 2012; Abrahams et al. 2013).

While these repositories of autism candidate genes serve as important resources to enable researchers to quickly identify and further study genes associated with ASD, any attempt to collect and annotate a comprehensive yet specific list of 'autism candidate genes' will face a number of problems of note. Foremost, due to the extreme heterogeneity in the clinical presentation of ASD, and the fact that ASD remains a qualitative clinical diagnosis, studies of autistic individuals have varying inclusion/exclusion criteria even when standardized protocols are used. For instance, some studies will exclude individuals with significant intellectual disability in order to study 'intrinsic autism,' while others will include cases with even severe intellectual disability. Such variability between studies has the potential to confound the assumption that these lists include genes purely associated with ASD and not one of its many co-morbidities or other endophenotypes.

Secondly, the sample sizes in studies reporting candidate gene associations with ASD vary widely and in general are underpowered (O'Roak and State, 2008), potentially biasing the identified candidates toward those with large effects. Moreover, the population structure in the various patient cohorts studied also varies considerably, challenging the interpretation and significance of the reported candidates because of the differing background allele frequencies in different ethnic groups. Furthermore, once genes are reported as significantly associated with ASD based on achieving a minimum threshold for significance, their inclusion in many databases of ASD candidates treats them equally (Basu et al. 2009, Matuszek et al. 2009), despite their varying significance of association. Even newer attempts to develop 'evidence

scores' for annotated candidates (Xu et al. 2012; Abrahams et al. 2013) remain incomplete, non-systematic, and static to the continuously updated understanding of ASD genetic risk.

As a result, attempts to prioritize ASD candidate genes, either for further investigation, to assess gene-gene interactions, or to infer information about the genetic architecture of ASD, are inevitably biased by the manner in which the individual variants were implicated in ASD initially. It will be important for future work to attempt to quantify in an unbiased manner the strength of ASD candidate association based on the strength of the genetic evidence supporting their involvement in ASD. However, until such a resource is available, one unbiased method to construct sets of autism candidate genes for further study (and the method employed in this work) is simply to include all variants associated to date that are annotated in expert-curated autism databases, remaining agnostic to how they were initially included.

Therefore, in order to further investigate relationships between and among as broad a set of 'autism candidate genes' as possible in this work, I included all ASD candidates that had been annotated by expert curation and published in peer-reviewed databases. The reasoning for this approach was to remain agnostic with regard to type of association with ASD in order to avoid introducing my own biases on top of those already existing, as detailed above. This approach resulted in an ASD candidate gene list that included genes implicated based on i) rare copy number variants and single gene disruptions enriched in cohorts of individuals with ASD as compared to controls ii) genes identified to be mutated in patients with syndromes in which some significant proportion of patients has ASD iii) small risk-conferring genes with common polymorphisms in the general population that have been identified in genome-wide association studies, and iv) 'functional' candidates that have been shown to be mis-expressed in autistic post-mortem brain tissue.

In the work described in Chapters 2.1 – 2.3, I assessed for relationships among these autism candidate genes via gene expression. As the work in this thesis evolved over time, so too did available databases of autism candidate genes. As a result, the autism candidate gene list in Chapter 2.1 is derived only from the AutDB database, as that was the only published resource available at the time the work in Chapter 2.1 was initiated. In Chapters 2.2 and 2.3, however, a larger set of ASD candidate genes was compiled for analysis, as two new databases became

available by the onset of those studies. The full list of ASD candidate genes used in Chapters 2.2 and 2.3 (herein named "*ASD list"*) was created by combining (taking the union) lists from three main ASD genes databases: AutDB, Autism Genetics Database, and AutKB-484, and the resulting list is included in the Appendix as **Table A2**. These databases each independently annotated genes that had previously been associated with autism, and although largely overlapping, there are some differences. As a result, the final combined *ASD list* consisted of 455 unique ASD candidate genes (**Table A2**). The subset of genes assessed in Chapter 2.1, derived only from the AutDB database, are annotated with an asterix in **Table A2**, and are presented as a separate table in the supplementary material (**Supplementary Table S1**).

Description of Normal Human Neurodevelopmental Gene Expression Dataset

In order to assess the expression properties of these autism candidate genes across normal human neurodevelopment, I searched the literature for repositories of human brain gene expression. In the following four sub-chapters (Chapters 2.1-2.4), the work described is based on publically-available gene expression data provided by the Allen Institute for Brain Science, the "BrainSpan Atlas of Human Development" (www.brainspan.org). This database represents the largest and most comprehensive human brain gene expression dataset currently available. The BrainSpan Atlas contains next-generation RNA sequencing (RNA-seq) data from neurologically normal post-mortem human donor brains. This data was downloaded in its entirety and analyzed as described in each sub-chapter (see respective Methods sections below).

The process of collecting donor tissue, sectioning the donor brains into anatomical areas, extracting RNA, performing RNA-sequencing, and aligning the RNA-seq data to the human reference genome (Build HG 37) was all carried out by the Allen Brain Institute. The details of this are available on the Allen Institute website, and is included as **Supplementary File S2**.

The resulting RNA-seq dataset that was used for this analysis consisted of expression values aligned to composite gene models, and given in units of reads per kilobase of exon model per million mapped reads (RPKM, Mortazavi et al. 2008). Upon download of the entire database, genes whose RPKM values were likely to represent noise rather than actual sequenced reads

were discarded by removing any gene that did not have at least one expression value greater than or equal to five RPKM in any of the tissue samples. The remaining dataset consisted of 13,563 expressed genes that were used for analysis.

The full BrainSpan atlas contains data generated from over 40 developing and adult post-mortem brains. However, many of these brains are missing data from some of the regions profiled. Therefore, for this analysis only donor brains with complete data were retained, resulting in a final dataset of 30 total donor brains used here. As shown in **Figure 2.0.1**, the donor brains analyzed span pre-conception through adulthood and contain an approximately equal distribution of male and female donors. To maintain consistency with other recent large-scale studies of human brain developmental gene expression to allow for appropriate comparison of this work with standard approaches in the field, we grouped donor brains into the same seven developmental stages previously described (Kang et al. 2011; Colantuoni et al. 2011, see **Figure 2.0.1**). This strategy resulted in the data from at least four separate donor brains binned together per time point. All available demographic information about each of the 30 donor brains individually is included in the Appendix (**Table A1**).

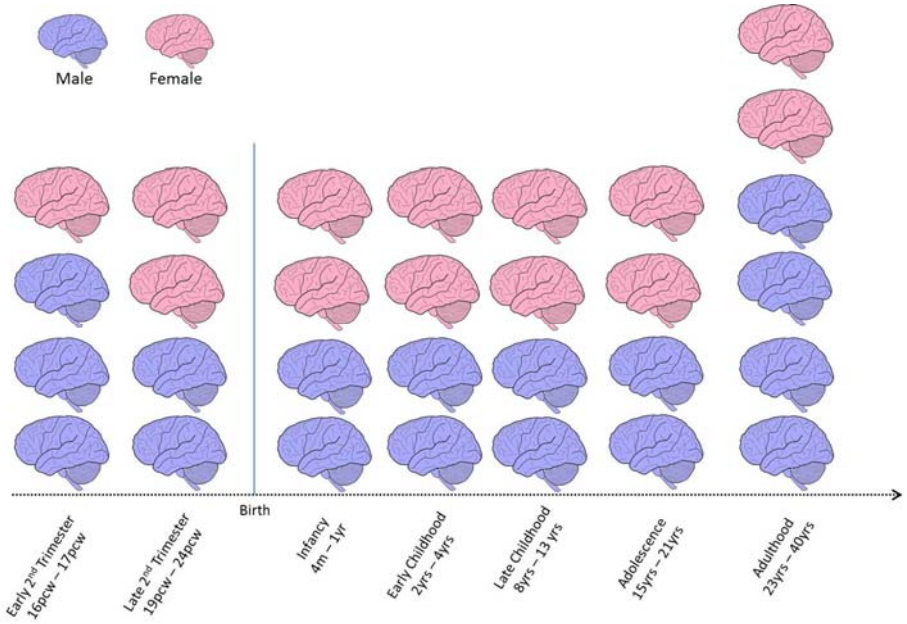


Figure 2.0.1. Temporal description of the number and sex of the assessed brains. The data were grouped into 7 developmental stages. Blue colored brains indicate the donor was male and pink indicates female. Pcw, post-conceptional weeks; m, month; yrs, years.

From each donor brain, RNA-sequencing was performed on RNA extracted from 16 separate brain regions (**Figure 2.0.2**). These regions included the cerebellar cortex (CBC), medial-dorsal nucleus of the thalamus (MD), striatum (STR), amygdala (AMY), hippocampus (HIP),

and 11 areas of the neocortex (NCX): orbitofrontal (OFC), medial frontal (MFC), dorsolateral frontal (DFC), ventrolateral frontal (VFC), primary motor (M1C), primary somatosensory (S1C), inferior parietal (IFC), primary auditory (A1C), superior temporal (STC), inferior temporal (ITC), and primary visual cortex (V1C).

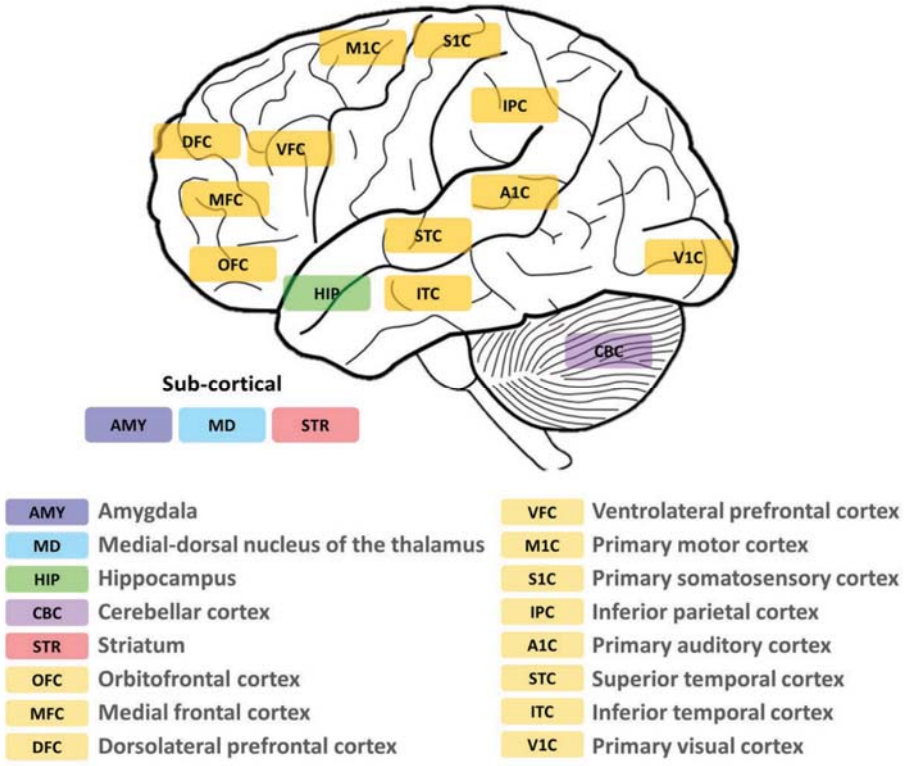


Figure 2.0.2. Graphical representation of brain regions assessed. Representation of the 16 structures that full BrainSpan dataset contains. Note: in chapters 2.1 and 2.4, only a subset of the brain regions were assessed as described in the respective Methods sections.

It is important to note that an assessment of only four brains per time point is a major limitation of this work. However, repositories of high-quality human donor brain tissue with multiple regions from the same individual are scarce (Abbott 2011). Consequently, while underpowered to detect differences of small effect size, this dataset represents the largest repository of gene expression data from human post-mortem tissue that is completely available. Moreover, all samples in this dataset were obtained using a single, standardized, acquisition process and run on the same gene expression platform, removing many of the confounds associated with combining various different RNA-seq datasets (Chu and Corey, 2012).

Despite the relatively small sample size, this is the largest dataset of human post-mortem brain gene expression. Furthermore, due to the uniquely sensitive nature of RNA-sequencing as compared to microarray, analysis of this dataset has reasonable ability to assess for differences in gene expression between groups. For instance, a recent report (Hart et al. 2013) estimating the power of RNA-seq to reliably detect differences in human tissue suggests that in order to identify a fold-change difference of 1.5 between sample groups, approximately 20 samples per group are required (at a power level (β) of 80% and Type 1 error rate (α) of 0.05, assuming a 0.4 coefficient of variation and a 20 million read count depth). Yet at present, no such human brain dataset exits with sample numbers this large. However, using the calculator provided by Hart et al., it can be estimated that the dataset used for this work (with 4-6 samples per group, **Figure 2.0.2**) has 60-80% power (β) to detect a fold change of 2.0 between groups (Hart et al. 2013). As the results presented in the studies of Chapter 2 are focused mainly on broad patterns of expression change not individual genes or miRNAs, the low power to detect all instances of actual differential expression is somewhat further minimized.

Studies of post-mortem human brain tissue are likely to continue to have small sample sizes owing to the nature of the work. Interesting, there is a substantial statistical literature on the assessment of extremely small sample sizes ($N < 5$), even with specific emphasis on neurobiological samples (Janusonis, 2009). In fact, it has been explicitly demonstrated using repeated simulations that the Student's t-test functions validly with small sample sizes, even as few as $N = 2$, provided the effect size being tested is large and the underlying population distributions are normal (de Winter, 2013). Furthermore, the study by de Winter further demonstrated that the t-test behaves similarly to other approaches for small sample size, such as rank transformations and Welch's test (de Winter, 2013).

Interestingly, the Student's t-test was originally developed specifically for small sample sizes (as an alternative to the z-test), and was first described with a use-case of $N = 4$ (Student, 1908; Zabell, 2008). Therefore, while working with extremely small sample sizes in this thesis is a limitation, as the chances of introducing Type II errors is relatively high (i.e. that true effects of small size will be missed), evidence suggests that the identified results should otherwise remain statistically valid provided the assumptions underlying parametric t-testing (normally distributed populations) hold true (de Winter, 2013), as is generally assumed to be

the case in human brain gene expression studies (Kang et al. 2011; Akula et al. 2014). As with all statistical inference, the chance of Type I error (i.e. false positive) remains, and the degree to which Type I error is likely depends on the resulting p-value, the ability to replicate the finding, and how surprising the results are in the given biological context. Throughout this work, I have attempted to minimize Type I error to the extent possible by correcting p-value for multiple testing comparisons, performing confirmatory RT-PCR, exploring the biological relevance of the identified genes in terms of the published literature and other bioinformatic approaches, and generally remaining conservative of the interpretations.

In summary, while larger sample sizes are always desirable in order to produce the most sensitive results, this dataset represents a unique opportunity to discover broad patterns of gene expression change of large effect that are unique to human neurodevelopment, a process that has largely been understudied and therefore limits efforts to better understand human disorders of neurodevelopment. While the small sample sizes used in this work have statistical power limitations, and future work should assess larger sample cohorts, there is still valuable insight into human neurodevelopmental genomics that can be obtained. Furthermore, throughout the work in this thesis I have made every attempt possible to incorporated additional analyses of the identified differentially expressed genes in order to place these results in biological context to better ensure their biological relevance. It will be important for future studies to replicate these efforts when more samples become available, but these studies represent some of the first to assess the functional genomics of autism candidate genes in human brain tissue and can therefore provide important initial insight.

--

2.1 Expression Profiling of Individual Autism Candidate Genes

2.1.1 Aim

Despite substantial efforts to uncover the genetic basis of ASD, the genomic etiology appears complex and a clear understanding of the molecular mechanisms underlying autism remains elusive. I hypothesized that focusing gene interaction networks on ASD-implicated genes that are highly expressed in the developing brain may reveal core mechanisms that are otherwise obscured by the genomic heterogeneity of the disorder. Here I report an *in silico* study of the gene expression profiles of ASD-implicated genes in the unaffected developing human brain. By implementing a biologically relevant approach, I identified a subset of highly expressed ASD-candidate genes from which interactome networks were derived. Strikingly, immune signaling through *NFκB*, *Tnf*, and *Jnk* was central to ASD networks at multiple levels of the analysis, and cell-type specific expression suggested glia—in addition to neurons—deserve consideration. This work provides integrated genomic evidence that ASD-implicated genes may converge on central cytokine signaling pathways.

2.1.2 Introduction

A major question in ASD research is how to reconcile the genetic and phenotypic heterogeneity of the disorder with the apparent convergence of molecular mechanisms into synaptic network abnormalities. One proposed unifying explanation posits that differences in gene expression in the developing brain could explain how many genes, each with a different contribution to proper formation of brain circuitry, could result in a single disorder with neural network dysfunction at its core (Geschwind 2008; Levitt and Campbell 2009). This model is underscored by the prototypical autism spectrum disorder, Rett Syndrome, in which mutations in the *Mecp2* gene result in global dysregulation of the transcriptome (Chahrour et al. 2008). Moreover, it has been shown that mutations in *Mecp2*—a transcriptional repressor—result in aberrant expression at many ASD-implicated loci (Samaco et al. 2005). To investigate this model, however, requires gene expression profiling of ASD-candidate genes in developing human brain tissue. At the onset of this work, a number of studies had investigated gene expression in post-mortem brain tissue of patients with ASD (Lintas et al. 2010), with three examining ASD brain tissue on a genome-wide scale (Purcell et al. 2001; Garbett et al. 2008; Voineagu et al. 2008). However, no study had explicitly described the

transcriptional profile of ASD-implicated genes. Furthermore, individual genes of interest that had been studied in human neurodevelopment were limited in developmental time points and brain regions investigated.

To investigate more thoroughly the notion that differences in expression of ASD-implicated genes underlies the complex genomics of the disorder, I hypothesized that focusing gene interaction networks on ASD-implicated genes with high expression in the developing brain may reveal core mechanisms that are otherwise obscured by the heterogeneity of all implicated loci. To do this, I mined the *BrainSpan* Atlas of Human Brain Development (Jones et al. 2009) for all genes implicated in ASD that were included in the database AutDB (Basu et al. 2009). I devised a biologically-driven computational approach to analyze differential expression across regions and development, and assessed cell-type specific expression using the Human Protein Atlas (Berglund et al. 2008). I discovered distinct molecular interaction networks using an enriched set of highly expressed genes, which implicated canonical immune signaling pathways at multiple levels of analysis as central to ASD.

2.1.3 Methods

Neurodevelopmental Disorder Databases

As described previously, autism candidate genes were obtained by using the complete AutDB database (**Table A1** and **Supplementary Table S1**, Basu et al. 2009). To concurrently investigate the relationship among autism and related neurodevelopmental disorders, and to serve as control datasets, I also obtained similar lits of Schizophrenia and Epilepsy-assoociated genes from the SzGene (Allen et al. 2008) and CarpeDB (Galperin 2005) databases, respectively (**Supplementary Tables S3 and S4**). In all three databases, some implicated regions are provisional loci, non-coding RNAs, pseudogenes, or otherwise not included in the *BrainSpan* Atlas and, therefore, were not considered.

BrainSpan Atlas of Human Brain Development

The *BrainSpan* Atlas was accessed on 2/16/2011 at www.brainspan.org, and the raw Gene Matrix .csv datafile was downloaded. I re-organized the data so that rows are genes and columns are developmental time points subdivided according to brain region. While the full dataset contains data from 16 different brain regions and up to age 40 (**Figures 2.0.1 and2.0.2**), the analysis in this Chapter considered only 11 brain regions that are most relevant

to autism (Schumann et al. 2011) and only donor brains up to 23 years old. This was done in order to decrease the computational burden of the analysis approach I developed. The brain regions assessed in this Chapter are: Dorsolateral Prefrontal Cortex (DLPC), Ventrolateral Prefrontal Cortex (VLPC), Medial Prefrontal Cortex (MPC), Orbital Prefrontal Cortex (OPC), Posterior Superior Temporal Cortex (PSTC), Inferior Lateral Temporal Cortex (ILTC), Hippocampus (Hipp), Amygdala (Amyg), Striatum (Stri), Cerebellum (Cere), and Primary Motor Cortex (PMC).

Gene Expression Analysis

Genes in AutDB, CarpeDB, and SZGene were parsed from the full database to create disease-specific expression datasets (**Supplementary Tables S5-S7**). Expression values were divided into quintiles and given corresponding colors for heat map creation (<20 RPKM, 20-40 RPKM, 40-60 RPKM, 60-80 RPKM, >100 RPKM). Genes were then assigned to one of five tiers within each brain region based on their highest level of expression across all time points, in a conservative attempt to analyze the expression data qualitatively. For example, if a gene is expressed at 150 RPKM at 24 weeks gestation (wg) and at 80 RPKM for all other time points, it is placed in the >100 RPKM tier. Based on results from established housekeeping genes (see Results), I considered a gene to be differentially expressed if it crossed more than three tiers. Because of this, genes in the top three tiers were considered to be "highly expressed," and were the focus of my subsequent analysis (Supplementary Tables S5-S7, "Highly Expressed Genes" tab).

Gene Ontology Enrichment Analysis

To test if a subset of genes implicated different Gene Ontology categories than a background set of all genes, I employed the Gene Ontology Enrichment Analysis and Visualization tool (Eden et al. 2009), accessed at http://cbl-gorilla.cs.technion.ac.il/. I specified the organism as *Homo sapiens*, chose the option for two unranked lists of genes, and set the p-value threshold to 0.01. Raw p-values were then converted to False Discovery Rate (FDR) q-values within the software.

Ingenuity Pathway Analysis

Integrated gene-network analysis for the AutDB, CarpeDB, and SZGene sets and on the highly expressed subsets were generated by Ingenuity Pathways Analysis (Version 8.8, Ingenuity® Systems, www.ingenuity.com). Each gene identifier was mapped to its corresponding gene object in the Ingenuity Pathways Knowledge Base. The gene lists were

overlaid onto a global molecular network developed from information contained in the Ingenuity Pathways Knowledge Base. These focus gene networks were then algorithmically generated based on their connectivity.

Canonical pathways analysis identified the pathways from the Ingenuity Pathways Analysis library of canonical pathways that were most significant to the data set. The significance of the association between the data set and the canonical pathway was measured in two ways: i) A ratio of the number of molecules from the data set that map to the pathway divided by the total number of molecules that map to the canonical pathway and ii) Fisher's exact test was used to calculate a p-value determining the probability that the association between the genes in the dataset and the canonical pathway is explained by chance alone. A p-value of less than 0.01 was considered significant. For comparison analysis between all disease genes and highly expressed genes, Benjamini-Hochberg multiple testing correction was used to calculate p-values, with 0.01 set as a significance threshold.

Functional network analysis identified the biological interactions that were most significant to the molecules in the network. The network molecules associated with biological functions and/or diseases in Ingenuity's Knowledge Base were considered for the analysis. Right-tailed Fisher's exact test was used to calculate a p-value determining the probability that each biological function assigned to that network was due to chance alone, with a threshold of 0.01 set for significance. A graphical representation of the molecular relationships between molecules was generated. Molecules are represented as nodes, and the biological relationship between two nodes is represented as an edge (line). All edges are supported by at least one reference from the literature. Nodes are displayed using various shapes that represent the functional class of the gene product. Edges are displayed as either solid or broken lines to describe the nature of the relationship between the nodes (solid for direct interaction, broken for an indirect interaction).

Human Protein Atlas

To compare expression data at the transcriptome level to protein-level expression, I also accessed the Human Protein Atlas (Berglund et al. 2008) at http://www.proteinatlas.org/. The Human Protein Atlas is a publicly available database cataloging the distribution of proteins in different normal human tissues, cancer types, and cell lines via validated antibody analysis. The data includes immunohistochemisty, western blot analysis and, for a large fraction of genes, a protein array assay and immunofluorescent based confocal microscopy. I utilized

the reported levels of antibody staining as given, except for genes that contained annotated expression results, which are reported instead.

2.1.4 Results

Evaluating Differential Expression in the BrainSpan Atlas

The *BrainSpan* Atlas reports the normalized reads per kilobase of exon model per million mapped reads (RPKM) units (Mortazavi et al. 2008); whereas primary RNA-seq analysis pipelines have the advantage of using raw read counts for statistical evaluation of differential gene expression. Thus, I first established a qualitative differential expression methodology that could directly interpret RPKM values with consistency and validity across different brain regions and time points. This allowed me to identify a subset of genes that were highly expressed directly from RPKM data.

To achieve this, I examined the expression profile of the top 15 genes determined by Hsiao *et al* as regularly expressed from 59 different whole-genome microarrays in 19 different tissue types (Hsiao et al. 2001). For 11 of the 15 genes there was consistency in expression across developmental time points and in different brain regions (**Supplementary Table S8**). To validate this approach further, I selected at random 10 canonical housekeeping genes representing 10 different cellular processes (Eisenberg and Levanon 2003). I observed consistent expression for all 10 of these genes across brain regions and time points (**Supplementary Table S9**). This resulted in a total of 21 housekeeping genes with constant expression (11 from Hsiao *et al* and 10 canonical), which were used to define normal biological variance in the *BrainSpan* dataset. To stratify the data based on expression, I grouped expression values into quintiles (<20 RPKM, 20-40 RPKM, 40-60 RPKM, 60-80 RPKM and >100 RPKM), as is often done for microarray expression data (Tebbenkamp et al. 2010). Of the 21 constantly expressed housekeeping genes, all vary within three consecutive quintile tiers. Based on these results, I concluded that genes crossing more than three tiers were likely to be significantly differentially expressed, as opposed to exhibiting normal biological variation. This initial approach demonstrated that reported RPKM values could be used qualitatively to assess differences in gene expression levels.

Notably, expression values at the 6 month time point were considerably lower for almost all genes and brain regions. This may be a function of lowered CNS transcriptional activity at this age, however a systematic error in sequencing is also likely. Since we were interested in

highly expressed genes, we were not concerned this would introduce false-positive results into our subsequent analysis.

Expression of Brain-specific Markers

Next, I analyzed genes of brain-specific markers (adhesion/elastic/filament proteins) with intermediate expression to further validate this method and gain insight into cell-type specific expression across brain regions and during different developmental time points. As seen in **Supplemental Table S10**, Keratin and Desmin—markers of epithelia and muscle, respectively—were not expressed as expected. Neurofilament (*Nefl*), a neuron-specific maker, showed high expression in most brain regions after 24 weeks gestation (wg). Notably, expression of *Nefl* was significantly lower in the cerebellum, which is consistent with my observation at the protein level (see below). Expression of Glial Fibrillary Acid Protein (*Gfap*), an astrocyte-specific intermediate filament, also showed high expression in all brain regions beginning at the fourth postnatal month, although markedly later in development than *Nefl*. Interestingly, Vimentin, a marker of mesenchyme-derived cells, exhibits a differential expression pattern with very high expression in the early developing brain (24 weeks gestation – 4 months). This may be a reflection of invading microglia, which are of mesenchymal origin and known to enter the developing brain during early embryogenesis (Ginhoux et al. 2010), and/or it may relate to the laying down of the vasculature and extracellular matrix early in development.

Expression Profile of Genes Implicated in ASD, Epilepsy, and Schizophrenia

I then parsed the database for all genes implicated in autism that were described in the database AutDB (**Supplementary Table S1** and Methods). To strengthen my approach and investigate the overlapping genetic and clinical aspects of schizophrenia and epilepsy with ASD, I also investigated all genes implicated in these disorders, which are cataloged in the databases SZGene and CarpeDB, respectively (see Methods). Non-redundant, protein-coding loci that were present in the atlas were included in this study, as summarized in **Figure 2.1.1**. Only 11 genes are shared by all three disorders. Gene ontology (GO) enrichment analysis of these 11 overlapping genes as opposed to all genes implicated in all three disease databases yielded many significant pathways mainly involved in the response to external stimuli and GABA metabolism (**Table 2.1.1**).

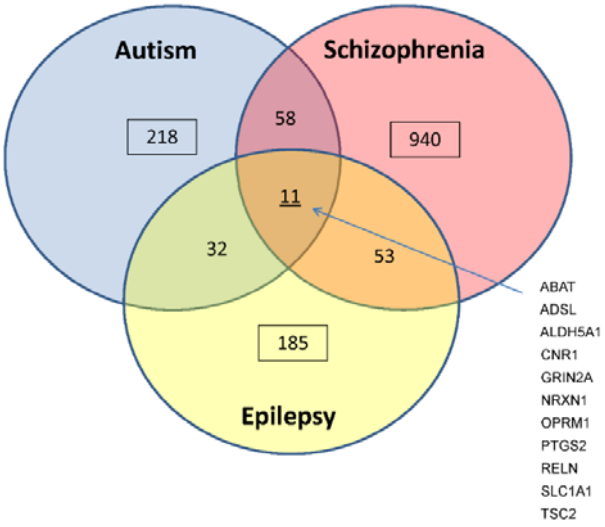


Figure 2.1.1. Summary of all genes analyzed from AutDB, CarpeDB, and SZGene. Number of genes and genes shared between disorders indicated.

Table 2.1.1. GO enrichment analysis of the 11 genes shared by autism, schizophrenia, and epilepsy.

GO Term	Description	P-value	FDR q-value	Enriched Genes
GO:0042220	response to cocaine	1.11E-05	1.89E-04	CNR1, OPRM1, ABAT, GRIN2A
GO:0014073	response to tropane	1.11E-05	9.44E-05	CNR1, OPRM1, ABAT, GRIN2A
GO:0014070	response to organic cyclic compound	6.99E-05	3.96E-04	CNR1, PTGS2, OPRM1, ABAT ,
GO:0009450	gamma-aminobutyric acid catabolic process	7.68E-05	3.26E-04	ABAT, ALDH5A1
GO:0051259	protein oligomerization	1.22E-04	4.15E-04	TSC2, ADSL , SLC1A1, ALDH5A1
GO:0043279	response to alkaloid	1.75E-04	4.96E-04	CNR1, OPRM1, ABAT, GRIN2A
GO:0009605	response to external stimulus	1.85E-04	4.49E-04	TSC2 , NRXN1, CNR1 , RELN, PTGS2, OPRM1, GRIN2A
GO:0051260	protein homooligomerization	1.99E-04	4.23E-04	TSC2, SLC1A1, ALDH5A1
GO:0009448	gamma-aminobutyric acid metabolic process	2.29E-04	4.33E-04	ABAT, ALDH5A1
GO:0032103	positive regulation of response to external stimulus	6.05E-04	1.03E-03	CNR1, PTGS2, OPRM1
GO:0010042	response to manganese ion	7.57E-04	1.17E-03	PTGS2, GRIN2A
GO:0042135	neurotransmitter catabolic process	7.57E-04	1.07E-03	ABAT, ALDH5A1
GO:0031622	positive regulation of fever generation	7.57E-04	9.90E-04	CNR1, PTGS2
GO:0031620	regulation of fever generation	7.57E-04	9.19E-04	CNR1, PTGS2
GO:0031650	regulation of heat generation	7.57E-04	8.58E-04	CNR1, PTGS2
GO:0031652	positive regulation of heat generation	7.57E-04	8.04E-04	CNR1, PTGS2
GO:0009607	response to biotic stimulus	8.35E-04	8.34E-04	CNR1, PTGS2, OPRM1, GRIN2A

I constructed expression heatmaps for all genes by brain region and time-point by assigning each RPKM expression value to one of five quintiles, and then grouping genes into five expression tiers. Strikingly, for each of the three disease sets more than 55% of genes were never expressed above 20 RPKM, with the majority of these less than 5 RPKM (**Table 2.1.2**). For ASD candidate genes, greater than 70% were not expressed highly in each brain region. In each region, a large percentage of ASD-implicated genes had no detectable transcription

(<1 RPKM). For instance, in the hippocampus 46 out of 219 (21%) ASD-implicated genes had no detectable transcripts. Similar proportions were not detected in the cerebellum (52/219 or 24%) or dorsolateral prefrontal cortex (40/219 or 18%). While it is possible that these loci still have functional roles in ASD genomics via *cis*-regulation or other mechanisms, I reasoned that their inclusion in protein-interaction networks might obscure more prominent molecular mechanisms underlying ASD.

Table 2.1.2. Summary of differential gene expression across all brain regions.

Region	% of genes less than 20 RPKM			% of genes in Top 3 Tiers		
	Autism	Epilepsy	Schizophrenia	Autism	Epilepsy	Schizophrenia
DLPC	71%	59%	67%	7%	18%	16%
VLPC	73%	59%	67%	7%	17%	16%
MPC	77%	59%	69%	7%	18%	15%
OFC	74%	58%	69%	6%	20%	16%
Motor	74%	61%	68%	6%	18%	16%
PS Temp	74%	57%	69%	6%	19%	16%
IL Temp	73%	59%	68%	8%	21%	15%
Hippo	79%	63%	70%	5%	15%	14%
Amygdala	74%	56%	66%	5%	19%	16%
Striatum	73%	58%	67%	6%	19%	16%
Cerebellum	77%	62%	67%	8%	14%	15%

It is of note that the cerebellum and frontal cortex contained the greatest number of highly expressed "autism genes" and the temporal cortex had the greatest number of "epilepsy genes," whereas schizophrenia gene expression distributed more evenly throughout the brain. While much work in autism has focused on the hippocampus as a potential epicenter of pathology, we found the developing hippocampus had the fewest ASD candidate genes expressed at high levels, and none were specific for the hippocampus. Conversely, the cerebellum contained a unique set of six autism candidate genes that were not highly expressed in any other brain region. These included the canonical neurodevelopmental genes *Nlgn3* and *Reln*, two cell adhesion molecules, and 7-dehydrocholesterol reductase. This is intriguing since multiple imaging studies have implicated the cerebellum in the pathogenesis of autism (Schumann and Nordahl 2011). The *BrainSpan* Atlas parcels the frontal cortex into four subregions, yet the expression profile of ASD genes between them was similar. Only one gene (*Gabrb3*) was specific to the frontal cortex, and it was only present at high levels in the ventrolateral prefrontal cortex. Interestingly, this gene lies in the 15q11-13 imprinted region implicated in Prader-Willi and Anglemen Syndromes, and is one of the most reproducible loci identified in ASD genome-wide association studies (Buxbaum et al. 2002).

For the remainder of the analysis, I focused on genes in the top three expression tiers (at least one time-point >60 RPKM) as genes that are significantly highly expressed as compared to all ASD-implicated genes (based on my "housekeeping gene" results). This yielded 32 genes for autism, 42 for epilepsy and 212 for schizophrenia (**Fig. 2.1.2**). Autism shared eight highly expressed genes with schizophrenia, and only two with epilepsy (*Dcx* and *Cnr1*). GO enrichment of these nine shared genes did not identify any significant pathways. There was only one gene—Cannabinoid Receptor 1 (*Cnr1*)—implicated in all three disorders that is highly expressed in the developing brain. *Cnr1* expression is high mainly during gestation, and is most pronounced in the cerebellum and amygdala (**Supplementary Table S11**).

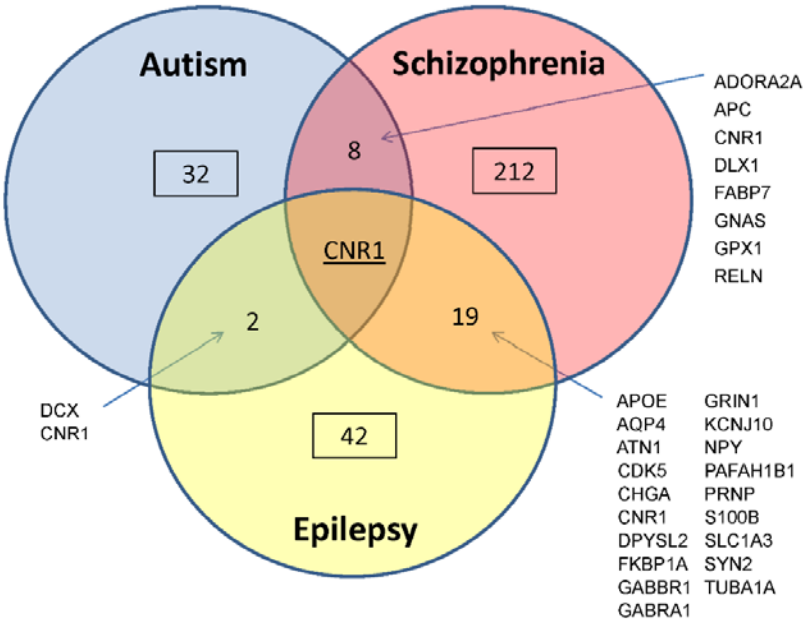


Figure 2.1.2. Summary of the subset of highly expressed genes identified.

Nine autism genes were highly expressed in all brain regions examined. These nine genes (*Fabp7, Gnas, Gpx1, Hnrnph2, Hras, Pdzd4, Rpl10, Sez6l2,* and *Tspan7*) had considerably higher expression than all other ASD genes (over 500 RPKM in many instances, **Supplementary Table S12**). Their temporal expression profiles were mostly constant across developmental stages, except for *Fabp7*, which exhibited drastic differential expression. *Fabp7* was expressed much higher than the other eight genes during almost all time-points, but was highest during the two gestational time points. Interestingly, *Fabp7* (Fatty acid binding protein 7) is known to interact with *Notch* in radial glia during development

(Anthony et al. 2005), and I subsequently found it to only be expressed in glia (see below). The temporal expression of the other 32 highly expressed genes varied considerably, but was biased toward high expression in the early time points analyzed.

Gene ontology enrichment of the 32 highly expressed autism genes revealed four new GO categories representing two significant processes—immune system regulation and apoptosis (**Table 2.1.3**). GO enrichment of the highly expressed Schizophrenia genes yielded a much different set of processes, mostly implicating cellular morphogenesis, but none involving the immune response (**Table 2.1.4**). The epilepsy dataset did not enrich for any significant functions when considering those genes that were highly expressed. This suggests that ASD-implicated genes with no or low expression in the developing brain may obscure functional pathway analysis, which otherwise implicates cytokine signaling.

Table 2.1.3. GO enrichment analysis of highly expressed autism genes.

GO Term	Description	P-value	FDR q-value	Enriched Genes
GO:0002682	regulation of immune system process	0.0001	0.0004	HLA-A, APC, ADORA2A, GNAS, MEF2C, PRKCB, CNR1, CADM1, GPX1
GO:0006915	apoptosis	0.0009	0.0018	APC, MEF2C, ADORA2A, PRKCB, CADM1, HRAS
GO:0012501	programmed cell death	0.0009	0.0012	APC, ADORA2A, MEF2C, PRKCB, CADM1, HRAS
GO:0031347	regulation of defense response	0.0009	0.0090	HLA-A, MEF2C, ADORA2A, CNR1, CADM1 , GPX1

Network Analysis

Next, I set out to determine if the genes identified as being highly expressed in the developing brain implicate different functional networks as compared to all genes associated with these diseases. To do so, I utilized integrated gene-network analysis using the curated Ingenuity Pathway Analysis (IPA) database. Initially I searched for canonical pathways for each disorder, comparing the highly expressed gene sets to all disease-associated genes (**Tables 2.1.5-2.1.7**). This analysis implicated many new canonical pathways from the set of highly expressed genes not seen in the full dataset analysis. For autism, this included corticotrophin releasing hormone signaling, g-protein and phospholipase C signaling, and neutrophil cytokine signaling. The new pathways implicated in schizophrenia included synaptic long-term potentiation and axon guidance signaling, and in epilepsy semaphorin signaling and the splicesome cycle. Interestingly, there are no canonical pathways shared between the three disorders when the entire set of implicated genes is considered, but analysis of the *highly expressed sets* implicates Reelin Signaling in Neurons as common to all three

disorders. Further investigation of this pathway (**Supplementary Fig. S13**) shows almost all molecules are implicated in at least one of these three neurodevelopmental disorders.

Table 2.1.4. GO enrichment analysis of highly expressed schizophrenia genes.

GO Term	Description	P-value	FDR q-value	Enriched Genes
GO:0048812	neuron projection morphogenesis	0.00001	0.00010	APC, PTPRZ1 , CTNNA1, MAPK8IP2 , CDK5, NR4A2, RTN4R, TSPO, OMG, SNAP25, PAFAH1B1, ADORA2A, CCK , DRD2, APOE, WNT7B , S100B , NTNG2 , APP
GO:0032990	cell part morphogenesis	0.00002	0.00010	APC, PTPRZ1, CTNNA1, MAPK8IP2, CDK5, NR4A2, RTN4R, TSPO, OMG, SNAP25, PAFAH1B1, ADORA2A, CCK, DRD2, APOE, WNT7B , S100B, NTNG2, APP
GO:0048858	cell projection morphogenesis	0.00002	0.00010	APC, PTPRZ1, CTNNA1, MAPK8IP2, CDK5, RTN4R, NR4A2, TSPO, OMG, SNAP25, PAFAH1B1, ADORA2A , CCK, DRD2 , APOE, WNT7B, APP, NTNG2, S100B
GO:0032989	cellular component morphogenesis	0.00007	0.00020	PPP3R1, APC, PTPRZ1, CTNNA1, MAPK8IP2, RELN, CDK5, RTN4R, NR4A2, TSPO, OMG, PAFAH1B1 , SNAP25, ADORA2A, SLC1A3, CCK, DRD2 , ATP2B2 , APOE, NTNG2, CAP2 , S100BB, APP
GO:0007409	axonogenesis	0.00007	0.00020	APC, PTPRZ1, CTNNA1, NR4A2, RTN4R, TSPO, SNAP25, PAFAH1B1 , CCK, DRD2, WNT7B, APOE, APP, NTNG2, S100B
GO:0090066	regulation of anatomical structure size	0.00008	0.00020	ADORA2A, SPTAN1, NTS, AKT1, CAPZA2 , GSN, APOE, ATP2B2 , AGT, NEFM, GPX1
GO:0051129	negative regulation of cellular component organization	0.00063	0.00110	APC, SFRP1 , MAG, YWHAH, NGFR, MAPT, GSN, RTN4R, OMG, PAFAH1B1, CLU, SPTAN1, CAPZA2, HSPA1B, RTN4, MBP
GO:0032535	regulation of cellular component size	0.00065	0.00100	SPTAN1, CAPZA2, GSN, AKT1, ATP2B2, NEFM
GO:0010721	negative regulation of cell development	0.00083	0.00110	CTNNA1, SFRP1, PAX6, MAG, YWHAH, NGFR, RTN4R, OMG, TSPO, DLX1, RTN4, MBP
GO:0007417	central nervous system development	0.00083	0.00100	MAL, ADORA2A, PTPRZ1, UGT8, MOG, RELN, PAX6, NGFR, GSTP1, S100B, MBP, ATN1
GO:0030030	cell projection organization	0.00084	0.00090	APC, PTPRZ1, CTNNA1, MAPK8IP2, CDK5, RTN4R, NR4A2, TSPO, OMG, PAFAH1B1, GNAO1, SNAP25, ADORA2A, CCK, DRD2, L1CAM, AKT1, WNT7B, APOE, ATP2B2, NTNG2, APP, S100B
GO:0031344	regulation of cell projection organization	0.00096	0.00100	APC, MAG, YWHAH, NGFR, MAPT, RTN4R, NPTN, OMG, PAFAH1B1, CNR1, AKT1, FEZ1, APOE, RTN4, NEFM, MBP

Table 2.1.5. Canonical Pathways implicated in ASD when considering all genes versus highly expressed genes. ** indicates the pathway was implicated in both sets. #indicates the pathway was common to all three disorders. P-values shown are corrected for multiple testing using the Benjamini-Hochberg method.

Canonical Pathways Derived from All AutDB Genes	B-H Corrected P-Value	Canonical Pathways Derived from 32 Highly Expressed ASD Genes	B-H Corrected P-Value
Serotonin Receptor Signaling	4.4E-08	Corticotropin Releasing Hormone Signaling	4.5E-05
Reelin Signaling in Neurons**	5.4E-06	G-Protein Coupled Receptor Signaling**	2.5E-04
HER-2 Signaling in Breast Cancer	5.0E-05	Role of NFAT in Cardiac Hypertrophy	3.1E-04
cAMP-mediated Signaling**	6.9E-05	Reelin Signaling in Neurons**#	3.7E-04
G-Protein Coupled Receptor Signaling**	7.6E-05	Factors Promoting Cardiogenesis in Vertebrates	5.0E-04
Virus Entry via Endocytic Pathways**	1.7E-04	α-Adrenergic Signaling	5.1E-04
Macropinocytosis Signaling	1.8E-04	cAMP-mediated Signaling**	5.1E-04
Axonal Guidance Signaling	2.2E-04	Virus Entry via Endocytic Pathways**	5.5E-04
PTEN Signaling	3.8E-04	G Beta Gamma Signaling	5.9E-04
GABA Receptor Signaling	4.3E-04	Phospholipase C Signaling	7.2E-04
Glioblastoma Multiforme Signaling	5.9E-04	Cholecystokinin/Gastrin-mediated Signaling	7.8E-04
PI3K/AKT Signaling	9.1E-04	fMLP Signaling in Neutrophils	9.5E-04

Table 2.1.6. Canonical Pathways implicated in schizophrenia when considering all genes versus highly expressed genes. ** indicates the pathway was implicated in both sets. #indicates the pathway was common to all three disorders. P-values shown are corrected for multiple testing using the Benjamini-Hochberg method.

Canonical Pathways Derived from All Schizophrenia Genes	B-H Corrected P-Value	Canonical Pathways Derived from 212 Highly Expressed Schizophrenia Genes	B-H Corrected P-Value
Glutamate Receptor Signaling**	1.0E-32	Glutamate Receptor Signaling**	1.3E-12
Amyotrophic Lateral Sclerosis Signaling**	4.0E-23	Reelin Signaling in Neurons**#	4.4E-09
Neuropathic Pain Signaling In Dorsal Horn Neurons	7.9E-22	cAMP-mediated Signaling	1.5E-08
CREB Signaling in Neurons**	1.0E-21	14-3-3-mediated Signaling	4.3E-08
Role of Macrophages, Fibroblasts and Endothelial Cells in Rheumatoid Arthritis	1.3E-20	Axonal Guidance Signaling	6.6E-08
Role of Osteoblasts, Osteoclasts and Chondrocytes in Rheumatoid Arthritis	1.3E-19	p70S6K Signaling	8.7E-08
G-Protein Coupled Receptor Signaling	4.0E-18	CREB Signaling in Neurons**	9.1E-08
Human Embryonic Stem Cell Pluripotency	1.3E-17	Synaptic Long Term Potentiation	1.9E-07
Serotonin Receptor Signaling	5.0E-17	Myc Mediated Apoptosis Signaling	1.1E-06
Glucocorticoid Receptor Signaling	7.9E-17	Amyotrophic Lateral Sclerosis Signaling**	1.1E-06

Table 2.1.7. Canonical Pathways implicated in epilepsy when considering all genes versus highly expressed genes. ** indicates the pathway was implicated in both sets. #indicates the pathway was common to all three disorders. P-values shown are corrected for multiple testing using the Benjamini-Hochberg method.

Canonical Pathways Derived from All Epilepsy Genes	B-H Corrected P-Value	Canonical Pathways Derived from 42 Highly Expressed Epilepsy Genes	B-H Corrected P-Value
GABA Receptor Signaling**	4.7E-09	Reelin Signaling in Neurons**#	1.3E-06
Neuropathic Pain Signaling In Dorsal Horn Neurons	2.1E-06	GABA Receptor Signaling**	2.8E-04
Reelin Signaling in Neurons**	2.6E-05	Semaphorin Signaling in Neurons	7.1E-03
β-alanine Metabolism	2.2E-04	Spliceosomal Cycle	7.2E-03
Glutamate Receptor Signaling**	7.4E-04	Glutamate Receptor Signaling**	1.1E-02
Calcium Signaling	8.5E-04		
Cellular Effects of Sildenafil (Viagra)	1.1E-03		
Butanoate Metabolism	2.7E-03		
Hepatic Cholestasis	5.5E-03		
Glutamate Metabolism	6.0E-03		

Unbiased gene-network analysis was then constructed in IPA, to identify connectivity networks derived from the enriched gene set compared to those derived from all autism-associated genes. Overlaying derived networks based on connectivity revealed that the two networks constructed from the highly expressed ASD genes are central to all networks obtained from all ASD-associated genes (**Figure 2.1.3**). In the first central network (**Figure 2.1.4**), *NFκB*, *Jnk*, and *Mapk* are hubs. Network 2 from the highly enriched set also contains *NFκB* as a hub, in addition to *Tnf*, *TgfB1* and *Myc* (**Figure 2.1.5**). Taken together, these enriched networks, which are the most inter-connected of all ASD-derived networks, have at their core fundamental cytokine signaling molecules not previously implicated as ASD susceptibility loci. These may serve as potential final common pathways through which the heterogeneous ASD-implicated genes ultimately converge. Moreover, this represents a third, independent level of analysis whereby the highly expressed ASD genes implicate immune signaling pathways that are not apparent when the full set of ASD-associated genes is considered.

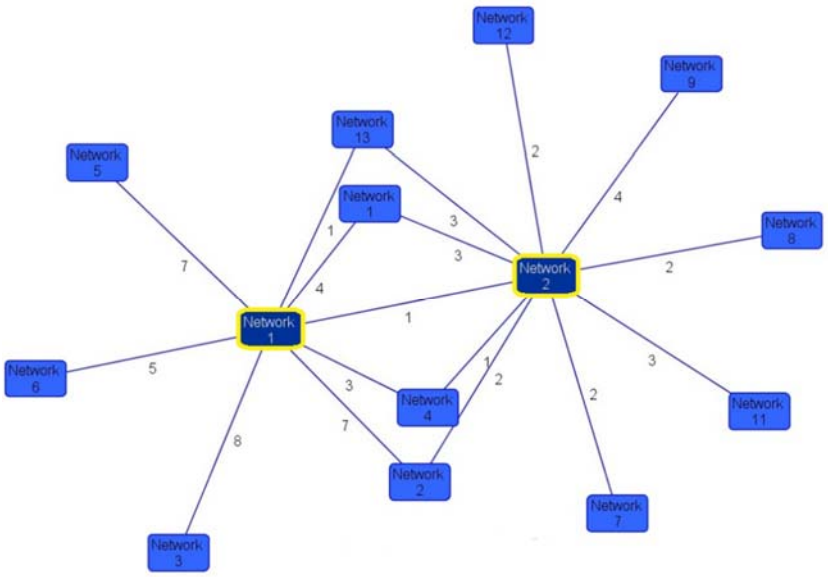


Figure 2.1.3. Overlapping gene-networks in ASD. Networks 1 and 2 (yellow border) were derived from the highly expressed ASD gene set. All other networks were derived from the set of all ASD-implicated genes. Orphaned networks (no edges) were excluded. Edge values represent number of interactions between networks.

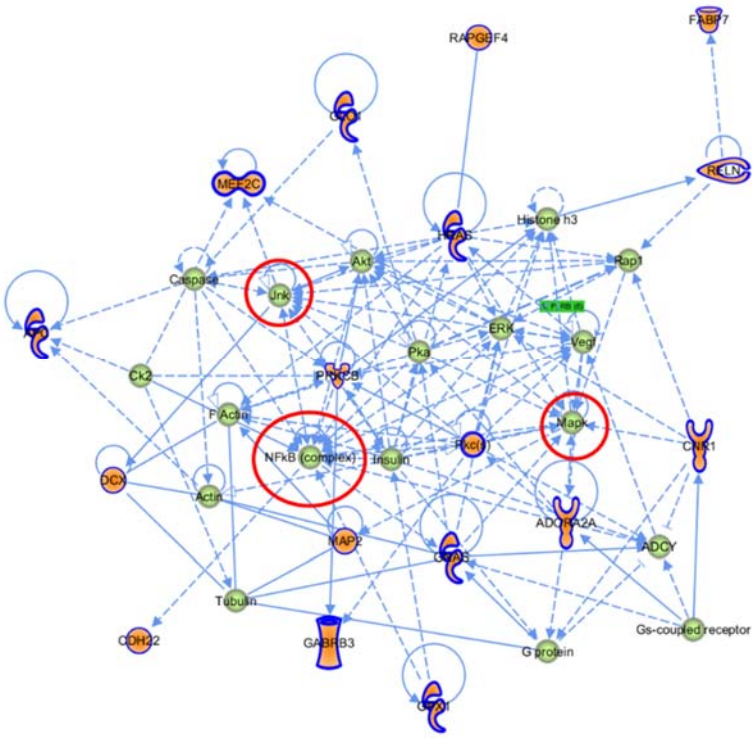
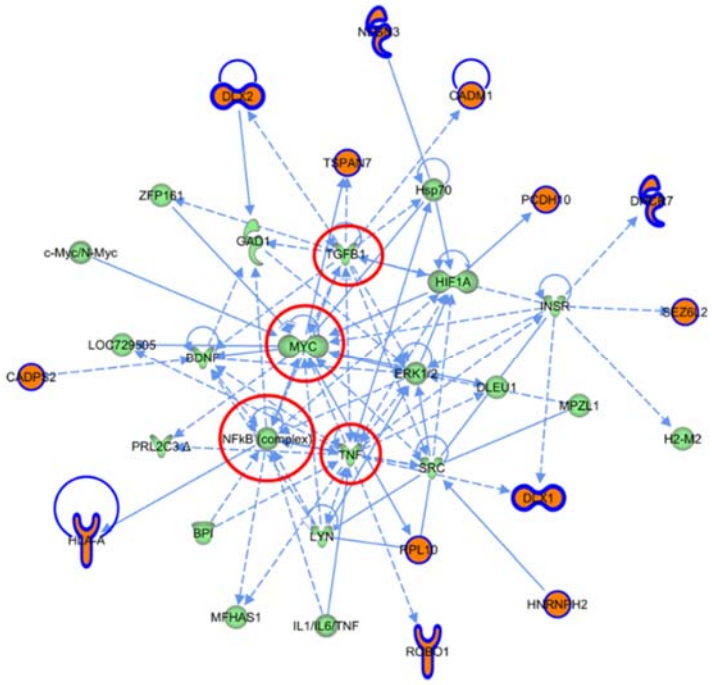


Figure 2.1.4. Network 1 derived from the ASD highly expressed gene set. Orange genes are those present in AutDB that are highly expressed.

Figure 2.1.5. Network 2 derived from the ASD highly expressed gene set. Orange genes are those present in AutDB that are highly expressed.



Similar analysis comparing ASD-associated gene networks specific to brain regions did not result in a significant clustering by region, nor were there central network nodes. Considering only genes expressed highly during gestational time points, we did not observe any new pathways or networks not already implicated using all time points. Gene-network analysis of the epilepsy and schizophrenia gene sets did not result in centrality of the highly expressed networks as we observed in ASD, perhaps reflecting the less heterogeneous nature of these disorders.

Correlating Gene Transcription with Cell-type Specific Protein Expression

Next, I was interested in correlating the ASD gene transcriptome results with protein expression levels in a cell-type specific manner. To do so, I mined the Human Protein Atlas database for the 32 highly expressed autism genes (see Methods, **Table 2.1.8**). Surprisingly, many of the highly expressed ASD genes were mainly detected in glia not neurons, and/or in specific layers of the cerebellum. A similar proportion of genes exhibited neuron-specific protein expression. For instance, the gene *Gnas*, a complex locus known to be imprinted and express antisense and non-coding transcripts (Peters and Williamson 2008), does not appear to make detectable protein in the CNS, yet is one of the 9 most highly expressed ASD

transcripts. Similarly, *Fabp7*, which was *the* most highly expressed ASD-associated gene, was *only* detected in glia. Moreover, *Cnr1*—the one highly expressed gene shared by all three disorders—is most highly expressed in glial cells and the molecular layer of the cerebellum. These results suggest investigation of cell-type specific expression in ASD will be an important undertaking, and consideration of non-coding RNAs in ASD pathogenesis is warranted as well. A limitation of this assessment is that there are not currently published data on glial-subtype specific protein expression patterns, and therefore I was unable to assess for specific types of glial cells.

Table 2.1.8. Cell-type specific protein expression of highly expressed ASD genes from the Human Protein Atlas database.

	Neurons			Glia			Cerebellum		
	Cortex	Hipp	Lat Vent	Cortex	Hipp	Lat Vent	Purkinje	Granular	Molecular
FABP7	-	-	-	++	+++	++	-	-	++
GNAS	-	-	-	-	-	-	-	-	-
GPX1	+	+	-	++	++	++	++	-	-
HNRNPH2	+++	+++	+++	+++	++	++	++	+++	+++
HRAS	+++	+++	+++	+++	+++	+++	-	+++	+++
PDZD4	++	+	++	+	+	+++	++	-	++
RPL10	+++	+++	+++	+++	++	++	+++	++	+++
TSPAN7	+++	++	++	-	-	-	-	-	-
MAP2	+++	+++	+++	-	-	-	+++	+++	+++
PRKCB	++	++	++	-	-	+	++	+++	++
MEF2C	+++	+++	+++	+++	+++	+++	+++	+	+++
RAPGEF4	+++	++	-	++	+	++	+++	+	++
APC	+	+	++	++	++	++	++	++	+++
DCX	+++	+	+	++	++	++	++	++	+++
RIMS3	+	+	+	-	-	-	+	++	-
ROBO1	++	++	++	+++	++	++	++	++	++
GLO1	++	++	++	++	++	++	-	-	+
DLX2	+++	++	++	+	+	+	++	+	-
CNR1	++	+	+	+++	+++	+++	+	++	+++
PCDH10	++	++	+	+++	+++	+++	+	++	+++
NLGN3	++	+++	+	-	-	-	+++	+++	-
RELN	+	+	+	-	-	-	+	-	-
CADM1	-	-	-	-	-	-	-	+++	-
CDH22	+	-	-	+	+	+	-	++	-

Data is reported as presented in the Atlas: +++ for "strong" expression, ++ for "moderate," + for "weak," and – for "negative." These highly expressed ASD genes were not present in the database: *Sez6l2, Gabrb3, Hsd11b1l, Hla-A, Dlx1, Adora2a, Cadps2,* and DHCR7. Lat vent = Lateral Ventricle, Hipp = Hippocampus.

Analysis of ASD Transcriptome Data

Lastly, I was interested in considering these findings in the context of the major three published transcriptomics studies on ASD brain tissue (Purcell 2001; Garbett 2008; Voineagu 2011). I examined the results of all three studies to determine how many of the genes that are differentially expressed in ASD brain tissue have previously been implicated in ASD through inherited or *de novo* DNA mutations. Surprisingly, in each study only ~5% of genes that were significantly differentially expressed in ASD brains were previously implicated in ASD (**Table 2.1.9**). This is particularly intriguing because both the analysis presented here of known ASD candidate genes and two of the three gene expression studies all implicate immune signaling pathways in ASD, even though most ASD-implicated genes profiled in this study are not dysregulated in ASD brain tissue

Table 2.1.9. Correlation of AutDB genes with published transcriptome studies in ASD brain.

	Garbet *et al* 2008	Voineagu *et al* 2011	Purcell *et al* 2001
Brain tissue studied	STC	STC, PFC, Cerebellum	Mainly Cerebellum
# of samples	6 ASD, 6 Ctrls	29 ASD, 29 Ctrls (cortex) 11ASD, 10 Ctrl (cerebellum)	10 ASD, 23 Ctrls
Transcriptome Profiling Method	U133 Plus 2.0 GeneChip (Affymetrix)	Ref8 v3 Array (Illumina)	Clontech Array and UniGEM V2 Array
# of genes dysregulated in ASD	130	444	30
Main findings	↑ Immune-related genes ↓ Genes involved in neuronal development	Genes converge on immune and synapse modules	↑ AMPA-type glutamate receptors
Dysregulated genes in AutDB	4/130 (3%)[1]	21/444 (4.7%)[2]	1/31 (3.2%)[3]

[1] SDC2, SLC9A9, DLX1, AHI1
[2] CD44, CDH10, DLX1, DPP6, GABRB3, HLA-A, KCNMA1, MET, NOS2A, PRKCB1, PTGS2, SCN1A, SLC25A12, NLGN4Y, CADM1, A2BP1, AHI1, PCDH10, PDZD4, CADPS2, SLC9A9
[3] CNR1
STC = Superior temporal cortex, PFC = Pre-frontal cortex, cere = cerebellum

2.1.5 Discussion

In an attempt to integrate the genomic heterogeneity underlying the complex etiologies of common neurodevelopmental disorders, I report here the analysis of expression from genes implicated in autism, schizophrenia, and epilepsy across the developing human brain. Sakai *et al* recently constructed a protein interactome network using a yeast two-hybrid screen on a subset of ASD candidate genes (Sakai et al. 2011), but no study had yet attempted to derive molecular pathways underlying ASD by investigating as large of a set of ASD candidate genes.

To do so, I first described gene ontology, canonical pathways, and interactome networks for all genes implicated in ASD that are cataloged in the database AutDB. Then, I developed a biologically relevant methodology to extract a subset of highly expressed ASD-implicated genes from the *BrainSpan* Atlas of Human Brain Development. I found that interactome analysis placed the two networks derived from highly expressed ASD candidate genes at the center of all ASD gene networks. Closer inspection of these networks revealed *NFκB*, *Jnk*, *MapK*, *TNF*, *TGF-B*, and *Myc* as central hubs. These central networks were supported by evidence at two other levels of the analysis (gene ontology and canonical pathways). Taken together, these findings integrate a large set of genes implicated in ASD and suggest that they may converge onto classical cytokine signaling pathways. While other transcriptomics studies on ASD tissue have implicated immune system signaling in ASD pathogenesis, these findings suggest that the ASD-implicated genes *themselves* may also be related to these functions.

Interestingly, there is also mounting evidence at the cellular and tissue levels that more in depth investigation of an immune component is warranted in ASD (Goines and Van de Water 2010). For instance, multiple studies have demonstrated altered cytokine profiles in ASD patients (Croonenberghs et al. 2002; Molloy et al. 2006), and altered *TGF-B* concentration in serum and CSF correlates with disease severity (Ashwood et al. 2008). Others have described various autoimmune phenomena including autoantibodies to neural antigens and maternal-fetal cross-reactive neural antibodies (Braunschweig et al. 2008). There is also indication of altered innate cellular immunity in ASD, such as differences in gene expression and altered response to immunostimlulatory ligands in both natural killer and monocytic cells from ASD patients (Enstrom et al. 2009; Enstrom et al. 2010). Post-mortem brain tissue from ASD patients shows increased microglial density in grey matter, an activated morphology,

and secretion of a cytokine profile consistent with a pro-inflammatory state, most prominent in the cerebellum (Vargas et al. 2005; Morgan et al. 2010). Moreover, microglia from MeCP2- null mice—a model of Rett Syndrome—produce a conditioned media that damages synaptic connectivity via a glutamate-excitotoxicity mechanism (Maezawa and Jin 2010). While all of this work provides post-hoc evidence for altered immune response in ASD, the results presented here suggest a direct link between genes implicated in ASD based on DNA mutations and molecular pathways involved in immune signaling.

This considerable attention to the immune response in ASD research has resulted in two prevailing theories: one suggests exogenous factor(s) stimulate neuro-inflammation during development, while the other postulates autoimmune activation causes ASD pathology (Pardo et al. 2005; Derecki et al. 2010). However, it is equally possible that the mutations described in ASD result in aberrant signaling regulation of immune cells during neurodevelopment. This could result in cell-autonomous activation and/or improper response to otherwise nominal stimuli, such as occurs in the autoinflammatory syndromes (Kastner et al. 2010). Alternatively, as glia are increasingly implicated in normal formation of synaptic connectivity (Bolton and Eroglu 2009)—and these results demonstrated a significant proportion of ASD-implicated genes appear to be glial-specific—it is possible that genomic aberrations ultimately funnel through core signaling pathways of glial cells to disrupt formation of neural networks independent of an inflammatory mechanism. In support of this notion, a number of recent reports have demonstrated that these same cytokine signaling pathways are central to proper brain development (Kacimi et al. 2011; Awasaki et al. 2011). Furthermore, signaling through the NFkB pathway has been shown to be important in synaptic plasticity independent of inflammation (Mattson 2005).

Interestingly, the canonical cytokine receptors, toll-like receptors (TLRs), are abundantly expressed not just in the peripheral immune system but also in the brain (Visser et al. 2006). Microglia express the full repertoire of TLRs, but some of these receptors are also present in neurons (Lehnardt, 2010). Experimental evidence suggests that distinct TLRs regulate neural plasticity and development in neurons. For example, TLR3 inhibits neural progenitor cell proliferation in the embryonic mouse telencephalon and regulates axonal growth (Cameron et al. 2007; Lathia et al. 2008). In addition, TLR8 is involved in injury and neurite outgrowth associated with neural development (Ma et al. 2006). Similarly, TLR2 and TLR4 play a role in adult neurogenesis of the hippocampus (Okun et al. 2010). Furthermore, a recent study that

systematically characterized the expression of TLR1–9 during pre- and postnatal development of the mouse brain demonstrated these receptors tend to have distinct developmental time-dependent patterns (Kaul et al. 2012). The fact that distinct TLRs exert specific expression patterns over time in the developing mouse brain suggests a physiological relevance of specific TLRs in vertebrates in brain development.

Moreover, two of three genome-wide expression studies in autistic brain tissue concluded that the most prominent transcriptome changes were related to neuro-immune disturbances. In the Garbett *et al* study, the most significant functional pathway implicated was NFκB signaling (Garbett et al. 2008). The most comprehensive transcriptomics study of ASD post-mortem brain to date concluded that one of two significant co-expression networks is involved in immune function (Voineagu et al. 2011). While the results presented in this chapter are only a first step in linking common molecular interaction pathways to the underlying genetic heterogeneity of ASD, they provide integrated genomic evidence, which is supported by these transcriptomics, cell, and tissue level studies that further investigation into cytokine signaling in ASD is needed.

2.1.6 Conclusion

In summary, the work reported in this chapter demonstrates the spatial and temporal expression profile of genes implicated in autism spectrum disorders, in addition to the genetically and phenotypically related neurodevelopmental disorders schizophrenia and epilepsy. I discovered that a large proportion of ASD-implicated genes are not expressed in the developing human brain, and a significant number appear to be mainly expressed in glial cells. Integrated gene-network analysis, gene ontology enrichment, and canonical pathways investigation of a subset of highly expressed ASD genes all implicate central immune signaling pathways as common to the heterogeneous interactome of the implicated genes.

--

2.2 Co-expression Network Analysis of Autism Candidate Genes

2.2.1 Aim

As hundreds of diverse genes have been implicated in ASD, an important first step toward understanding the functional genomics of this disorder was to analyze individual expression profiles of ASD candidates individually, as was presented in Chapter 2.1. However, understanding how so many genes, each with disparate function, can all be linked to a single clinical phenotype remains unclear. To address this, I hypothesized that understanding functional relationships between autism candidate genes during normal human brain development may provide convergent mechanistic insight into the genetic heterogeneity of ASD. To test this hypothesis, I analyzed in collaboration the co-expression relationships of 455 genes implicated in autism using the *BrainSpan* database. We discovered modules of ASD candidate genes with biologically relevant temporal co-expression dynamics, which were enriched for functional ontologies related to synaptogenesis, apoptosis, and GABA-ergic neurons. Furthermore, we also constructed co-expression networks from the entire transcriptome and found that ASD candidate genes were enriched in modules related to mitochondrial function, protein translation, and ubiquitination. Hub genes central to these ASD-enriched modules were further identified, and their functions supported these ontological findings. Overall, our multi-dimensional co-expression analysis of ASD candidate genes in the normal developing human brain suggests the heterogeneous set of ASD candidates share transcriptional networks related to synapse formation and elimination, protein turnover, and mitochondrial function.

2.2.2 Introduction

Mechanistic understanding of how ASD candidate genes relate to the neurobiology of autism is a difficult task, since genes encode multiple highly complex functions at different stages of development and across different regions of the brain. Moreover, the set of genes implicated in ASD is highly heterogeneous, and many of their functions are completely unknown. Furthermore, understanding how disruption in different genes with disparate functions still results in a common clinical phenotype makes developing common targeted biomarkers and treatments for ASD challenging. Therefore, in addition to attempts to identify genes that are causative for ASD, and to understand their individual expression profiles, it is equally

important to understand how ASD candidate genes may relate to each other during human neurodevelopment in order to identify potential shared molecular pathways.

A global survey of ASD gene co-expression patterns across normal human neurodevelopment could therefore facilitate our translation of ASD candidate genes to ASD candidate pathways, but this has not yet been undertaken. A recent study that assessed autism gene co-expression patterns in two adult human brains is an important step toward this goal (Ben-David and Shifman 2012), but as autism is a neurodevelopmental disorder it is imperative to understand the relationship of autism candidate genes in a developmental context. Conversely, other studies have explored the expression profiles of individual ASD candidates in human brain development (Kang et al. 2011), but lack an assessment of the relationships among these ASD candidates and how they relate to global transcriptional pathways important in brain development.

Transcriptome-based studies of the developing human brain have previously been limited in the sample size, number of brain structures analyzed, and developmental time points assessed, hampering the ability to evaluate the genetic contributors to neurodevelopmental disease comprehensively (Abrahams et al. 2007; Ip et al. 2010; Johnson et al. 2009; Somel et al. 2010; Sun et al. 2005). However, the recent availability of broad developmental surveys of gene expression, which cover many brain regions over multiple developmental stages, can greatly facilitate such analysis (Kang et al. 2011). The *BrainSpan* Atlas presents a unique opportunity to understand the spatial and temporal co-expression properties of ASD candidate genes.

We developed a biologically driven computational approach to deduce functional relationships among this diverse set of genes. We first discovered modules of ASD candidates with biologically relevant temporal co-expression dynamics. These modules were related to the processes of synaptogenesis, apoptosis, and the neurotransmitter γ-aminobutyric acid (GABA). Then, we created a transcriptome-wide co-expression network from all genes expressed in the brain, to discover significant 'Molecular Interaction Modules,' and demonstrated that ASD candidate genes are enriched only in modules related to the processes of synaptogenesis, mitochondrial function, protein translation, and ubiquitination. Lastly, we identified hub genes within the ASD-enriched Molecular Interaction Modules, whose

functions supported our ontological results, and which may serve as additional ASD candidate genes. Our analysis of this multi-dimensional expression data suggests pathways previously independently implicated in autism are related to each other through shared neurodevelopmental transcriptional networks.

2.2.3 Methods

Developing Human Brain Transcriptome Data and ASD Gene List

We downloaded the *BrainSpan* transcriptional atlas from www.brainspan.org, processed the data, and grouped samples into developmental timepoints as was described above (page 51, Figure 2.0.1). We were able to design an efficient computational algorithm to analyze the co-expression relationships, therefore all 16 brain regions and all 30 donor brains were analyzed as described above (**Figure 2.0.1 and 2.0.2**, page 51-52). In this analysis, the list of autism candidate genes used was the full 455 ASD gene list detailed in **Appendix Table A2**.

Co-expression Netowrk Creation and Statistical Analysis

In order to identify functional relationships between ASD candidate genes, we investigated patterns of gene co-expression change across developmental stages between each pair of genes from the *ASD list*. First, the correlation between each pair of ASD genes was calculated *separately* within each of the seven developmental stage bins, based on the Spearman's rank correlation between the two genes across all brain regions. For each gene-pair, this resulted in a correlation value for each of the seven developmental stages, representing the brain-wide transcriptional similarity between the genes at each developmental stage (**Figure 2.2.1**). Gene-pairs were retained only if they had an absolute correlation value *greater than 0.8* in at least one developmental stage. We used the Spearman's Rank Correlation as it focuses more on the similarity in the change of gene expression, as opposed to similarity in the absolute values of gene expression.

Second, the surviving gene-pairs were hierarchically clustered into distinct modules based on the similarity of their correlation profiles over time (using the Euclidean distance between the profiles and a complete linkage to merge clusters). Finally, the correlation pattern for each module was summarized by averaging all the gene-pair correlation patterns included in the respective module. It is worth noting that the patterns within the modules represent changes

in co-expression across development (which should not be confused with actual expression levels of genes).

Lastly, in order to assess the biological relevance of these networks, we created 10,000 random networks, each consisting of 455-randomly selected genes from this dataset. We constructed co-expression networks from each of these random sets in the same manner as described above, and then determined how many gene-pairs remain above the correlation threshold of 0.8 in each random co-expression network. We then employed the Hypergeometric probability test to determine if the number of gene pairs with correlation above 0.8 in the ASD co-expression network is more than would be expected by chance based on these 10,000 random co-expression networks. This analysis indicated that the ASD co-expression networks display a significant enrichment for co-expression relationships ($p = 1x10^{-4}$) as compared to the random networks, indicating that our threshold of 0.8 is likely to be capturing true biological relationships in the data (**Appendix Figure A3**).

Gene set enrichment and Gene Ontology enrichment analysis

Enrichment of transcriptome-wide Molecular Interaction Modules for ASD candidate genes and cell-type specific genes was assessed using the hypergeometric probability density function (hygepdf) in MATLAB R2011a (The MathWorks, Inc.). The resulting p-values were corrected for multiple testing using false discovery rate (FDR). All results reported are the –log10 of FDR-corrected p-values, and only p-values < 0.001 were considered significant.

Gene list were assessed for shared biological pathways by testing for enrichment of gene ontology terms (GO) using DAViD Bioinformatics Resources 6.7 (Huang et al. 2009). The complete list of expressed genes in this study's dataset (13,563 genes) was used as the background. Only gene ontology terms with a Benjamini-Hochberg multiple testing corrected p-value < 0.01 are presented as significant.

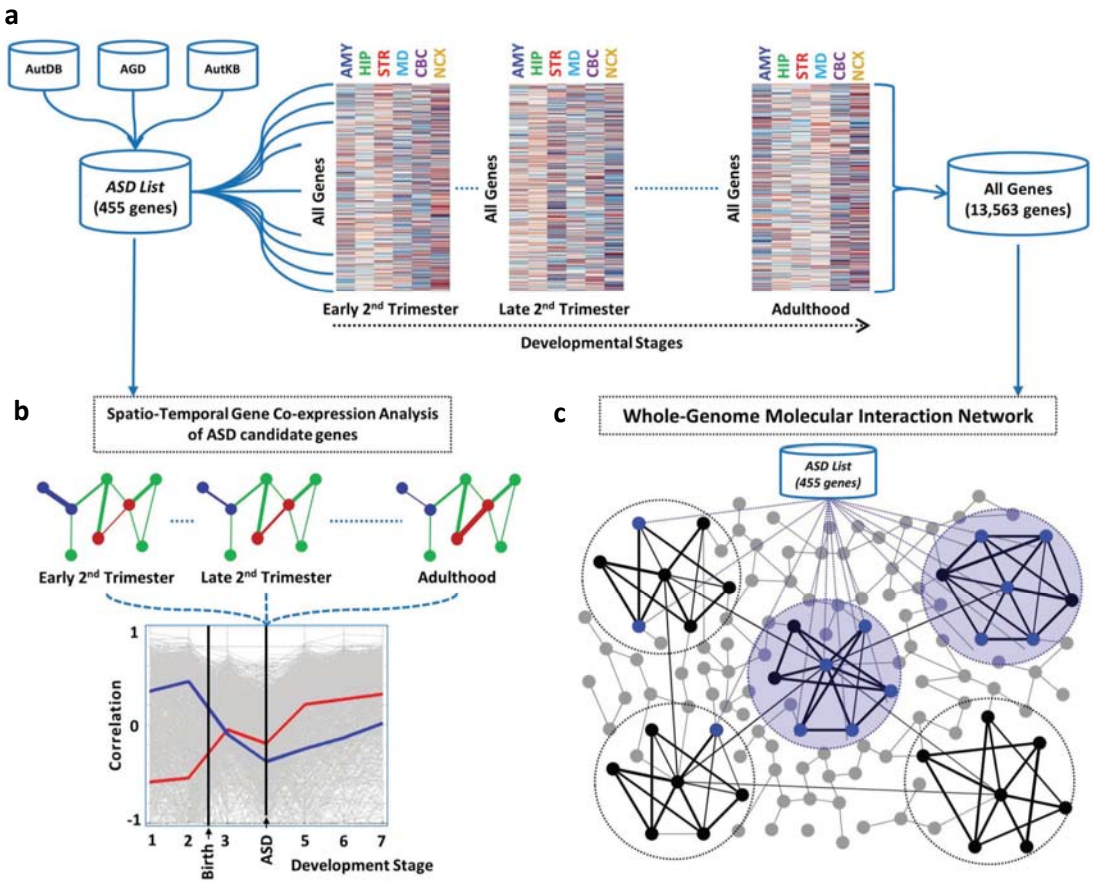


Figure 2.2.1. Graphical Representation of methodology used in this analysis. (a) Each heat-map shows the expression of all genes across six representative brain regions in three representative developmental stages. (b) A co-expression network of ASD candidate genes was generated for each developmental stage by correlating the expression vectors across brain regions. The blue gene-pair represents two genes that are moderately correlated at early developmental stages, but gain correlation through development. Stronger correlation is represented by a thicker edge between the two nodes. By contrast, the red gene-pair represents two genes that lose correlation over development. The lower panel shows the correlation patterns of all gene-pairs in the network (grey) across development. Correlation patterns of the blue & red pairs are shown in respective colors. Birth and the average age of ASD diagnosis are indicated. (c) The transcriptome-wide Molecular Interaction Network was constructed based on the pairwise correlation between each pair of genes expressed in the *BrainSpan Atlas* (13,563 genes). Each node in the network represents a gene while the weighted edges represent correlations between genes based on their expression across all samples. Nodes were clustered into modules (dashed circles). Genes from the *ASD list* are highlighted within each module (blue nodes). Blue circles indicate modules that are significantly enriched in genes from the *ASD list*.

2.2.4 Results

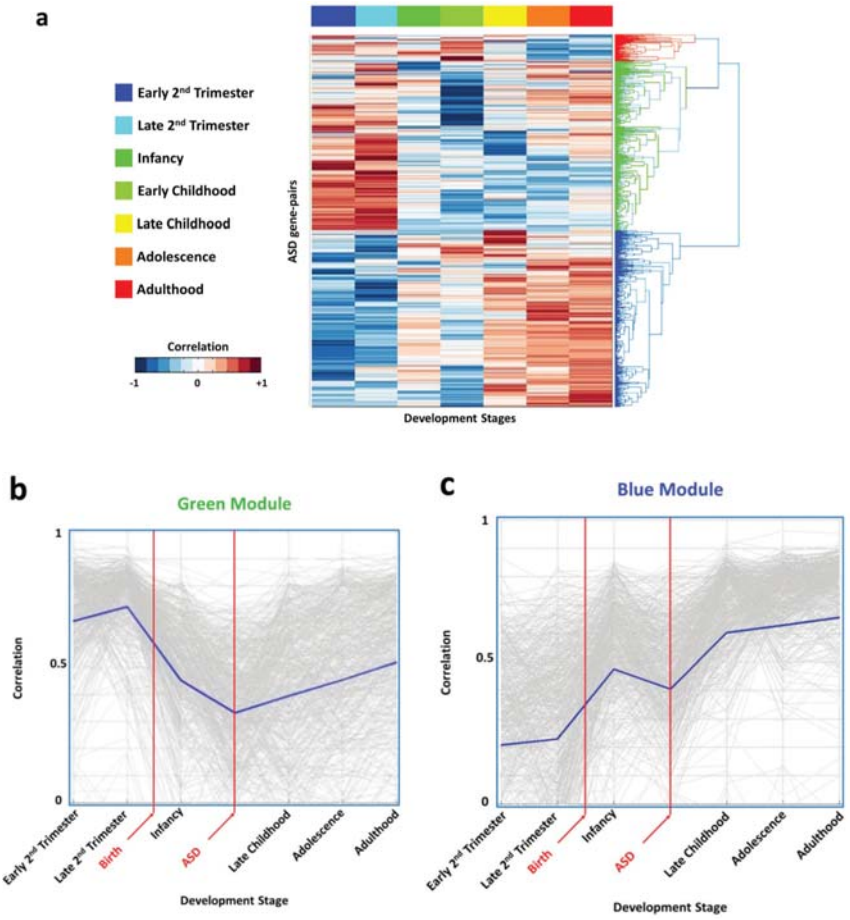
Part 1: Spatio-temporal Gene Co-expression Analysis of ASD Candidate Genes

ASD gene modules display distinct temporal dynamics around birth

Figure 2.2.2a shows the hierarchical clustering of the retained ASD gene-pairs. In total there were 103,285 pair-wise correlations between the 455 ASD candidate genes in the *ASD list*, of which 1,168 remained after applying the stringent threshold of an absolute correlation *greater than 0.8*. The surviving gene-pairs clustered into three distinct modules. Two of these modules, the "Green" module and the "Blue" module, displayed distinct correlation patterns relative to pre- versus post-natal development. The Green module (**Figure 2.2.2b**) consisted of gene-pairs that lose correlation in the middle stages of development (infancy and childhood); that is, each pair of genes within the Green module has highly correlated spatial expression profiles at prenatal developmental stages but this correlation is lost at birth. In contrast, the Blue module (**Figure 2.2.2c**) consisted of gene-pairs that gain correlation during development. These genes do not show correlation at prenatal stages but progressively increase correlation throughout postnatal development. The "Red" module did not show any coordinated pattern of expression over developmental time (not shown).

To characterize these modules further, we used the gene ontology enrichment analysis tool David 6.7 (39) to discover whether genes in these modules related to specific molecular mechanisms, cellular pathways or disease annotation terms. The top significantly enriched terms (Benjamini-corrected p-values < 0.01) are summarized as shown in **Figure 2.2.3**. All the three modules were enriched for annotation terms related to *neuron projection*, *synapse*, *synaptic transmission* and *behavior*. The three modules were also enriched for disease terms including *mental retardation* and *epilepsy*. The Green and Blue modules were significantly enriched for *neuron differentiation*, *cell morphogenesis,* and *learning/memory*. The Green module was specifically enriched in functional terms related to *regulation of apoptosis* and *regulation of cell death*, while the Blue module was specifically enriched in terms related to *ion channel*, *neurotransmitter receptor activity* and *GABA receptor activity*. **Supplementary Table S3** includes the full list of enriched gene-annotation terms for these two modules.

Figure 2.2.2. Spatio-temporal Gene Co-expression Analysis of ASD candidate genes. (a) Heat-map of the temporal correlation pattern of ASD gene-pairs (rows) through different developmental stages (columns). The dendrogram to the right shows the clustering of ASD gene-pairs into three modules (Red, Green and Blue). (b-c) The average correlation pattern of gene-pairs in the Green module (b) shows loss of correlation at childhood, whereas the average correlation pattern of gene-pairs in the Blue module (c) shows progressive gain of correlation across development.



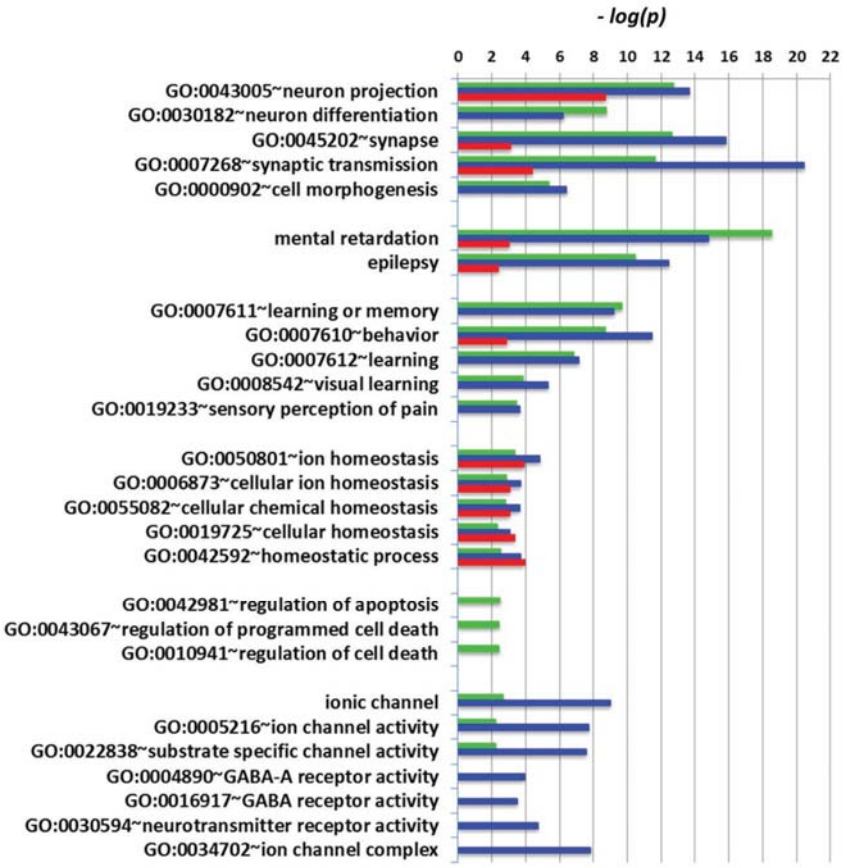
None of the GO terms that were significantly enriched in the three ASD modules showed any significant enrichment in modules from 10 randomly created sets. We also assessed how many gene-pairs remained after thresholding them on co-expression (absolute correlation > 0.8 at any developmental stage) in 10,000 random gene sets of 455 genes. These results showed that the number of gene-pairs remaining after thresholding the ASD list (1,168 gene-pairs) is highly significant ($p = 10^{-4}$).

Modules of ASD candidate genes are enriched in neurons

We then assessed if these modules were enriched in specific brain cell types. Lists of cell-type specific genes were obtained from a previously published work (Cahoy et al. 2008).

These lists included 1,465 neuron-, 1,529 oligodendrocyte-, and 1,829 astrocyte-specific genes. ASD candidate gene modules were assessed for enrichment of these cell types using the hypergeometric probability test (see Methods). Both the Green and Blue modules were significantly enriched in neurons, whereas the Red module demonstrated no significant enrichment, as shown in **Figure 2.2.4**.

Figure 2.2.3. Gene Ontology terms enriched in each of the three modules. Bars are colored according to the module's name. Data is presented as the $-\log_{10}$(p-value) with Benjamini-correction applied.



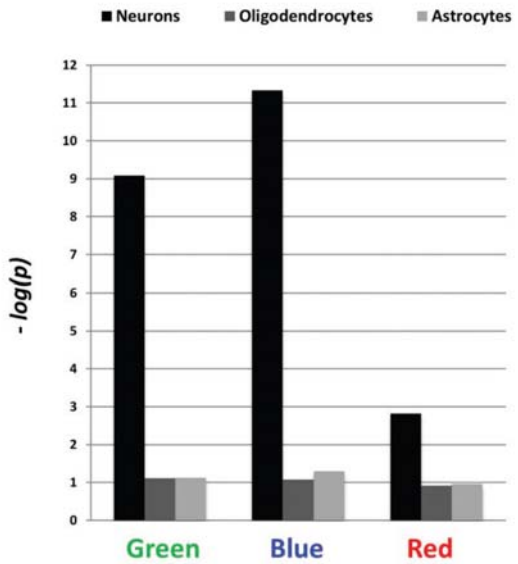


Figure 2.2.4. Enrichment scores for each of the ASD modules in neurons, astrocytes and oligodendrocytes. Data is presented as the $-\log_{10}$(p-value) with Benjamini-correction applied.

Part 2: Enrichment of ASD Candidate Genes in Transcriptome-wide Molecular Interaction Modules

Given the marked genetic heterogeneity of ASD and the large number of genes involved, it is also important to understand the role of ASD candidate genes in normal brain development within the context of the whole transcriptome, as sub-networks of the entire brain transcriptome may be perturbed by the ASD candidates. An analysis of these sub-networks could reveal ASD-related pathways that would be missed by analyzing the ASD candidates alone, as it is unlikely that all ASD candidate genes have been identified to date. Moreover, this top-down approach allows the identification of other genes that might also relate to ASD. Therefore, we performed a transcriptome-wide co-expression network analysis to identify functionally related gene modules throughout the normal developing brain transcriptome ('Molecular Interaction Modules'). Then, we assessed whether these modules were specific to distinct brain regions or developmental stages, and if they were related to specific pathways, cellular processes, or disease annotation terms. Finally, we determined if ASD candidate genes were enriched in any of the resultant Molecular Interaction Modules.

No evidence for region-specific modules

The transcriptome-wide co-expression network was constructed from all genes expressed in the brain (13,563 genes), based on their expression profile across all samples (480 samples, i.e. all brain structures and developmental stages). Genes were hierarchically clustered based on Spearman's rank correlation and complete linkage between pairs of genes. The resulting network consisted of 32 modules of varying size (from 36 to 1,386 genes), as shown in **Figure 2.2.5a**. Visual analysis of the heat-map and average expression patterns of member genes from each of the 32 modules demonstrated that none were specific to particular anatomical regions. This observation is consistent with the results from a similar dataset of human brain development assessed by microarray (Ben-David and Shifman 2012). We did not observe any pre/post natal specific expression patters in any of the 32 modules.

Modules enriched for ASD genes relate to synaptogenesis, protein turnover, and mitochondria

The resulting transcriptome-wide co-expression modules were then assessed for enrichment of genes belonging to the *ASD list* using the hypergeometric probability test. Four modules—Magenta, Brown, Orange, and Purple—were significantly enriched for ASD candidate genes (FDR-corrected p-values < 0.001), as shown in **Figure 2.2.5b**. The Magenta module (**Figure**

2.2.6a) contained genes highly co-expressed during early childhood. The Brown module (**Figure 2.2.6b**) included genes with low co-expression during childhood and differential spatial co-expression at late developmental stages. The Orange Module (**Figure 2.2.6c**) contained genes with progressively increasing co-expression during development. Finally, the Purple module (**Figure 2.2.6d**) included genes with varied co-expression during development and high differential spatial expression in adolescence and adulthood.

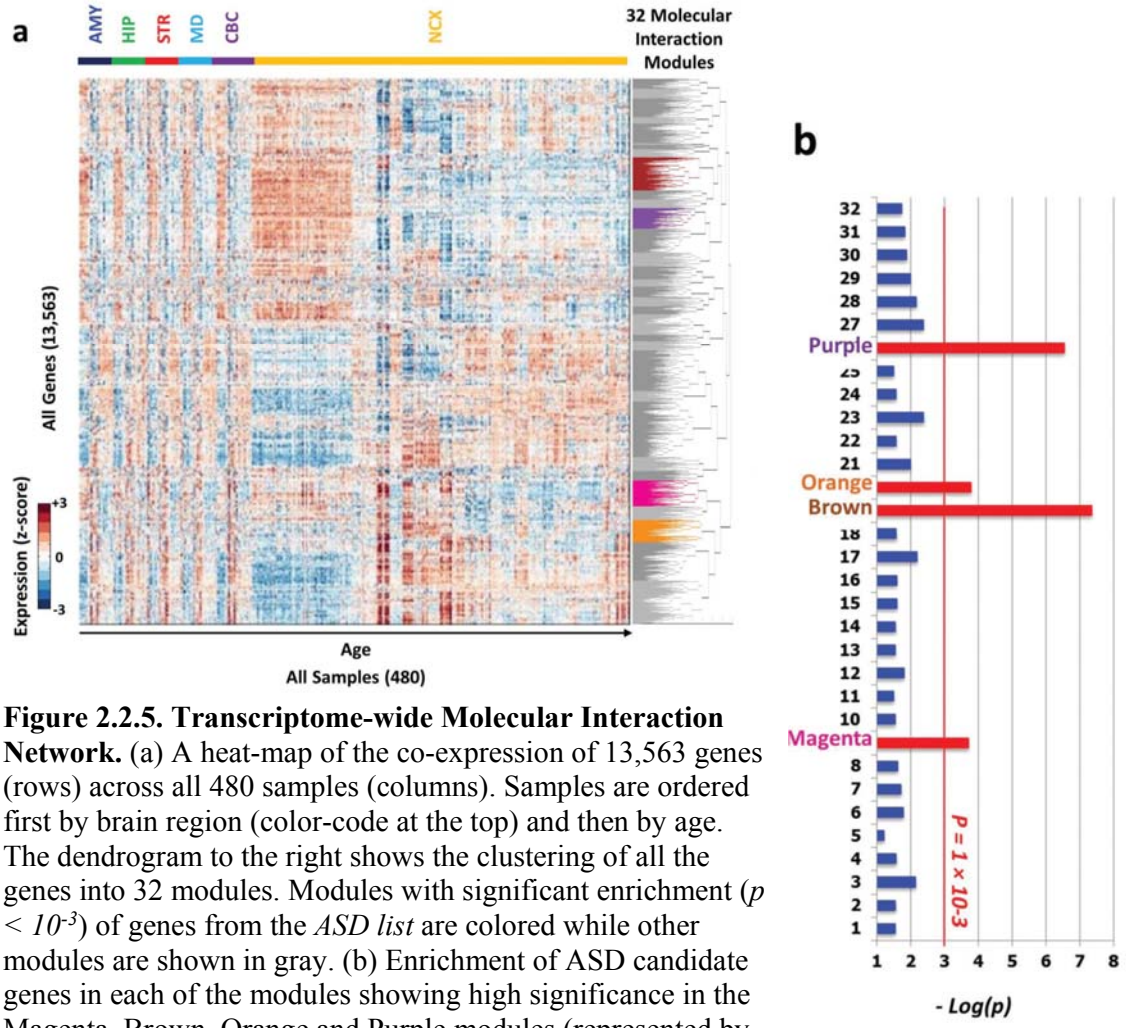


Figure 2.2.5. Transcriptome-wide Molecular Interaction Network. (a) A heat-map of the co-expression of 13,563 genes (rows) across all 480 samples (columns). Samples are ordered first by brain region (color-code at the top) and then by age. The dendrogram to the right shows the clustering of all the genes into 32 modules. Modules with significant enrichment ($p < 10^{-3}$) of genes from the *ASD list* are colored while other modules are shown in gray. (b) Enrichment of ASD candidate genes in each of the modules showing high significance in the Magenta, Brown, Orange and Purple modules (represented by $-log_{10}(p)$, *FDR-corrected*).

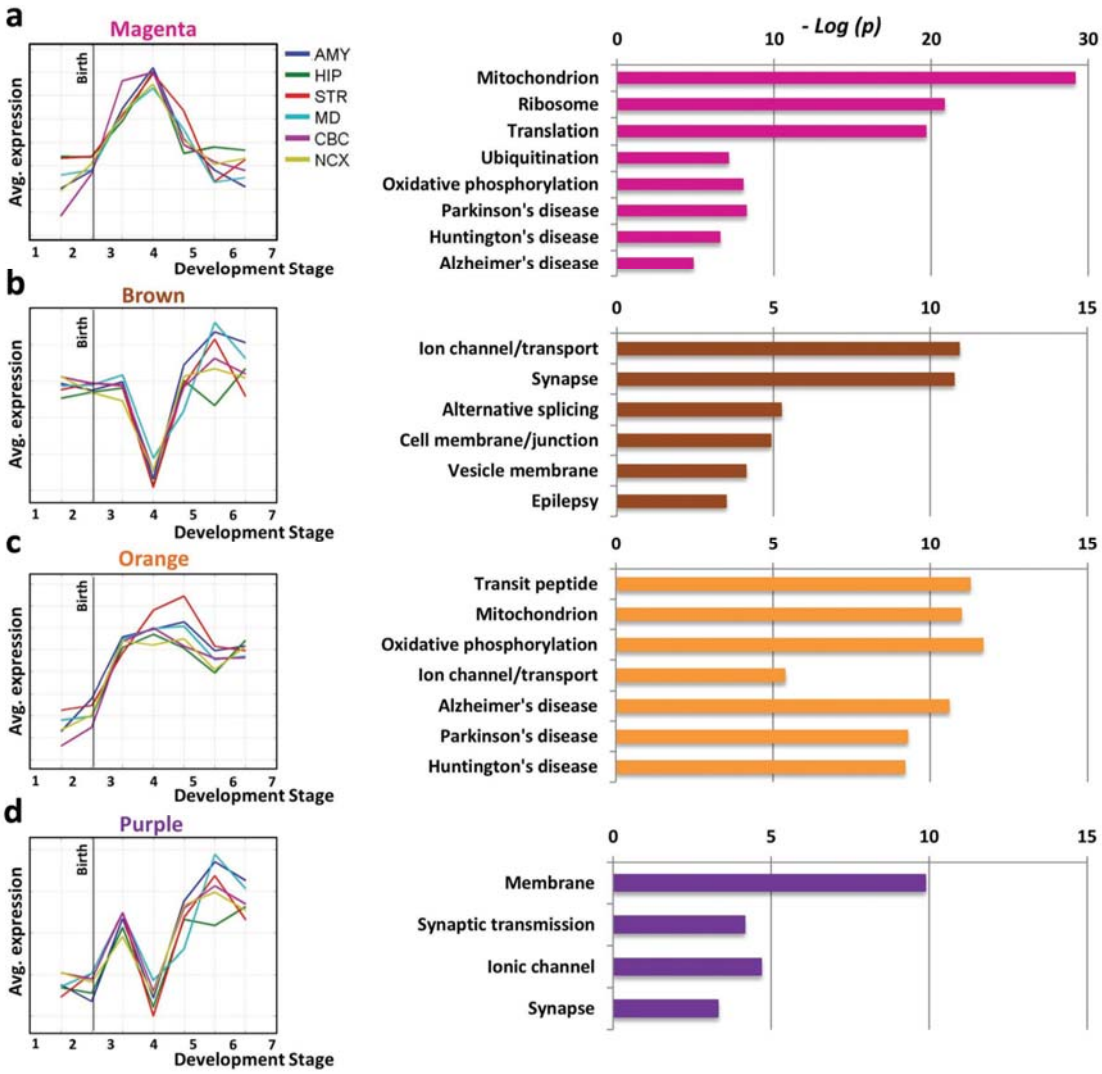


Figure 2.2.6. ASD Modules. (a) Left: Average co-expression pattern of the Magenta module genes across different brain regions (different plot colors). Right: Top GO terms enriched in the Magenta module. (b) Left: Average co-expression pattern of the Brown module genes. Right: Top GO terms enriched in the Brown module. (c) Left: co-Average expression pattern of the Orange Module genes. Right: Top GO terms enriched in the Orange module. (d) Left: Average expression pattern of the Purple module genes. Right: Top GO terms enriched in the Purple module. All enrichment values are represented by $-log_{10}(p)$, *Benjamini-corrected*.

Then, these ASD-enriched modules were tested for enrichment of gene ontology terms, as shown in **Figure 2.2.6 (**right panel**)**. The Magenta and Orange modules were significantly enriched for *mitochondrial* processes. Additional GO terms that were significantly enriched in the modules included *ribosome* and *protein translatio*n, *transit peptide*, *ubiquitination,* and *alternative splicing*. Significant enrichment for *synapse* was also found in the Brown module

and the Purple module. Enrichment of ASD candidate genes into transcriptome-wide synapse modules further supports our previous finding of ASD modules (Green and Blue modules), above, which were also related to synaptogenesis. Neurological disease terms were also significant in the ASD-enriched modules: *epilepsy* (Brown module), *Parkinson's* (Magenta and Orange modules), *Alzheimer's* (Magenta module and Orange modules) and *Huntington's* (Magenta and Orange modules).

ASD-enriched Molecular Interaction Modules are Mainly Neuronal

Each module was also tested for enrichment of specific neural cell populations (i.e. neurons, oligodendrocytes, and astrocytes), as described earlier. Three out of the four ASD-enriched modules were enriched for neurons (*Magenta, Brown and Purple modules*), as shown in **Figure 2.2.7**. The Orange module, which was related to mitochondrial functioning, was highly enriched in astrocytes but not neurons. This finding is of relevance, as multiple recent studies have implicated glia, and specifically astrocytes, in the brain pathology of autistic subjects (Cao et al. 2012; Lioy et al. 2011).

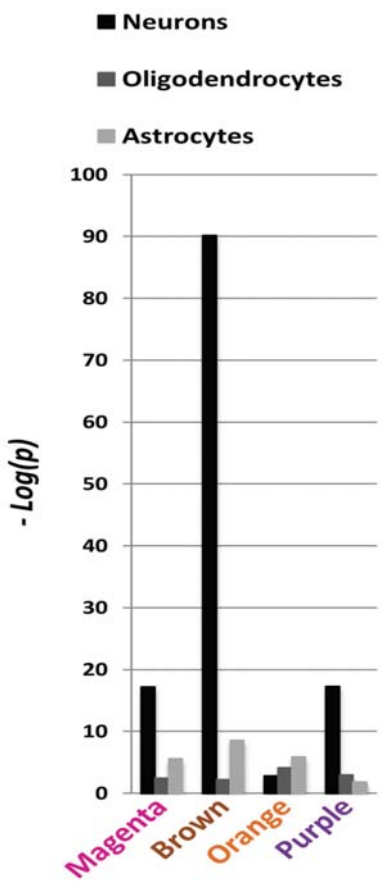


Figure 2.2.7. Enrichment of the ASD Modules in Cell-type Specific Genes. Enrichment of ASD-enriched modules in: neurons, oligodendrocytes and astrocytes (represented in $-log_{10}(p)$, *FDR-corrected*).

ASD-enriched Molecular Interaction Module hub genes provide molecular targets

An alternative approach to annotate the function of each ASD-enriched module is to analyze the genes with the strongest correlations within each module. It has been shown that within an interaction network, genes with the most connections to other genes, termed hub genes, are informative for the network as a whole, and are potential high yield therapeutic targets (Barabasi et al. 2011). The strongest correlations within a module were explored using Cytoscape v2.8 (Smoot et al. 2011). First, each ASD-enriched module (Magenta, Brown, Orange and Purple) was imported as a graph with genes acting as nodes and pair-wise correlations between genes representing edges between the nodes. **Figure 2.2.8** shows a subset of the connected nodes within each graph.

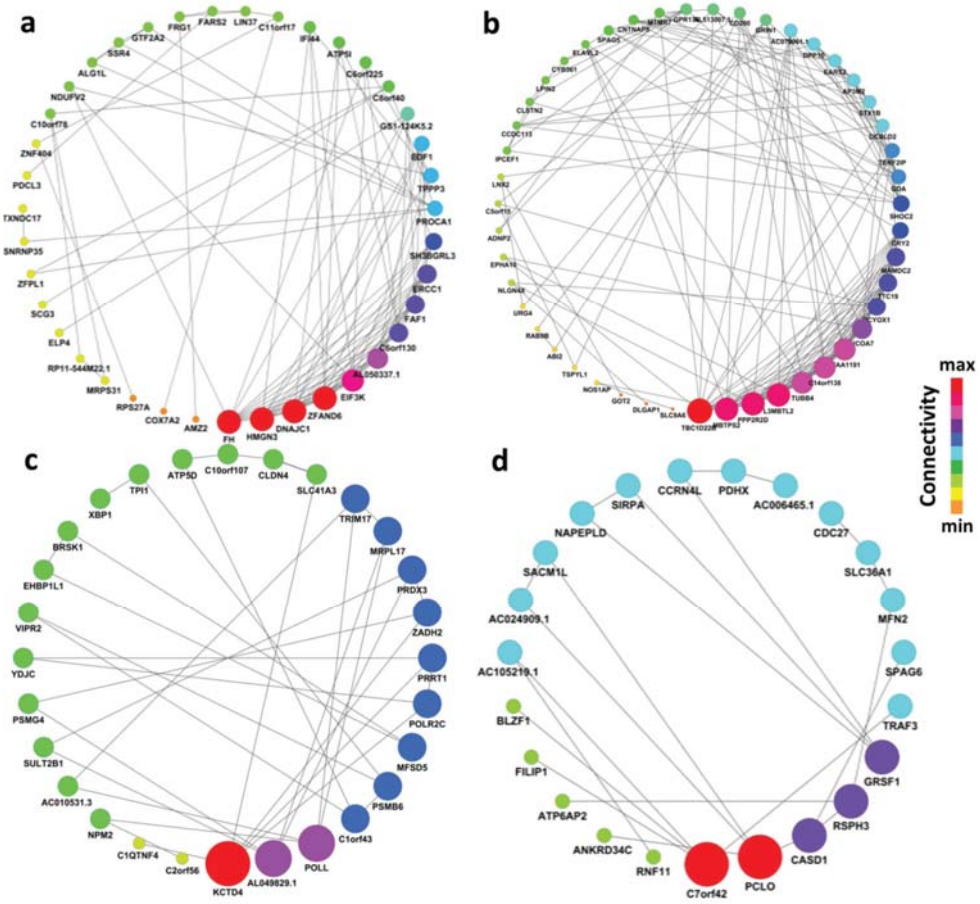


Figure 2.2.8. Hub Genes of ASD Modules. Each of the four ASD-enriched modules is presented with the Degree Sorted Circle layout of Cytoscape, with the nodes' size and color reflective of the level of connectivity within the network. For clarity, edges with correlation values smaller than 0.9 are removed. (a) Top connected genes of the Magenta module. (b) Top connected genes of the Brown module. (c) Top connected genes of the Orange module. (d) Top connected genes of the Purple module.

The 10 most highly connected nodes (genes) within each graph were extracted and their putative functions determined by manual curation of the literature. Among these most highly connected hub genes, a number were of note. The most striking observation was that most of the highly connected hub genes in the Magenta and Brown modules are known to function in the processes of *chromatin remodeling*, *transcription*, or *translation* (*HMGN3, EIF3K, ZFAND6, DNAJC1, C6orf130, ERCC1, LCMBT2, MBTPS2, KIAA1191, C14orf138, GDA,* and *NCOA7*). This result is in line with the gene ontology enrichment for these modules. A number of other central hub genes are involved in *intra-cellular signaling pathways* (*PROCA1, TBC1D22B, PPP2R2D, HACE1*), and a few are known to function as *membrane ion channels* (*PRRT1, KCTD4, SLC26A1, KCNA4*). In addition, a number of hub genes function in *apoptosis* or *myeloid/microglia* cell processes (such as *RNF11, CD200, and FAF1*). These hub gene functions largely recapitulate the ontologies of their respective networks, supporting our enrichment results and highlighting potential critical regulatory molecules of these networks.

2.2.5 Discussion

In order to gain insight into the molecular pathogenesis of ASD, we present a biologically-driven computational approach to analyze a heterogeneous set of genes previously independently implicated in ASD, to understand if they may relate to each other through shared functional genomics mechanisms. The main goal of this work is to understand if ASD candidate genes relate to common cellular/molecular pathways when considered in the context of transcription during normal human brain development. Identifying such pathways has profound implications for understanding the pathophysiology of ASD, especially since the majority of ASD patients do not have an identifiable genetic mutation (Huguet et al. 2013). Yet those patients without an identifiable mutation are still likely to have alterations in the same pathways, although the alterations may be caused by environmental, epigenetic, or other non-genetic factors.

We intentionally analyzed a very broad collection of genes associated with ASD, in an attempt to understand if there are cellular or molecular pathways that may represent final common mechanisms across all patients. Despite the fact that some of the genes in our ASD list are essentially causative for ASD (for instance, single gene mutation syndromes such as Fragile X), while others are not as strongly associated, we have weighted all genes equally to

avoid bias toward more severely-affected patient cases. Future work could attempt to weight genes differently within the co-expression networks to study different genetic subtypes of autism.

We discovered subsets of ASD candidate gene modules that displayed biologically-relevant co-expression dynamics, which were enriched for the processes of synaptogenesis, apoptosis, and GABA-ergic signaling. In addition, we assessed for functional genomic relationships between ASD candidate genes and the entire developing human brain transcriptome. This analysis revealed that ASD candidate genes are enriched within transcriptome-wide modules related to synaptogenesis, mitochondrial function, alternative splicing, protein translation, and ubiquitination. By identifying gene modules that have similar expression patterns in the brain (regardless of time period), we were able to infer that they are likely functioning in similar pathways. This allowed us to infer which cellular and molecular mechanisms are likely to be disrupted in autism. We also demonstrated the cell-type specific enrichment of these modules is mostly in neurons. Although several brain regions have been highlighted in neuroimaging and connectivity studies of autistic brains (namely cortical regions and the cerebellum; Carper and Courchesne 2005; Courchesne and Pierce 2005), interestingly, none of the transcriptome-wide modules were specific to particular anatomical regions, which supports previous reports of the BrainSpan dataset via microarray (Kang et al. 2011). Finally, by assessing genes with the highest connectivity within the transcriptome-wide Molecular Interaction Modules that were enriched for ASD candidates, we identified hub genes that may represent critical regulatory molecules in these networks, and their functions further supported our enrichment findings.

The number of strongly connected gene-pairs from the ASD list were found to be highly significant ($p = 10^{-4}$), indicating that based on their significantly strong co-expression across development, those ASD-associated genes are functionally related. We discovered three subsets of ASD–associated genes with distinct co-expression profiles around birth, even though the co-expression network for each developmental stage was calculated separately to avoid any bias towards pre/post natal expression changes. All three of these modules were significantly enriched for the processes of synaptogenesis and behavior, in addition to the disease annotations of mental retardation and epilepsy. Two of the modules (the Green and Blue modules) were also significantly associated with cell morphogenesis, neuron

differentiation, and learning. Moreover, the Green module, which had highly correlated spatial co-expression at prenatal developmental stages with a dramatic loss of correlation at birth, was uniquely enriched for the process of apoptosis. Conversely, the Blue module displayed an opposite co-expression trajectory—poor correlation in expression prior to birth, but strong co-expression beginning in infancy and increasing through adulthood—and was uniquely related to GABA-ergic signaling and ion channels. The distinct, biologically relevant expression patterns of these two modules around birth, a developmental period with the greatest shifts in gene expression (Kang et al. 2011), suggests a key role of these networks in brain development and autism.

A finding of particular interest was that ASD-associated genes were highly co-expressed later in development in some of the identified modules (childhood and adulthood), although autism symptoms are generally apparent by age 2. However, this does not preclude the possibility that the pathways implicated by these modules are involved in ASD pathogenesis, as our analysis was on co-expression patterns, not absolute gene expression levels. It is possible that the genes in these modules are still expressed in early neurodevelopment, but that they are most strongly co-expressed with other genes in the same module later in life. Consequently, disruption of the integrity of these genes (through inherited mutations, de novo mutations, mis-expression, etc.) early in development is likely to disrupt the functions of those modules later in life.

The functional ontologies of these networks are all pathways previously implicated ASD. Disrupted synaptogenesis has been one of the most replicated findings in ASD research (Bourgeron 2009), and autism is largely considered to be a disorder that results from a convergence of factors into synaptic dysfunction (Zoghbi 2003). Our finding of multiple ASD gene co-expression networks enriched for the function of synaptogenesis is in line with these previous studies. Additionally, our analysis shows these same transcriptional networks are related to the processes of GABA-ergic signaling and apoptosis, which have been independently associated with ASD through various approaches. GABA-ergic neurons are the main inhibitory cell of the brain, and much research has suggested that an imbalance in the ratio of inhibitory to excitatory neurons may underlie autism at the cellular circuit level (Rubenstein and Merzenich 2003). Furthermore, a number of clinical trials are currently ongoing to test GABA-ergic modulators for the treatment of ASD (Spooren et al. 2012).

Likewise, apoptosis—and more specifically the pruning of overabundant neural connections in early development—has recently been shown to be a critical process in the developing mammalian brain (Paolicelli et al. 2011), and a number of studies have suggested this process may be aberrant in ASD (Maezawa et al. 2011; Sheikh et al. 2010). A delicate balance between formation of needed synaptic connections and pruning of overabundant connectivity (and their excitatory/inhibitor ratio) is a main component of early experience-dependent brain development, and both human and animal studies have previously shown deficiencies in these processes in ASD (Courchesne and Pierce 2005). Our results suggest these processes may relate to each other and to ASD candidate genes through shared transcriptional networks.

ASD candidate gene modules with distinct temporal co-expression profiles around birth, which are highly related to synaptogenesis, support the notion that the pathogenesis of ASD is strongly related to this process. Additionally, the demonstration that the same transcriptional networks are also related to GABA-ergic signaling and apoptosis—both also suggested to be aberrant in autism—suggests that these disparate pathways may relate to each other through underlying shared transcriptional networks, providing a potential mechanism for functional convergence of ASD candidate genes into common pathways underlying autism.

By incorporating the ASD candidate genes into the context of the entire brain transcriptome, our results suggest that the disruption of synaptogenesis in autism may also relate to underlying basic cellular processes —alternative splicing, protein translation, and ubiquitination— which have previously been implicated in ASD (Glessner et al. 2009; Kelleher and Bear 2008; Piton et al. 2012; Smith and Sadee 2011). Defects in protein translation in particular have recently been shown to be a prominent feature in multiple animal models of ASD (Gkogkas et al. 2013; Neves-Pereira et al. 2009; Santini et al. 2013).

Two transcriptome-wide modules that were enriched for ASD candidate genes were both related to mitochondrial function. A large body of evidence has associated mitochondria dysfunction with rare syndromic forms of autism (Rossignol and Frye 2012) and recent evidence suggests that altered mitochondrial gene expression may contribute to non-syndromic autism as well (Anitha et al. 2012a; Anitha et al. 2012b). Furthermore, these modules were also related to Huntington's and Alzheimer's disease, both known to have

mitochondrial defects associated with their pathogenesis (Sheng and Cai 2012). While the ASD-only gene modules in the first part of this study did not implicate mitochondrial function, significant enrichment of ASD genes in two different transcriptome-wide networks related to mitochondria suggests that additional ASD genes related to mitochondria may remain to be discovered, and our hub gene analysis provides potential high confidence candidates.

A number of other studies have also assessed gene co-expression networks as relate to autism, although none has used a set of ASD candidate genes as broad nor have they used a comprehensive gene expression profile of human neurodevelopment, yet the findings of these studies are largely in agreement with those presented here. For instance, a study of autism candidate genes was performed using a subset of the genes assessed in this Chapter (the AutDB list) and they were assessed for co-expression modules in adult mice. The authors demonstrated that the ASD candidate genes were more highly co-expressed than random sets of genes—a finding similar to ours—and that there were two 'cliques' of highly co-expressed ASD candidates that corresponded the GO functions of synaptic transmission and ion transport (Menashe et al. 2013). They also found these cliques to be enriched in the cerebellum in particular. A similar study, which assessed two adult human brains in the Allen Atlas dataset for gene co-expression modules enriched for ASD candidates described in the AutDB database, found enrichment of ASD candidates in a neuron-expressed module with GO functions related to synaptogenesis and neural plasticity, and a second module associated with endocytosis (Ben-David and Shifman, 2013).

While the phenotype of autism may ultimately result from dysfunctional synaptogenesis, it is possible that such fundamental cellular processes as protein translation, ubiquitination, alternative splicing, and mitochondrial function may underlie the synaptic dysfunction. Furthermore, this may help explain the incredibly variable clinical spectrum of autism, and account for the increased prevalence of other complex medical problems in both the brain and other systems that ASD patients experience (Levy et al. 2009). Moreover, a recent meta-analysis of *de novo* mutations in autism demonstrated enrichment for genes related to transcriptional regulation, and showed they have similar neurodevelopmental expression patterns to the Green and Blue modules of ASD candidates we identified (Ben-David and Shifman 2012). Whether and how defects in these basic cellular mechanisms result in altered

synaptogenesis, are a reaction to altered synaptogenesis, or are mutually- exclusive from synaptogenesis is unclear. However, our results suggest that a complex interplay between these processes and synaptogenesis are related to each other through overlapping co-expression networks.

A number of studies have assessed for changes in gene expression in post-mortem autistic brain directly (Lintas et al. 2012; Voineagu 2012). These studies have repeatedly shown that the autistic transcriptome is abnormally expressed compared to control brains across many different brain regions. The genes that are mis-expressed in autistic brains have been consistently demonstrated to be involved in pathways related to the synapse (Chow et al. 2012; Voineagu et al. 2011), immune response/apoptosis (Chow et al. 2012, Voineagu et al. 2011; Garbett et al. 2008), neurotransmitter receptors (Purcell et al. 2011), RNA splicing (Chow et al. 2012; Voineagu et al. 2011; Ziats and Rennert 2013), and mitochondrial function (Anitha et al. 2012b; Smith et al. 2012). These findings in autistic brain complement our results by showing that the ASD co-expression modules we discovered in the *normal* developing brain are functioning in the same pathways that are consistently disrupted in autistic brains.

Finally, the identified hub genes of ASD-enriched modules recapitulate the gene ontology analysis of these modules, strengthening the observation that basic cellular functions related to genome processing and mitochondrial function may represent a nexus in the genomic pathology of ASD. In addition, a number of hub genes relate to myeloid cells and apoptosis. There is a growing body of evidence implicating cytokine signaling, microglia-mediated synaptic pruning, and other immune-related processes in ASD (Maezawa et al. 2011), and this finding suggests the autism candidate genes may indirectly relate to processes that interact with these pathways through the transcriptional machinery. Furthermore, this supports our finding that the Green module of autism candidate genes relates to apoptotic pathways. However, because comprehensive lists of microglia-specific marker genes are not available, we were unable to assess for enrichment of ASD candidate genes into this cell type in this study. By highlighting individual genes that are most central in the identified molecular interaction networks, the hub gene analysis may provide potential additional high-yield ASD candidates for their respective transcriptional networks.

2.2.6 Conclusion

In summary, we have profiled the transcriptional co-expression networks of autism candidate genes throughout normal human brain development to identify modules of ASD candidate genes with biologically-relevant expression patterns. We have shown that these ASD modules are enriched for synaptogenesis, apoptosis, and GABA-ergic signaling, suggesting that pathways previously independently implicated in autism are related to each other through shared neurodevelopmental transcriptional networks. In addition, we expanded the analysis of ASD candidates to consider their relationship with the entire brain transcriptome. We demonstrated that ASD-enriched transcriptome-wide Molecular Interaction Modules are related to mitochondrial function, splicing, and protein turnover, which suggests further ASD candidates related to these functions may remain to be discovered.

Our comprehensive analysis of the global co-expression relationships between ASD candidates demonstrates that the various pathways implicated in autism separately may relate to one another when considered in a broader functional genomics framework. Furthermore, our Molecular Interaction Module analysis represents a valuable strategy to identify and prioritize other potential ASD candidate genes. Moreover, this approach can be used to assess genes from other complex neurodevelopmental and psychiatric disorders like schizophrenia, to uncover potential overlapping transcriptional pathways in the developing human brain among other gene sets.

--

2.3 Global Sex Differences in Gene Expression

2.3.1 Aim

Autism spectrum disorders affect significantly more males than females. Understanding sex differences in normal human brain development may provide insight into the mechanism(s) underlying this disparity; however, studies of sex differences in brain development at the genomic level are lacking. Here, I report an analysis of sex-specific gene expression from a recent large transcriptomic study of normal human brain development, to determine whether sex-biased genes relate to specific mechanistic processes. I discovered that male-biased genes are enriched for the processes of extracellular matrix formation/glycoproteins, immune response, chromatin, and cell cytoskeleton. I highlight that these pathways have been repeatedly implicated in autism and demonstrate that autism candidate genes are also enriched for these pathways. I propose that the overlap of these male-specific brain transcriptional modules with the same pathways in autism spectrum disorders may partially explain the increased incidence of autism in males.

2.3.2 Introduction

Among the varied genetic, cellular, and clinical phenotypes described in autism spectrum disorders, the predominance of males versus females affected has stood apart as one of the most replicated findings. The most recent analysis indicates that at least four males are affected for every one female (Centers for Disease Control and Prevention 2012), and among the least severely affected children with ASD, the male-to-female ratio is even higher (Gilberg et al. 2006). Multiple hypothesis have been put forth to explain this phenomena, including the notion that autistic children have brains that are more "masculine" (Baron-Cohen 2002; Baron-Cohen et al. 2005; Baron-Cohen et al. 2011). Implicit in such theories is the idea that by understanding the mechanism(s) contributing to the sex disparity of ASD, a better understanding of the underlying pathophysiology can be realized. While sex differences in ASD have mainly been studied at the behavioral level, molecular evidence that the autistic brain is somehow biased toward a male pattern of development is lacking.

Therefore, I was interested in assessing sex-biased patterns of normal human brain development at the genomic level, in an attempt to gain insight into processes that may underlie sex differences in ASD.

2.3.3 Methods

I re-analyzed gene expression data from a recent whole-genome transcriptomics study of normal human brain development by Kang and colleagues (Kang et al. 2011). This study performed genome-wide microarray analysis on unremarkable post-mortem human brain tissue from 16 brain regions spanning preconception to adulthood. While the Kang et al. study identified genes with male and female-specific patterns of expression and highlight individual genes of interest, a global assessment of sex-specific gene expression differences was not undertaken. Therefore, I assessed male versus female gene expression differences in aggregate to determine if common biological pathways were over-represented. I used the gene ontology function DAVID 6.7 (Huang et al. 2009) to analyze all sex-biased genes in aggregate (i.e. male versus female differentially expressed genes from all brain regions and developmental time points; **Supplementary Table S15**). Additionally, I repeated this analysis using two other gene ontology databases: Ingenuity Pathways Analysis and GeneGO. Only functional ontologies with a Benjamini-Hochberg multiple testing correction p-value < 0.05 were considered significant.

2.3.4 Results

The genes with female-biased expression patterns were not significantly enriched for any particular functional categories. However, genes with male-biased patterns of expression were enriched for the processes of: glycoproteins/extra-cellular matrix (ECM), immune response, nucleosome/chromatin, and cell cytoskeleton (**Table 2.3.1**). Repeating the analysis using both Ingenuity Pathways Analysis and GeneGO yielded similar results (**Supplementary Tables S16-S19**).

This finding is incredibly interesting in light of previous studies of autistic brain, as these pathways are consistently implicated in ASD. For instance, gene-expression studies from post-mortem autism brain tissue have demonstrated over-expression of immune response pathways in ASD (Voineagu et al. 2011; Garbett et al. 2008). Furthermore, both genetic and cellular studies have suggested synaptogenesis is impaired in autism through cell adhesion and cell-ECM binding pathways (Betancur et al. 2009). Moreover, recent large sequencing studies repeatedly identified a chromatin-remodeling gene as one of only a few reaching genome-wide significance (Neale et al. 2012; O'Roak et al. 2012). Interestingly, however, when I compared the list of genes with male biased expression to those aberrantly expressed in ASD brain (Voineagu et al. 2011) or otherwise implicated in ASD (Xu et al. 2012), no

genes overlapped. This suggests that perhaps individual genes themselves do not relate normal sexually dimorphic brain development to sex differences in ASD incidence, but rather suites of genes that funnel into common sex-specific brain development pathways may become perturbed in ASD.

To test this hypothesis, I also performed GO enrichment analysis on autism candidate genes, to determine if these male sex-biased pathways overlap with those implicated by ASD candidates. I assessed in aggregate putative autism genes curated in the AutKB database (both 'syndromic' and 'non-syndromic' sets; Xu et al. 2012). I found that all of the male sex-biased pathways were also implicated by the ASD candidate genes (**Supplementary Table S20**). Taken together, these results demonstrate that pathways—but not the individual genes—implicated in autism overlap with normal male-specific modules of the developing brain. Therefore, I propose that shared transcriptional modules, which influence both normal male brain development and the pathogenesis of ASD, may partially explain the increased incidence of autism in males.

Of unique interest is the finding of an immune system module in normal male-specific brain transcription. While peripheral immune responses have a significant gender dimorphism (Shames 2002), the role of immune activation in autism spectrum disorders and its relation to sex differences has remained unclear. Based on a lack of enrichment for autism susceptibility genes in immune-related whole-genome co-expression networks that derived from genes differentially expressed in the autistic brain, Voineagu et al. concluded that immune findings in autistic brain are environmental rather than genetic (Voineagu et al. 2011). However, my results suggest that a transcriptional program utilizing immune system components somehow may contribute to making a normal human brain more "male." To further explore my results in this context, I parsed the autism susceptibility genes into those that had been identified on the basis of microarray or proteomics studies ('expression set') versus those that had been identified via studies assessing inherited or *de novo* DNA mutations ('inherited set'), such as genome-wide association, copy number variation, linkage, and other association analysis (**Supplementary Table S21**). This showed that immune-related modules were only significant in the expression set, but not the inherited set, of ASD candidate genes (**Supplementary Tables S22 and S23**). It is intriguing to speculate then, that while disruption of immune pathways in ASD may not be inherited, as Voineagu *et al* propose and these results support, male brains may still be more susceptible to environmental insults

because of their reliance on these immune-related pathways for normal male-specific brain development.

2.3.5 Conclusion

Overall, these results suggest the hypothesis that sex-specific transcriptional modules may make males more susceptible to neurodevelopmental disorders that result from aberrations—both inherited and environmental—in these pathways. As these four particular pathways are repeatedly implicated in ASD and overlap with pathways implicated by ASD candidate genes, I propose that normal sex-differences in the functional genomics underlying human brain development may partly explain the significantly higher incidence of ASD in males. Notably, if these four modules are more prominent in normal male versus female brain development, than any disorder resulting from aberrations in these pathways would be expected to have increased incidence in males. It will be important for future studies to further assess the role of male-specific gene expression programs as they relate to co-expression, dysregulation, and interaction with autism-candidate genes in the developing brain, in addition to those of other neurodevelopmental and psychiatric disorders with sex-biased incidence.

Table 2.3.1. List of all significant gene ontology results from analysis of male genes.

Annotation Cluster 1	Enrichment Score: 2.89		
Category	Term	Raw PValue	Benjamini Corrected Pvalue
SP_PIR_KEYWORDS	glycoprotein	2.51E-09	6.62E-07
SP_PIR_KEYWORDS	signal	6.00E-08	7.92E-06
UP_SEQ_FEATURE	signal peptide	7.24E-08	3.60E-05
UP_SEQ_FEATURE	glycosylation site:N-linked (GlcNAc...)	4.79E-07	1.19E-04
SP_PIR_KEYWORDS	Secreted	1.20E-06	1.06E-04
SP_PIR_KEYWORDS	extracellular matrix	1.59E-06	1.05E-04
SP_PIR_KEYWORDS	disulfide bond	2.37E-06	1.25E-04
GOTERM_CC_FAT	GO:0005578: extracellular matrix	5.23E-06	9.05E-04
UP_SEQ_FEATURE	disulfide bond	8.91E-06	1.48E-03
GOTERM_CC_FAT	GO:0031012: extracellular matrix	1.12E-05	6.47E-04
GOTERM_CC_FAT	GO:0005583: fibrillar collagen	8.87E-05	3.06E-03
GOTERM_BP_FAT	GO:0030199: collagen fibril organization	5.14E-05	1.00E-02
SP_PIR_KEYWORDS	Skin	3.38E-04	9.87E-03
PIR_SUPERFAMILY	PIRSF002255: collagen alpha 1(I) chain	6.80E-04	4.65E-02
GOTERM_CC_FAT	GO:0005576: extracellular region	7.36E-04	1.80E-02
GOTERM_CC_FAT	GO:0044421: extracellular region part	8.24E-04	1.57E-02
SP_PIR_KEYWORDS	Ehlers-Danlos syndrome	1.82E-03	3.63E-02
GOTERM_CC_FAT	GO:0044420: extracellular matrix part	1.95E-03	3.31E-02
GOTERM_CC_FAT	GO:0005581: collagen	2.33E-03	3.60E-02
SP_PIR_KEYWORDS	collagen	2.38E-03	4.40E-02
Annotation Cluster 2	Enrichment Score: 2.79		
Category	Term	Raw PValue	Benjamini Corrected Pvalue
GOTERM_BP_FAT	GO:0009611: response to wounding	1.01E-04	1.23E-02
GOTERM_BP_FAT	GO:0006952: defense response	1.46E-03	6.90E-02
GOTERM_BP_FAT	GO:0006954: inflammatory response	2.93E-02	4.70E-01
Annotation Cluster 3	Enrichment Score: 2.78		
Category	Term	Raw PValue	Benjamini Corrected Pvalue
GOTERM_CC_FAT	GO:0000786: nucleosome	7.83E-06	6.77E-04
SP_PIR_KEYWORDS	nucleosome core	8.96E-06	3.94E-04
INTERPRO	IPR007125: Histone core	1.08E-05	3.11E-03
GOTERM_BP_FAT	GO:0006334: nucleosome assembly	3.01E-05	2.91E-02
GOTERM_BP_FAT	GO:0031497: chromatin assembly	3.68E-05	1.20E-02
GOTERM_CC_FAT	GO:0032993: protein-DNA complex	4.71E-05	2.04E-03
GOTERM_BP_FAT	GO:0065004: protein-DNA complex assembly	4.74E-05	1.16E-02
GOTERM_BP_FAT	GO:0034728: nucleosome organization	5.36E-05	8.74E-03
GOTERM_BP_FAT	GO:0006323: DNA packaging	1.92E-04	1.86E-02
SP_PIR_KEYWORDS	chromosomal protein	2.16E-04	7.09E-03
KEGG_PATHWAY	hsa05322: Systemic lupus erythematosus	2.63E-04	1.28E-02
GOTERM_BP_FAT	GO:0006333: chromatin assembly	2.99E-04	2.42E-02
GOTERM_BP_FAT	GO:0043933: macromolecular complex	5.16E-04	3.56E-02
SP_PIR_KEYWORDS	isopeptide bond	5.94E-04	1.56E-02

GOTERM_CC_FAT	GO:0000785: chromatin	8.09E-04	1.74E-02
Annotation Cluster 4	Enrichment Score: 2.55		
Category	Term	Raw PValue	Benjamini Corrected Pvalue
GOTERM_BP_FAT	GO:0030199: collagen fibril organization	5.14E-05	1.00E-02
GOTERM_BP_FAT	GO:0032964: collagen biosynthetic process	5.02E-04	3.72E-02
Annotation Cluster 6	Enrichment Score: 2.25		
Category	Term	Raw PValue	Benjamini Corrected Pvalue
GOTERM_BP_FAT	GO:0030036: actin cytoskeleton organization	3.45E-05	1.68E-02
GOTERM_BP_FAT	GO:0030029: actin filament-based process	5.69E-05	7.96E-03
GOTERM_BP_FAT	GO:0007015: actin filament organization	1.64E-04	1.77E-02
GOTERM_CC_FAT	GO:0015629: actin cytoskeleton	2.01E-04	5.78E-03
SP_PIR_KEYWORDS	actin binding	1.66E-03	3.60E-02

--

2.4 Identification of differentially expressed miRNAs and their Relation to ASD

2.4.1 Aim

Non-coding RNAs, and miRNAs in particular, are known to play a role in modulating gene expression, and have been circumstantially implicated in ASD. However, a comprehensive assessment of miRNA expression in the developing human brain has not been performed, hampering attempts to both understand broad patterns of miRNA-mediated gene expression regulation during neurodevelopment, and to explore individual miRNAs of interest to neurodevelopmental disorders like ASD. Therefore, in this chapter I describe my results of the most comprehensive assessment of broad miRNA expression patterns during human neurodevelopment to date, and their relation to neurodevelopmental disorders. Finally, I assessed specific miRNAs of interest to syndromic causes of ASD to gain father insight into the potential role miRNAs may play in the pathogenesis of ASD functional genomics.

2.4.2 Introduction

Human neurodevelopment requires coordinated expression of thousands of genes, exquisitely regulated in both spatial and temporal dimensions, to achieve the proper specialization and inter-connectivity of brain regions. Consequently, the dysregulation of complex gene networks in the developing brain is thought to underlie many neurodevelopmental and psychiatric disorders (Oldham et al. 2008). In order to understand these pathologic gene expression changes, it is critical to achieve a comprehensive understanding of normal gene expression *regulation* throughout human neurodevelopment. While broad surveys of gene expression across the developing human brain have recently been described (Kang et al. 2011), the molecular regulators of this gene expression—most notably microRNAs—have only been assessed in a few brain regions or developmental periods (Shao et al. 2010; Hu et al. 2011; Somel et al. 2010; Somel et al. 2011). As microRNAs are increasingly recognized as fundamental to brain developmental processes and neurologic diseases (Qureshi et al. 2012), a comprehensive understanding of their expression dynamics throughout human brain development is important.

Therefore, I analyzed the differential expression of all microRNAs (miRNAs) detected by RNA-sequencing of 82 neurologically-normal post-mortem human brain tissue samples, which derived from 18 individual donor brains spanning 4 months through 19 years of age (see Methods). Donor samples were grouped into four developmental time windows (infancy, early childhood, late childhood, and adolescence, **Table 2.4.1**). Six distinct brain regions were assessed: four regions of the prefrontal cortex, the hippocampus, and the cerebellum. I also assessed for differential miRNA expression between males and females in the prefrontal cortex. Then, I identified putative gene targets of the differentially expressed miRNAs, determined if these gene targets were enriched for particular functional processes, and assessed if the identified targets were enriched for genes associated with common neurodevelopmental, psychiatric, and neurodegenerative diseases. Finally, I explored in depth the potential miRNA-mediated regulation of three genes that are known to be causative of syndromes with ASD as a major component. The results presented here further implicate miRNAs in the functional genomics of ASD and other neurodevelopmental disorders.

Table 2.4.1. Developmental periods and average number of donor tissue samples assessed.

Developmental Period	Ages	Avg. Samples per Region
Infancy	4 months – 1 year	3.5
Early Childhood	2 – 4 years	3.0
Late Childhood	8 – 13 years	2.8
Adolescence	15 – 23 years	4.3

2.4.3 Methods

miRNA Data and Pre-processing

Data was downloaded at: download.alleninstitute.org/ brainspan/MicroRNA. The full dataset contained 1620 miRNAs measured across 215 brain samples. Only brain samples originating from the orbitofrontal prefrontal cortex (OFC; Brodmann's Area (BA) 11), dorsolateral prefrontal cortex (DFC; BA 9, 46), medial prefrontal cortex (MFC; BA 32, 33, 34), ventrolateral prefrontal cortex (VFC; BA 44,45), hippocampus (HIP), or cerebellum (CER) were retained. For analysis between sexes, brain regions were aggregated from the prefrontal cortex samples. Next, miRNAs with read counts likely to be noise rather than true reads were removed, in order to increase subsequent statistical power; this has been demonstrated not to affect the dispersion model used to calculate differential expression (Anders and Huber 2010). Importantly, I did this prior to any analysis of the data. To do so, the sum total of

read counts for each miRNA across all 82 samples was calculated. miRNAs with zero total reads were immediately discarded (58 miRNAs). Next, miRNAs were ordered from most to least reads (range 1 to 41,540,463) and the dispersion of read counts was plotted for visualization. Then, any miRNA with a total read count less than 60 was discarded, resulting in 902 retained miRNAs (**Supplementary Table S24**).

miRNA Differential Expression Analysis

Differentially expressed miRNAs were discovered using the edgeR package (Robinson et al. 2010) run in the R programming environment. The edgeR user guide was followed as detailed in the "classic analysis" section. This method was chosen to evaluate differential expression because its performance is intermediately conservative among various RNA-seq analysis packages (Robles et al. 2012). miRNAs were considered to be significantly differentially expressed between groups only if the false discovery rate (FDR) p-value was < 0.05 and the absolute $\log_2$ of Fold Change (FC) was > 1.5. Differentially expressed miRNAs were identified across three dimensions: spatial, temporal, and by sex. Spatial miRNAs were differentially expressed between two anatomic brain regions within one developmental time period. Temporal miRNAs were differentially expressed over developmental time within the one anatomic brain region. Sex-biased miRNAs were differentially expressed between male and female prefrontal cortex samples within one time period (data was combined from all four prefrontal cortex regions).

Downstream analysis of miRNA targets

miRNAs that were determined to be differentially expressed temporally or by sex were further analyzed for putative target genes under their control. To do so, I used the target prediction algorithms of TargetScanHuman 6.2 (Lewis et al. 2005) and miRDB (Wang 2008). I considered as significant only those targets that were predicted by both algorithms. Gene ontology (GO) enrichment analysis of the target genes was performed using DAVID Bioinformatics Resources 6.7 Functional Annotation Tool (Huang et al. 2009). Gene ontologies were considered significant only if their Benjamini-Hochberg multiple testing corrected p-value was < 0.05. GO enrichment analysis was performed on lists of aggregated targets (all time periods) that were brain region specific.

Test for enrichment of disease-associated genes

To determine if the target genes of differentially expressed miRNAs may relate to neurological diseases, I assessed for their enrichment into disease-related gene sets. Disease related gene sets were downloaded from the Genotator database (Wall et al. 2010). Enrichment was tested using the Hypergeometric probability distribution function in Excel. The population universe (i.e. all protein-coding genes in the human genome) was set to 20,687 (Dunham et al. 2012). A success in the Hypergeometric function test was a gene that was both a predicted miRNA target and previously associated with a disorder. P-values were corrected for multiple testing by applying the conservative Bonferorri method. Enrichment was only considered significant if the Bonferorri-corrected p-value was < 0.01.

2.4.4 Results

In total, I discovered 75 miRNAs differentially expressed across developmental time *within* brain regions (absolute $\log_2$ fold change > 1.5 and FDR < 0.05, **Supplementary Table S25**). Similar to previously described changes in gene expression, the greatest differential expression of miRNAs occurred during the transition from infancy to early childhood (**Figure 2.4.1**). The dorsolateral prefrontal cortex exhibited the greatest number of differentially expressed miRNAs (35 miRNAs) and the cerebellum a similar amount (22 miRNAs); the hippocampus and other regions of the prefrontal cortex each displayed less than five differentially expressed miRNAs. In contrast, differential expression of miRNAs *between* brain regions increased over developmental time (**Figure 2.4.2**, **Supplementary Table S26**). This finding is opposite previously described patterns of mRNA expression, which has been shown to become more globally similar between brain regions over development (Kang et al. 2011).

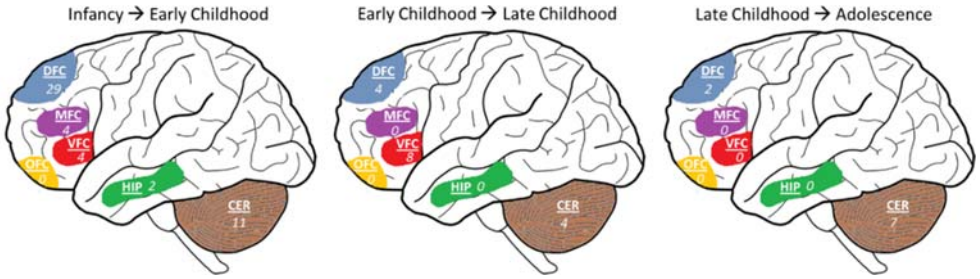


Figure 2.4.1. Number of differentially expressed miRNAs within each region over development.

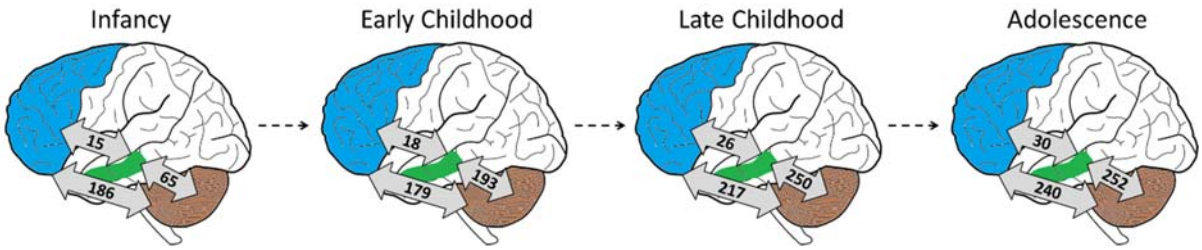


Figure 2.4.2. Number of differentially expressed miRNAs between brain regions over development. Light blue represents the prefrontal cortex, green represents the hippocampus, and the cerebellum is shaded in brown.

As many neurodevelopmental disorders display a significant sex-bias in their prevalence (Jazin et al. 2010; Abel et al. 2010; Werling and Geschwind 2013), I also assessed for differential miRNA expression by sex in the prefrontal cortex. I discovered 40 miRNAs with significant sex-biased expression differences between the prefrontal cortex of males and females (**Table 2.4.2**, **Supplementary Table S27**). Strikingly, 93% were more highly expressed in females, again a trend opposite to that of sex-biased gene expression (Kang et al. 2011). Furthermore, the majority of sex-biased miRNA expression occurred in adolescence (65%), suggesting that miRNA-targeted gene expression differences in the prefrontal cortex of males versus females becomes most pronounced around puberty.

Table 2.4.2. Differentially expressed miRNAs between male and female prefrontal cortex over development.

	Up-regulated in Males	Up-regulated in Females	Total
Infancy	1	1	2
Early Childhood	1	9	10
Late Childhood	0	2	2
Adolescence	1	25	26
Total	3	37	**40**

To explore the potential biologic and pathogenic roles of the differentially expressed miRNAs, I identified putative targets of the temporally and sex-biased differentially expressed miRNAs. I then assessed for enrichment of gene ontology categories in all lists of putative target genes. Overall, miRNA target genes were highly related to the process of transcription regulation in almost all lists (**Supplementary Tables S28-S32**). This finding is in line with the well-known function of miRNAs as master regulators of gene expression networks (Chen et al. 2007; Hobert 2008), and underscores the importance of identifying these key hubs of brain transcriptomes. Additionally, putative gene target lists were enriched

for biological processes relating to nervous system development, synaptogenesis, and other basic intracellular processes.

Of particular note was the functional enrichment of miRNA targets that were differentially expressed between male and female prefrontal cortex. In addition to the processes implicated in all lists, the sex-biased targets were further enriched for Wnt signaling and transforming growth factor-beta (TGF-β) pathways. This result suggests these pathways may partially underlie normal behavioral differences in executive functioning between males and females. Furthermore, these two pathways are implicated in neurological disorders with sex-biased differences in prevalence (Freese et al. 2010; Krieglstein et al. 2011), and therefore may relate this sex disparity to underlying miRNA expression differences during normal brain development.

Next, I assessed for enrichment of miRNA targets among genes previously implicated in various neurological and psychiatric disorders that have significant genetic etiology (see Methods). I tested for enrichment of genes involved in epilepsy, three neurodevelopmental disorders (autism, schizophrenia, and bipolar disease), three neurodegenerative disorders (Alzheimer's, Huntington's, and Parkinson's diseases), and three psychiatric diseases (major depressive disorder, post-traumatic stress disorder, and obsessive-compulsive disorder). The enrichment of all gene lists significant for various disorders is shown in **Figures 2.4.3** and **2.4.4**. The three neurodevelopmental disorders (ASD, Schizophrenia, and Bipolar) showed a nearly identical enrichment pattern, among many categories. In contrast, there was almost no enrichment for neurodegenerative disease lists. Similarly, the neuropsychiatric disorders showed no enrichment for miRNA target genes, except for major depressive disorder, where the pattern was similar to the neurodevelopmental disorders.

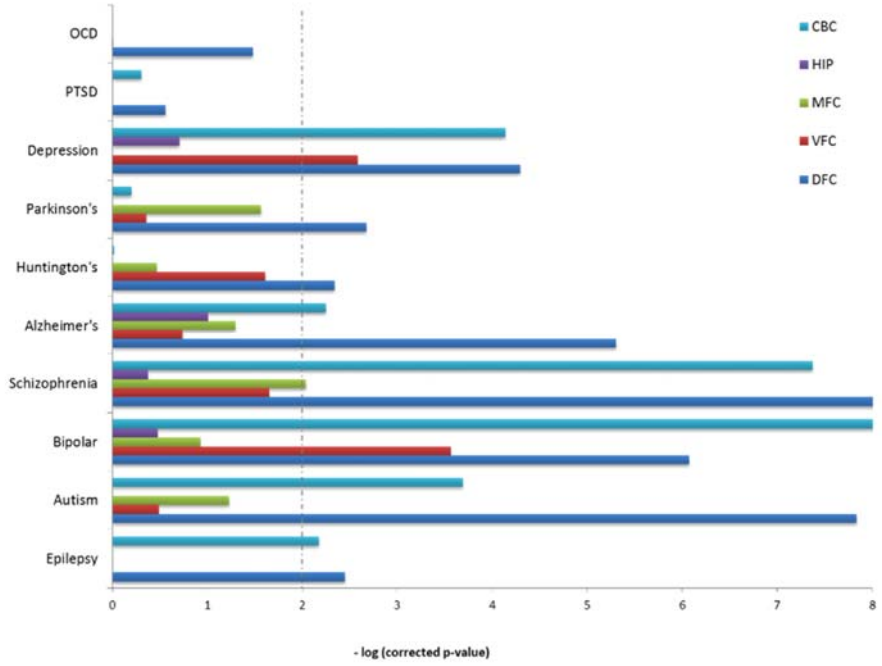


Figure 2.4.3. Enrichment of differentially expressed miRNA target genes by brain region for disease associated genes. Dashed line indicates significance (corrected p-value < 0.01).

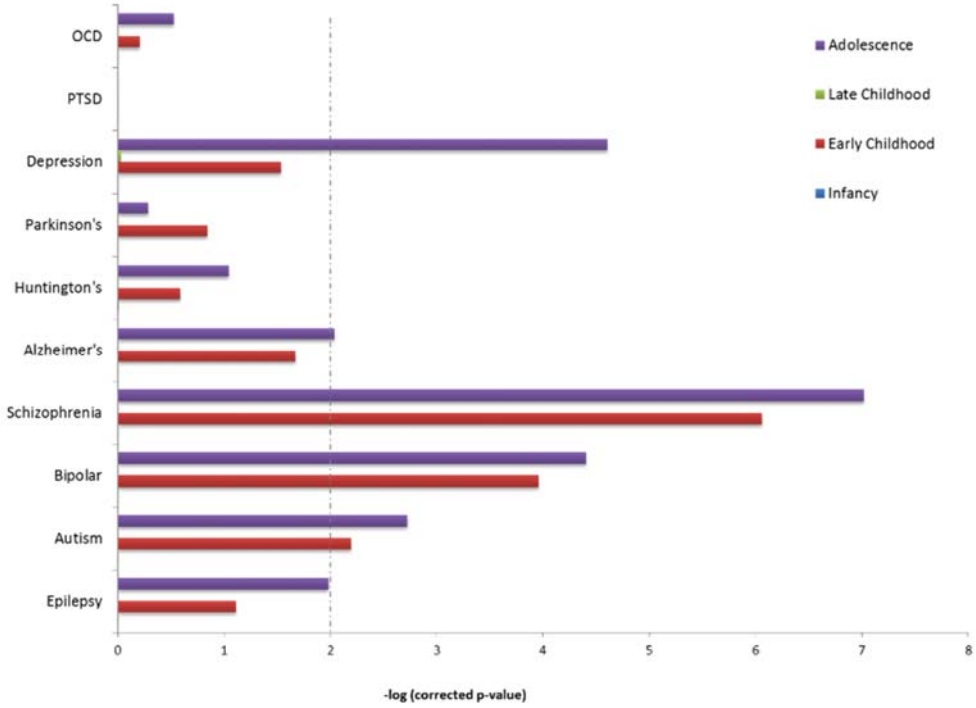


Figure 2.4.4. Enrichment of differentially expressed miRNA target genes among male versus female sets for disease associated genes. Dashed line indicates significance (corrected p-value < 0.01). Note that there was little to no enrichment of Infancy and Late Childhood time periods.

Finally, I explored the temporal-spatial correlation between three high-confidence autism candidate genes (*PTEN*, *BDNF*, and *MECP2*) and their experimentally-known regulator miRNAs (Mellios and Sur 2012). To do so, the Pearson correlation was calculated between each ASD candidate gene and its cognate miRNA by brain region across all of developmental time (**Figure 2.4.5a-c**). This analysis showed that miRNA-mediated gene suppression appears to be region specific, and that this region specificity is unique to each gene-miRNA pair, as significant anti-correlations were only found in certain brain regions and these regional patterns differed between the three genes. For instance, both *PTEN* and *BDNF* appear to be significantly down-regulated by their cognate miRNAs in the dorsolateral and ventrolateral prefrontal cortices, but not in other brain regions; whereas *MECP2* appears to be regulated by its miRNA mainly in the cerebellum. Furthermore, different mature isoforms of the same precursor miRNA sometimes display opposite correlations (e.g. miR212-3-p vs. miR21205-p in panel C, DFC), providing strong evidence for highly-specific miRNA-mediate gene repression.

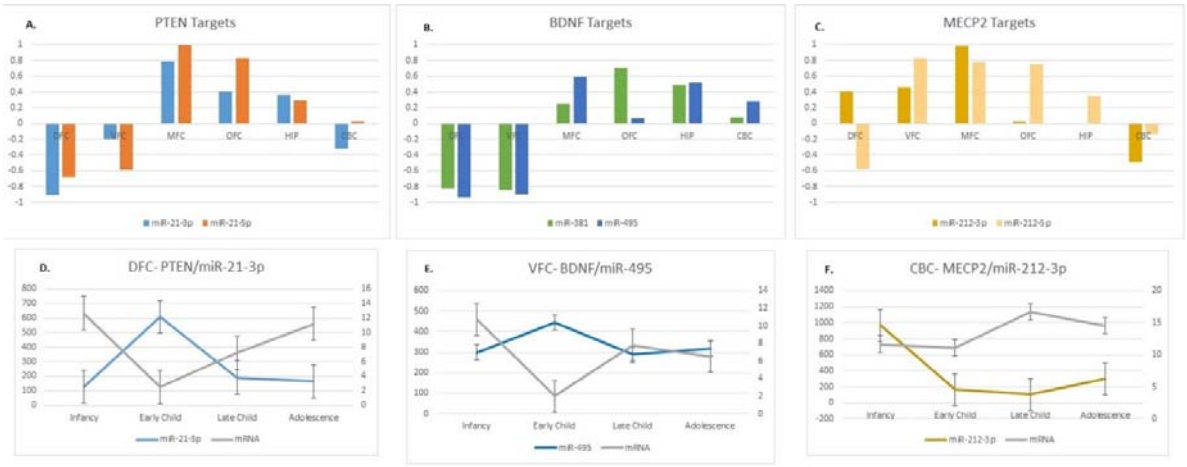


Figure 2.4.5. Temporal, spatial, and isoform-specific miRNA regulation of three autism candidate genes. a-c. Pearson correlation analysis between three autism candidate genes and their known miRNAs by brain region across all of developmental time. **d-f**. Temporal profile of significant miRNA-gene expression pairs.

Assessing the temporal profile of significant miRNA-gene expression pairs shows that miRNA-mediated gene suppression also appears to be time-period specific (**Figure 2.4.5d-f**). For instance, miR-21-3p appears to modulate expression of PTEN throughout development in the DFC (**d**), whereas miR-212-3p appears to only modulate *MECP2* expression in the

cerebellum after infancy (**f**). Taken together, these results further illuminate the critical regulatory roles that miRNAs play in modulating expression of ASD candidate genes.

2.4.5 Discussion

In summary, the work described in this chapter represents the most comprehensive assessment to date of spatio-temporal miRNA expression in the developing human brain. I identified miRNAs differentially expressed both within and between brain regions, and demonstrated that the greatest shifts in miRNA expression occur shortly after birth. However, unlike global gene expression patterns, miRNAs become more differentially expressed between brain regions over time, potentially driving regional specialization as the brain matures. Target genes under putative control by region-specific differentially expressed miRNAs are most related to the processes of transcription regulation and neurodevelopment, highlighting the central function of these miRNAs to brain transcription networks. Additionally, sex-biased expression of miRNAs increases in the prefrontal cortex around puberty, and the pathways related to sex-biased target genes are further enriched for Wnt signaling and TGF-β pathways. Common neurodevelopmental disorders with complex genetic etiologies are highly related to genes targeted by these miRNAs, but this was not found for genes related to neurodegenerative or other neuropsychiatric diseases with adult onset. Examining the specific relationship between three high confidence ASD candidate genes and their experimentally-known miRNAs showed that miRNA-mediated gene silencing appears to be highly temporally and spatially specific, and even isoform-specific miRNA regulation appears during neurodevelopment. These results highlight the importance of miRNAs in understanding the functional genomics of ASD, and suggest that future work should more closely examine the role of miRNAs in ASD molecular pathogenesis.

This study has a number of important limitations. First, the total sample size is 18 donor brains, potentially limiting the statistical power. Unfortunately, this problem is prevalent throughout human neurosciences research owing to the lack of large repositories of human post mortem brain tissue (Button et al. 2013). Therefore, it will be important for future studies to replicate and aggregate the data presented here with larger datasets when they become available. Additionally, while computational prediction of miRNA targets based on sequence homology is an effective discovery tool, individual miRNAs of interest will require *in vitro* or *in vivo* experimental validation of their targets.

2.4.6 Conclusion

In conclusion, while thousands of genes are differentially expressed throughout human neurodevelopment, I have identified a set of miRNAs with differential spatio-temporal and sex-biased expression patterns that may regulate these expression changes. A number of the identified miRNAs are of note for their known role in neurodevelopmental processes.

For instance, miR-9, which I found to be increased in expression nearly 5-fold in the hippocampus of early childhood samples as compared to infants (FDR = 0.0039), is known to be a critical regulator of neural progenitor migration and proliferation (Delaloy et al. 2011). Intriguingly, I did not observe increased miR-9 expression in any other brain region during post-natal development, which is to be expected, as the hippocampus was the only region assessed that contains neural stem cells after embryonic development (Song et al. 2002). Similarly intriguing was the finding of significant differential expression of miR-103 between the prefrontal cortex of males and females in adolescence (fold change 1.73, FDR = 0.0041). MiR-103 has been demonstrated to regulate expression of the insulin like growth factor (IGF) family of proteins (Liao and Lonnerdal 2010), of which IGF-2 is known to exhibit genomic imprinting—the phenomena of expressing an allele from either the paternal or maternal DNA but not both—and has unique, brain-region specific imprinted expression patterns (Pham 1999). This is particularly interesting given that microRNAs are one of the main mechanisms by which genomic imprinting is maintained (Delaval and Feil 2004), and imprinting mechanisms could partially account for the significant sex bias seen in neurodevelopmental disorders like ASD (Skuse 2000). These examples further support the notion that the identified miRNAs are likely critical regulators of neurodevelopmental transcriptional processes.

The targets of these differentially expressed miRNAs are highly enriched for genes related to transcriptional regulation, neurodevelopmental processes, and common neurodevelopmental disorders. Furthermore, inter-regional expression differences of miRNAs appear to increase over development. These results suggest the identified miRNAs are likely hubs of critical brain developmental and pathologic transcriptional processes.

Chapter 3. Functional Genomics Studies of Autistic Post-mortem Brain Tissue

In Chapter 2, I studied autism candidate genes during normal human neurodevelopment in an attempt to discover potential shared molecular and cellular mechanisms through which the hundreds of genes implicated in ASD may ultimately converge upon to result in the common clinical phenotype. My results repeatedly implicated a number of pathways that had previously been independently linked to ASD: immune and cytokine signaling, glia, synaptogenesis, transcription/translation, mitochondrial function, and non-coding RNAs.

In this chapter, I performed three studies directly assessing autistic post-mortem brain tissue for defects in these processes, in order to corroborate the mechanisms identified in Chapter 2 that were inferred based on ASD candidate gene function during normal human neurodevelopment. First, I assessed for aberrant expression of a class of ncRNAs not previously examined in autism—long non-coding RNAs. My results demonstrated that lncRNAs that are differentially expressed in autistic brain are preferentially located near genes involved in neurodevelopment, and that broad patterns of transcriptional dysregulation in autistic brain become apparent when assessing mRNAs and lncRNAs in parallel. Next, I performed a direct assessment of glial marker genes in autistic brain tissue, demonstrating that both microglia and astrocyte specific cell surface markers are altered in autistic brain in a manner similar to previously published studies that assessed glial cell numbers in ASD via other techniques. Finally, I performed the first assessment to date of the mitochondrial transcriptome in ASD brain tissue, and show that altered mitochondrial genes are related to apoptosis in ASD prefrontal cortex, in addition to oxidative metabolism—potentially linking previously separate lines of evidence implicating glia/apoptosis and metabolism in ASD.

The work in this chapter provides direct evidence in ASD brain tissue that the processes I discovered ASD candidate genes are most highly related to in normal neurodevelopment (Chapter 2) are in fact abnormal in autistic brain. This both underscores the importance of the approach taken in Chapter 2 to discovering ASD pathways, and provides further evidence for these processes being involved in the pathogenesis of ASD.

Post-Mortem Autistic Brain Tissue Dataset

The following three studies discussed in this Chapter (3.1 – 3.3) were conducted on human post-mortem brain tissue from individuals with autism and matched controls. Frozen brain tissue was obtained from the NICHD Brain and Tissue Bank for Developmental Disorders at the University of Maryland, Baltimore, MD. The collection protocol at the University of Maryland, Baltimore was reviewed and approved by the Institutional Review Board of that institute. This source obtained consent to use brain tissue for research from each patient or their guardian prior to his/her death, and their protocol was approved by an appropriate Institutional Review Board. No patient-specific identifiable information was obtained.

Because multiple brain regions have been implicated in ASD, we requested tissue from two separate brain areas that have been consistently demonstrated as abnormal in autism—the prefrontal cortex and the cerebellum. We further requested that if possible, the prefrontal cortex and cerebellum samples should originate from the same donor brain, in order to be able to make intra-individual comparisons. We obtained post-mortem prefrontal cortex (PFC) brain tissue from five individuals with autism and from four healthy controls (**Table 3.0.1**). We obtained post-mortem cerebellum brain tissue from four individuals with autism and four healthy controls. While many of PFC/Cerebellum samples were obtained from the same donor brain, this was not available for all donors. The tissue samples in which the matching PFC or Cerebellar sample are included in this dataset are marked with an asterix in **Table 3.0.1**.

All cases were Caucasian males, and case controls were matched by age as closely as possible. The average post-mortem interval (PMI) was not significantly different between autistic and control tissue samples (**Table 3.0.1**; ASD = 17.9hrs, ctrl = 13.2hrs, p-value = 0.16). This remained true after sub-stratifying by brain region (ASD PFC = 19.4hrs, ctrl PFC = 13.2hrs, p-value = 0.28; and ASD cerebellum = 16.0hrs, ctrl cerebellum = 13.25hrs, p-value = 0.48). RNA isolated from post-mortem brain tissue was generally of high quality, and the RNA Integrity Number (RIN) was not significantly different between autism and controls (ASD = 5.84, ctrl = 6.18, p-value = 0.67). The RIN was also not significantly different after sub-stratifying by brain region (ASD PFC = 4.92, ctrl PFC = 5.80, p-value = 0.13; and ASD cerebellum = 7.00, Ctrl cerebellum = 6.65, p-value = 0.85).

Table 3.0.1. Clinical characteristics and RNA quality of autistic and control samples.

Sample#	UMB #	Diagnosis	Brain Area	Age (yrs)	PMI (hrs)	RNA Quality		
						A260/280	A260/230	RIN
1	5308*	Autism	PFC	4.5	21	2.051	2.266	4.9
2	1349	Autism	PFC	5.6	39	2.044	2.232	4.3
3	5144	Autism	PFC	7.2	3	2.058	2.271	5.4
4	5302*	Autism	PFC	16.3	20	2.031	2.238	4
5	4999*	Autism	PFC	20.8	14	2.039	2.232	6
6	4670*	Control	PFC	4.6	17	2.048	2.276	5.2
7	1185	Control	PFC	4.7	17	2.026	2.243	4.7
8	4898	Control	PFC	7.7	12	2.056	2.183	5.9
9	4848*	Control	PFC	16.7	15	2.044	2.185	6.7
10	4727*	Control	PFC	20.5	5	2.066	2.209	6.5
11	5308*	Autism	Cerebellum	4.5	21	2.087	1.781	7.3
12	4899	Autism	Cerebellum	14.3	9	2.077	2.314	9.3
13	5302*	Autism	Cerebellum	16.3	20	2.083	1.646	2.2
14	4999*	Autism	Cerebellum	20.8	14	2.081	2.114	9.2
15	4670*	Control	Cerebellum	4.6	17	2.088	2.161	6.1
16	4722	Control	Cerebellum	14.5	16	2.073	2.828	6.5
17	4848*	Control	Cerebellum	16.7	15	2.087	2.327	6.8
18	4727*	Control	Cerebellum	20.5	5	2.067	2.307	7.2

PFC: Prefrontal cortex; UMB: University of Maryland Brain Bank sample number; PMI: Post-mortem interval; RIN = RNA integrity number. *indicates both prefrontal cortex and cerebellum samples were present from the same donor brain.

A few brain samples were of lower RNA integrity than the others (for example, sample 4 and 13). To control for this I first ensured that the group differences in RNA quality were not significantly different (discussed above), as differential expression analysis was subsequently performed at the group level. Secondly, in the individual analyses that used these two samples (Chapter 3.2 and 3.3), I assessed for differentially expressed genes separately after removing these two samples from the analysis, and found there to not be any significant change in the results. Therefore, these samples were retained for the final analysis. As

compared to other reports studying autistic post-mortem brain tissue, the average RNA integrity values for these samples is good (Kang et al. 2011; Chow et al. 2012).

As is noted in the respective methods sections below, in Chapter 3.1 only a subset of this entire dataset was analyzed. In Chapters 3.2 and 3.3, all available post-mortem tissue was analyzed.

--

3.1 Long non-coding RNAs are Dysregulated in Autistic Prefrontal Cortex and Cerebellum

3.1.1 Aim

The autism spectrum disorders have a significant hereditary component, but the implicated genetic loci are heterogeneous and complex. Consequently, there is a gap in understanding how diverse genomic aberrations all result in one clinical ASD phenotype. Gene expression studies from autistic brain tissue have demonstrated aberrantly expressed protein-coding genes may converge onto common molecular pathways, potentially reconciling the strong heritability and shared clinical phenotypes with the genomic heterogeneity of the disorder. However, the regulation of gene expression is extremely complex and governed by many mechanisms, including non-coding RNAs. Yet no study in ASD brain tissue has assessed for changes in regulatory long non-coding RNAs, which represent a large proportion of the human transcriptome, and actively modulate mRNA expression. To assess if aberrant expression of lncRNAs may play a role in the molecular pathogenesis of ASD, I profiled over 33,000 annotated lncRNAs and 30,000 mRNA transcripts from post-mortem brain tissue of autistic and control prefrontal cortex and cerebellum by microarray. I detected over 200 differentially expressed lncRNAs in ASD, which were enriched for genomic regions containing genes related to neurodevelopment and psychiatric disease. Additionally, comparison of differences in expression of mRNAs between prefrontal cortex and cerebellum within individual donors showed ASD brains had more transcriptional homogeneity. Moreover, this was also true of the lncRNA transcriptome. These results suggest that investigation of lncRNA expression in autistic brain may further elucidate the molecular pathogenesis of this disorder.

3.1.2 Introduction

While transcriptome studies in autistic brain samples have demonstrated that aberrant expression of mRNA transcripts may represent a convergence of the heterogeneous genomics of ASD, none of these studies has concurrently assessed the regulatory RNAs that may underlie aberrant mRNA expression. Three studies in lymphoblast cell lines from autistic patients have shown that miRNAs are abnormal in autism (Talebizadeh et al. 2008;

Sarachana et al. 2010; Seno et al. 2011), and Abu-Elneel *et al* demonstrated differential expression of miRNAs in autistic cerebellum (Abu-Elneel et al. 2008). However, a novel class of regulatory RNAs, long non-coding RNAs, has recently been implicated in a number of fundamental gene regulatory events, but their role in autism molecular pathogenesis remains unknown.

Therefore, the purpose of this study was to determine if dysregulated expression of lncRNAs might play a role in the molecular pathogenesis of ASD. To do so, I profiled over 33,000 annotated lncRNAs in ASD patient post-mortem brain tissue (prefrontal cortex and cerebellum) using microarrays. In parallel, I also assessed for transcriptional differences in all known protein-coding mRNAs. I identified over 200 differentially expressed lncRNAs, which were oriented in or around protein-coding loci strongly enriched for brain development genes. Moreover, I discovered that the previously reported homogeneity of mRNA transcription within autistic brains is also observed within the lncRNA component of the transcriptome.

3.1.3 Methods

Brain Tissue

The human post-mortem brain tissue used in this study is a subset of all the samples described in **Table 3.0.1** above. As this was a pilot investigation of lncRNAs in autism, only two autistic prefrontal cortex samples and two autistic cerebellar samples were assessed (and compared to two matched control prefrontal cortex and two matched control cerebellar samples). These four donors' brains are detailed below in **Table 3.1.1.** Of note, from each of the four donors both prefrontal cortex and cerebellum samples were obtained in this analysis.

RNA Isolation and Quality Control

Total RNA was extracted by homogenizing samples in TRIzol® Reagent (Invitrogen) according to the manufacturer's protocol. RNA quantity was measured by NanoDrop ND-1000. Agilent Bioanalyzer 2100 was used to assess RNA integrity for each sample (**Table 3.1.1**).

Table 3.1.1. Characteristics of patients from whom brain samples were obtained.

UMB#	Sex	Race	Diagnosis	Age (yrs)	Age (days)	PMI (hrs)	RIN	Cause of death
5308	Male	Caucasian	Autism (ADI-R)	4	182	21	8.9-9.1	Skull fracture
5302	Male	Caucasian	Autism (Clinical)	16	119	20	8.8-8.9	DKA
4670	Male	Caucasian	Control	4	237	17	8.9-9.0	Commotio Cordis
4848	Male	Caucasian	Control	16	271	15	9.0-9.0	Drowning

RIN = RNA integrity number calculated from isolated RNA used for analysis; ADI-R = Autism Diagnostic Interview-Revised; PMI = Post-mortem interval; DKA = diabetic ketoacidosis

lncRNA Microarray

ArrayStar, Inc (Rockville, MD) Human lncRNA Microarray V2.0 was used and run by the service provider. The array contained 33,045 lncRNAs and 30,215 protein-coding transcripts. The lncRNAs were manually collected from the most authoritative databases such as RefSeq, UCSC knowngenes, Ensembl, and manually curated lncRNA literature sources (**Table 3.1.2**). The mRNAs were obtained from RefSeq (March 2011). Each transcript was represented by a specific exon or splice junction probe. Positive probes for housekeeping genes and negative control probes (i.e. scramble sequences) were also printed onto the array for hybridization quality control.

Table 3.1.2. Source of lncRNAs contained on ArrayStar lncRNA microarray.

Database	# of lncRNAs	Literature Source	# of lncRNAs
RefSeq (March 2011)	2,608	Khalil *et al* "lincRNAs"	3,289
UCSC Known Genes 4	10,380	Calin *et al* "T-UC RNAs"	962
Ensembl 37.59	23,383	Rinn *et al* Hox cluster ncRNA	407
H-invDB 7.0	2,568	Orom *et al* enhancer lncRNAs	3,019
RNAdb 2.0	1,492		
NRED (March 2011)	1,112		

Microarray Labeling, Hybridization, and Scanning

Sample labeling and array hybridization were performed according to the Agilent One-Color Microarray-Based Gene Expression Analysis protocol (Agilent Technology) with minor modifications. Briefly, mRNA was purified from 1 μg total RNA after removal of rRNA (mRNA-ONLY™ Eukaryotic mRNA Isolation Kit, Epicentre). Then, each sample was amplified and transcribed into fluorescent cRNA along the entire length of the transcripts without 3' bias utilizing a random priming method. The labeled cRNAs were purified by RNAeasy Mini Kit (Qiagen). The concentration and specific activity of the labeled cRNAs (pmol Cy3/μg cRNA) was measured by NanoDrop ND-1000. 1μg of each labeled cRNA was fragmented by adding 11μl of 10x Blocking Agent and 2.2μl of 25x Fragmentation Buffer, heating the mixture at 60°C for 30 min, then adding 55μl of 2x GE Hybridization buffer to dilute the labeled cRNA. 100μl of hybridization solution was dispensed into the gasket slide and assembled to the lncRNA expression microarray slide. The slides were incubated for 17 hours at 65°C in an Agilent Hybridization Oven. The hybridized arrays were washed, fixed, and scanned using the Agilent DNA Microarray Scanner (G2505B).

Microarray Data Processing

Agilent Feature Extraction software (version 10.5.1.1) was used to analyze acquired array images. Quantile normalization and subsequent data processing were performed using the GeneSpring GX v11.5.1 software package (Agilent Technologies). After quantile normalization of the raw data, lncRNAs and mRNAs that at least 4 out of 8 samples had flagged as "Present" or "Marginal" were chosen for further data analysis. Differentially expressed lncRNAs and mRNAs were identified through fold change (FC) filtering. Differentially expressed lncRNAs and mRNAs with statistical significance (as determined by two-tailed student's t-test < 0.05) were identified through Volcano Plot filtering. All microarray data was deposited into Gene Expression Omnibus (GEO) at the National Center for Biotechnology Information (NCBI), NIH under Series Number GSE36315.

qRT-PCR

Five randomly selected lncRNAs from among those showing the greatest fold change were chosen for confirmation via quantitative real time reverse transcriptase PCR (qRT-PCR). The selected lncRNAs and the primers used for qRT-PCR are described in **Supplementary**

Table S33. Five micrograms of total RNA was used for the synthesis of first strand cDNA using the SuperScript III First Strand cDNA Synthesis Kit (Invitrogen). qRT-PCR analysis was performed using ABI prism 7900 (Applied Biosystems) with SYBR Green expression assay system (Applied Biosystems). Normalized, relative gene expression was calculated using standard ΔΔCt methods using Applied Biosystem RQ Manager Software (v1.2). Each qPCR reaction was run three separate times, with technical triplicates in each reaction.

In silico Mapping Analysis

To assess for the potential *cis*-regulatory effects of the identified lncRNAs, I utilized the Genomic Regions Enrichment of Annotations Tool (Mclean et al. 2010). This program takes genomic coordinates as inputs and outputs nearby genes and their ontologies. Default settings were used for analysis on all probes detected as differentially expressed between ASD and Ctrl (both prefrontal cortex and cerebellum), with curated regulatory domains included.

Gene ontology enrichment analysis

To assess for functional categories that the genes identified as significantly differentially expressed in ASD implicated, I used the Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.7, accessed at: http://david.abcc.ncifcrf.gov/. GO categories were reported as significant only if the p-value after multiple testing corrections was <0.05.

3.1.4 Results

In total, 222 lncRNAs were differentially expressed between ASD and control samples (fold change > 2, $p<0.05$). Eighty-two of these were unique to prefrontal cortex, and 143 were unique to cerebellum (**Fig. 3.1.1**). The majority of differentially expressed lncRNAs in ASD were from intergenic regions (~60%), antisense to protein-coding loci (~15%) or within introns of protein coding genes (~10%), with the others representing overlapping transcripts from exons or introns in both sense and antisense directions. This distribution was not significantly different from the distribution of all lncRNAs detected by the array (**Fig. 3.1.2**). I confirmed a select number of the most highly differentially expressed lncRNAs between autism and controls by qRT-PCR analysis (**Fig. 3.1.3**).

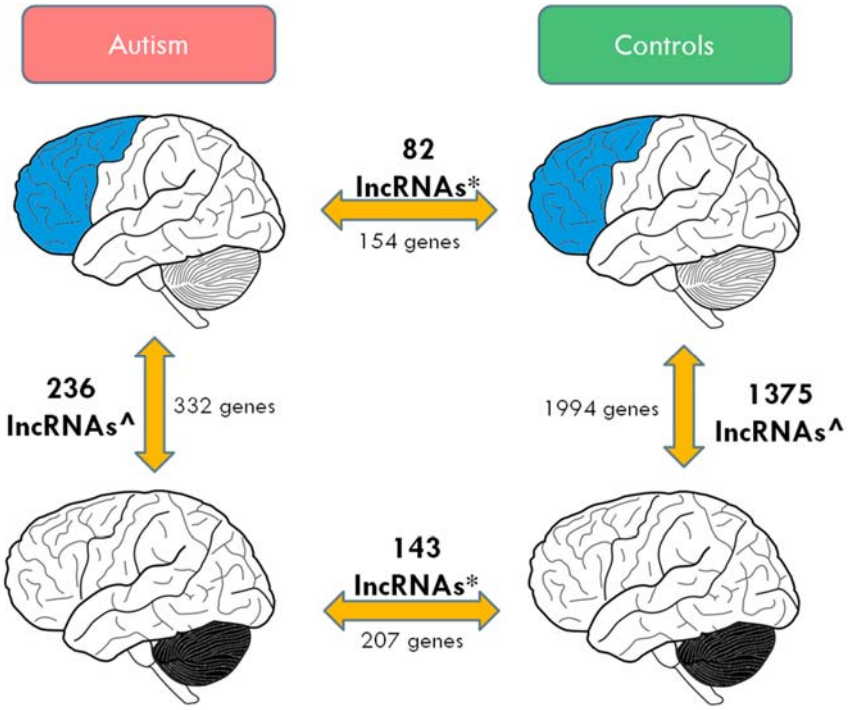
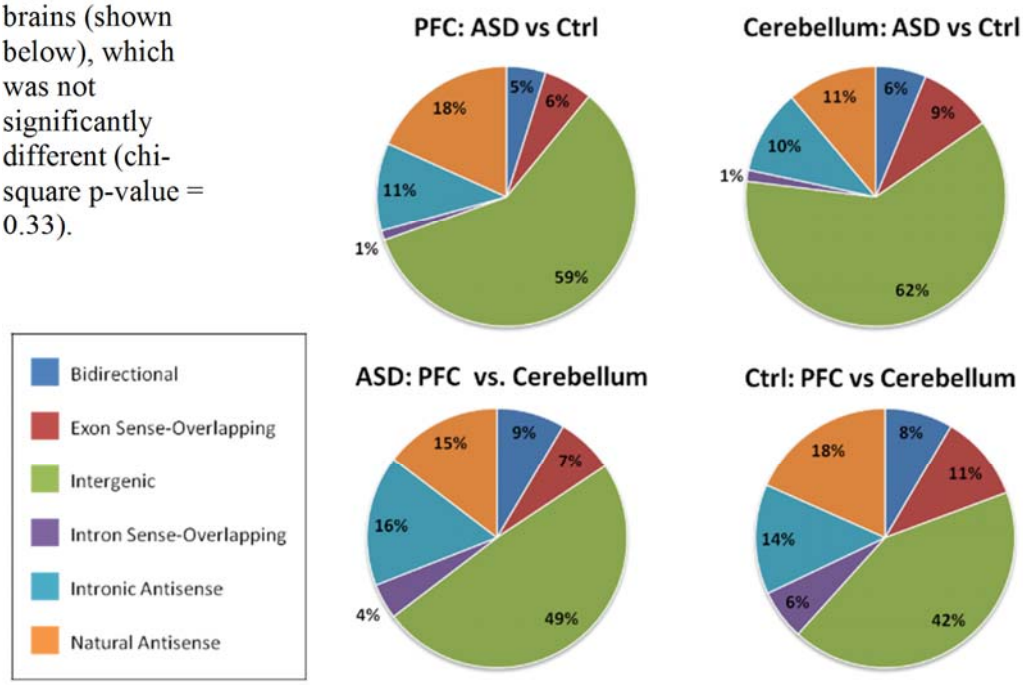


Figure 3.1.1. Summary of differentially expressed lncRNAs and mRNAs. *3 shared lncRNAs, ^99 shared lncRNAs.

Figure 3.1.2. Distribution of differentially expressed lncRNAs by genomic origin. All lncRNAs that were detected by the array in our samples were mapped to their genomic origin, and this distribution was compared to the 222 lncRNAs that were differentially expressed in ASD brains (shown below), which was not significantly different (chi-square p-value = 0.33).



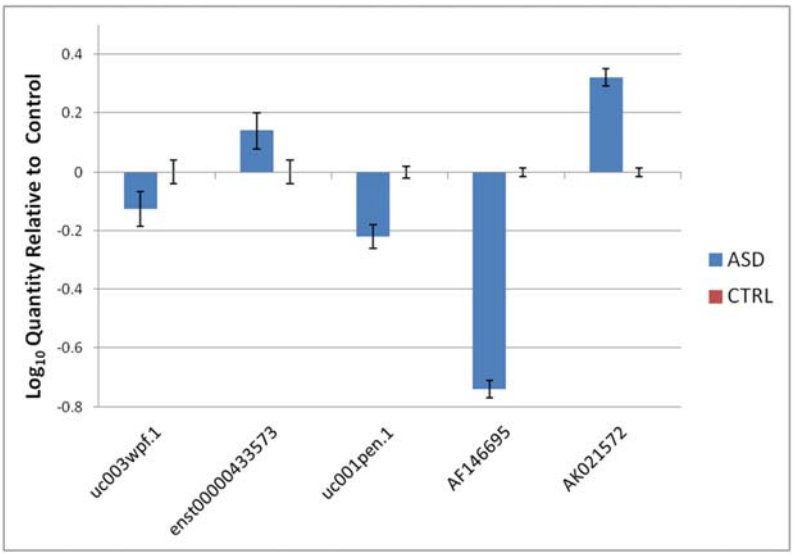


Figure 3.1.3. qRT- PCR Analysis of select lncRNAs. Five lncRNAs that were detected as differentially expressed by microarray were confirmed by qRT-PCR (with the same directional change).

Almost 50% of differentially expressed lncRNAs map to within 50 kilobases (kb) of an annotated gene, and greater than 90% map within 500kb of a known gene (**Fig. 3.1.4**). Mapping all differentially expressed lncRNAs to the nearest genes identified 381 protein-coding loci under putative *cis*-regulatory control by these lncRNAs. The ontologies of those loci implicated two functions: cerebral cortex cell migration and targets of microRNAs mir-103/107 (**Table 3.1.3**). These results are intriguing given that the prevailing cellular model of autism is a defect in neuronal connectivity (Geschwind and Levitt 2007), and that mir-103/107 has previously been implicated in CNS development (Moncini et al. 2011), Alzheimer's disease (Nelson and Wang 2010), and Schizophrenia (Santarelli et al. 2011). Although it is not possible to define the specific aspects of cell mirgration that are implicated here, owing to limitations in gene ontology enrichment analysis, it will be important for future work to investigate this complex processes in more detail. Eleven of these genes near differentially expressed lncRNAs have previously been implicated in ASD, and 18 have previously been shown to exhibit differential expression in ASD brain (**Table 3.1.4**).

Table 3.1.3. Gene ontology analysis of 381 mRNA loci nearby differentially expressed lncRNAs.

Ontology	Term Name	Binomial FDR Q-Value	Binomial Fold Enrichment	Hypergeometric FDR Q-Value	Hypergeometric Fold Enrichment
GO Biological Process	Cerebral cortex cell migration	3.16e-2	4.68	2.06e-2	12.82
MSigDB miRNA Motifs	Targets of miR-103/miR-107	5.88e-8	3.18	3.85e-2	3.30

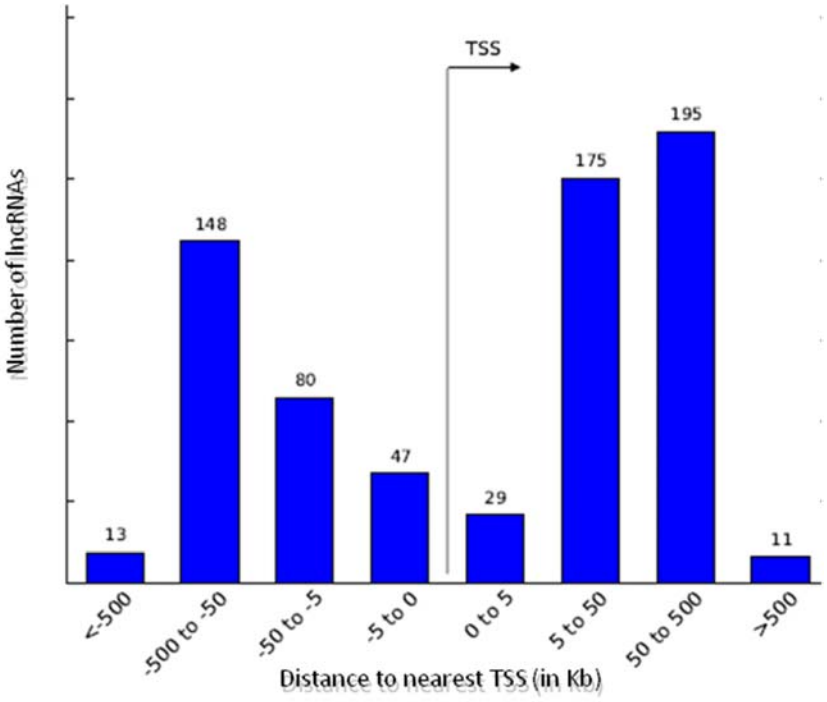


Figure 3.1.4. Relative orientation and distance to the nearest transcriptional start site (TSS) of all differentially expressed lncRNAs.

Table 3.1.4. Genes near differentially expressed lncRNAs that were previously implicated in ASD or shown to be differentially expressed in ASD brains.

Cataloged in AutDB or AGD	Differentially Expressed in ASD brains in Voineagu *et al* study	
DHCR7	DACH1	KIAA0427
DLGAP2	DUSP5	LGALS3
DRD3	ECE2	MAPRE2
HLA-A	FBLN2	NTSR2
RPL10	GSTT1	PFKP
SDC2	HLA-A	TNRC6A
SHANK2	HLA-H	TXNIP
UBE3A	IGFBPL1	
APBA2	ITPR1	
BDNF	KCNB1	
DLX6	KCNG1	

Ninety of the differentially expressed lncRNAs are oriented in or around a known protein-coding region (i.e. not intergenic). Of these, three are known imprinted loci in humans

(C9orf85, SLC4A2, and UBE3A). Interestingly, UBE3A is implicated in the genomic imprinting disorder Angelman Syndrome, which shares many features with ASD (Bonati et al. 2007). Surprisingly, however, only three of these 90 genes are also differentially expressed (RBM8a, ARL17A, KLF6), suggesting perhaps more complex mechanisms for many of these lncRNAs than simple *cis*-regulation.

The array also contained probes for known protein-coding transcripts, of which we detected 355 genes differentially expressed between ASD and controls, which were enriched for the process of alternative splicing (**Table 3.1.5**). This finding is in agreement with a recent large transcriptome study in autistic brains by Voineagu *et al*, where they demonstrated dysregulated splicing of *A2BP1*-dependent exons in ASD brains using RNA-seq (Voineagu et al. 2011).

Table 3.1.5. Gene ontology analysis for differentially expressed mRNAs between autism and control prefrontal cortex.

Category	Term	P-Value	Bonferroni	Benjamini	FDR
SP_PIR_KEYWORDS	alternative splicing	1.70E-04	3.90E-02	3.90E-02	2.10E-01

Because the samples assessed from the prefrontal cortex and the cerebellum were from the same patients, I had the ability to compare intra-individual differences in expression of both genes and lncRNAs between these regions, which has not previously been done. I detected almost 2,000 genes differentially expressed in control prefrontal cortex versus control cerebellum, which were highly enriched for gene ontology terms related to synaptogenesis (**Table 3.1.5**), but only 322 genes differentially expressed between ASD prefrontal cortex and cerebellum (**Fig. 3.1.1**). These results are also in agreement with the study by Voineagu and colleagues, where they observed more transcriptional homogeneity in ASD brains (Voineagu et al. 2011). In light of this, then, it was particularly intriguing to find that the number of lncRNAs differentially expressed within control brains was also much greater than lncRNAs differentially expressed within autism brains (1375 lncRNAs versus 236 lncRNAs, respectively).

Table 3.1.6. Gene ontology analysis for differentially expressed mRNAs within control prefrontal cortex versus cerebellum.

Category	Term	P-Value	Bonferroni	Benjamini	FDR
GOTERM_BP_FAT	synaptic transmission	1.70E-10	6.40E-07	6.40E-07	3.10E-07
GOTERM_BP_FAT	transmission of nerve impulse	2.60E-09	9.90E-06	4.90E-06	4.90E-06
GOTERM_CC_FAT	neuron projection	1.00E-08	5.80E-06	5.80E-06	1.50E-05
SP_PIR_KEYWORDS	alternative splicing	1.20E-07	8.00E-05	8.00E-05	1.80E-04
SP_PIR_KEYWORDS	cleavage on pair of basic residues	2.20E-07	1.50E-04	7.40E-05	3.20E-04
UP_SEQ_FEATURE	splice variant	2.20E-07	9.50E-04	9.50E-04	4.10E-04
GOTERM_BP_FAT	cell-cell signaling	6.60E-07	2.50E-03	8.20E-04	1.20E-03
GOTERM_CC_FAT	synapse	2.30E-06	1.30E-03	6.40E-04	3.30E-03
GOTERM_MF_FAT	substrate specific channel activity	2.10E-06	2.50E-03	2.50E-03	3.30E-03
GOTERM_CC_FAT	postsynaptic membrane	2.50E-06	1.40E-03	4.70E-04	3.60E-03
SP_PIR_KEYWORDS	postsynaptic cell membrane	3.90E-06	2.70E-03	9.00E-04	5.90E-03
GOTERM_CC_FAT	dendrite	4.10E-06	2.40E-03	5.90E-04	6.10E-03
SP_PIR_KEYWORDS	developmental protein	4.40E-06	3.00E-03	7.60E-04	6.60E-03
GOTERM_MF_FAT	calcium ion binding	4.80E-06	5.90E-03	2.90E-03	7.80E-03
GOTERM_CC_FAT	synapse part	6.70E-06	3.80E-03	7.70E-04	9.80E-03
GOTERM_MF_FAT	channel activity	7.30E-06	8.90E-03	3.00E-03	1.20E-02
GOTERM_MF_FAT	passive transmembrane transporter activity	8.00E-06	9.70E-03	2.40E-03	1.30E-02
SP_PIR_KEYWORDS	synapse	9.80E-06	6.70E-03	1.40E-03	1.50E-02
GOTERM_BP_FAT	regulation of nervous system development	1.10E-05	3.90E-02	9.90E-03	2.00E-02
GOTERM_BP_FAT	regulation of neuron projection development	1.60E-05	6.00E-02	1.20E-02	3.00E-02

3.1.5 Discussion

While there have been multiple studies of the mRNA transcriptome in ASD, this chapter describes the first assessment of regulatory lncRNAs in autism post-mortem brain tissue. I identified lncRNAs that are differentially expressed in ASD brain tissue, and show they are enriched for genomic loci involved in neurodevelopment and psychiatric disease. Notably, *trans*-regulatory mechanisms of these lncRNAs are likely to be major contributors to their cellular importance. Future studies using knock-down or over-expression techniques in a relevant model system would be a reasonable approach to uncover potential *trans*-regulatory effects.

Furthermore, both the lncRNA and the mRNA transcriptome appear to be more differentially expressed within control brains (between prefrontal cortex and cerebellum) as compared to ASD brains. This finding is particularly interesting in the context of imaging studies of autistic brain, where it has been suggested that anatomically distinct regions of the autistic brain are less specialized from each other than in healthy subjects (Minshew and Keller 2010). It is intriguing to speculate that perhaps less "genomic differentiation" between brain regions in autism underlies these imaging findings.

Owing to the very small sample size assessed in this work, this study should be considered a pilot assessment until the results can be replicated in other autistic post-mortem brain samples.

3.1.6 Conclusion

In summary, these results identify lncRNAs that are aberrantly expressed in autistic brain, and suggest that perhaps lncRNAs contribute to dysregulation of protein-coding loci in ASD, and/or that a fundamental defect in genome-wide transcriptional regulation—including non-coding regions of the genome—underlies ASD molecular pathology. Future studies will need to replicate and expand these findings in more patient samples, but this initial evidence suggests that the lncRNA component of the transcriptome deserves attention in autism.

--

3.2 Aberrant Glial Marker Expression in Autistic Brains

3.2.1 Aim

The cellular mechanism(s) underlying autism spectrum disorders are not completely understood, but ASD is thought to ultimately result from disrupted synaptogenesis. However, studies have also shown that glial cell numbers and/or function are abnormal in post-mortem brain tissue from autistic patients. Yet direct assessment of glial cells in post-mortem human brain tissue is technically challenging, which has limited glial research in human ASD studies. Therefore, we attempted to determine if glial cell-type specific markers may be altered in autistic brain tissue in a manner that is consistent with known cellular findings, such that they could serve as a proxy for glial cell numbers and/or activation patterns.

3.2.2 Introduction

Two separate bodies of work have independently implicated both synaptogenesis and microglia/astrocyte dysfunction in ASD. However, it is not clear how these separate lines of evidence may converge into a common mechanism in the autistic brain that ultimately results in the shared clinical phenotype. Because separate studies have shown that microglia and astrocytes play critical roles in sculpting developing synapses during normal neurodevelopment (Eroglu and Barres 2010; Paolicelli et al. 2011), it is reasonable to hypothesize that inherent defects or aberrant numbers of microglia and astrocytes in the developing autistic brain may be causative of the synaptic abnormalities by affecting the proper wiring of developing neuronal connections. However, because appropriately-preserved post-mortem autistic brain tissue is lacking (Abbott 2011), cellular-level studies assessing glial numbers and activation in human autistic brains have been limited. Moreover, quantification of cell numbers in postmortem tissue by stereology is technically challenging, further limiting the ability of researchers to assess the few appropriately-preserved tissue samples that are available. Finally, no studies have concurrently specifically assessed for microglia, astrocytes, and neurons in the same set of autistic brain samples. As a consequence, a comprehensive understanding of the relationship between glial and neuron cells in autistic brains is needed.

Therefore, the purpose of this study was two-fold. First, we sought to determine if microglia, astrocyte, and neuron-specific markers were altered in post-mortem autistic brain tissue, in order to further investigate the role of glia in ASD. Then, we determined if glial and neuronal cell-type specific marker expression patterns are consistent with known cellular-level findings, because gene expression studies of post-mortem human brain are often easier to perform than cell-level studies, and therefore this approach may serve as a valuable 'screening' assay to infer relative cell proportions.

To do so we compared internally-normalized mRNA expression levels of microglial, astrocyte, and neuronal cell-type specific marker genes in post-mortem brain tissue from patients with autism and healthy controls. Our results provide further evidence for a role of glia in autism pathology, and suggest that assessment of internally-normalized glial cell-type specific markers may serve as a proxy for relative cellular distributions.

3.2.3 Methods

Post mortem brain samples

The post-mortem brain tissue assessed in this study is the full dataset describe in Table 3.0.1 above (page 112). This dataset contains five prefrontal cortex samples from autistic donors and five prefrontal cortex samples from age/sex matched controls, and four cerebellum samples from autistic donors and four cerebellum samples form age/sex matched controls. When possible, the same donor was used to obtain prefrontal cortex and cerebellum samples from, as is noted in Table 3.0.1 with an asterix.

RNA isolation and quality control

RNA isolation and quality control analysis were performed as previously described in the proceeding Chapter (see Chapter 3.1, Methods). Briefly, total RNA was extracted using TRIZOL Reagent (Invitrogen) according to the manufacturer's protocol. Quantification of RNA was performed using a NanoDrop ND-1000, and RNA integrity was assessed using an Agilent Bioanalyzer 2100 (**Table 3.0.1**).

Reverse transcriptase reaction

Total RNA (1ug) was used in a 20µL reverse transcriptase reaction to synthesize cDNA with SuperScript3 Reverse Transcriptase (Invitrogen) according to the manufacturer's protocol. Briefly, 1µg of total RNA added to an aqueous solution containing 250ng/µl of random hexamer (Operon) and 10mM deoxyribonucleotide triphosphate. The RNA was denatured for 5 minutes at 65 °C and then snap cooled on ice for 2 minutes. To each sample mixture was added 0.1M DTT, 5x First-Strand Buffer (250mM Tris-HCl, 375mM KCl, 15 mM $MgCl_2$), RNaseOUT Recombinant Ribonuclease Inhibitor (40 Units/µl), and SuperScript3 Reverse Transcriptase (200units/µl). The reaction was carried out under the following conditions: 25 °C for 5 minutes, 50 °C for 60 minutes, and 70 °C for 15 minutes. The cDNA produced from the reaction was diluted to 0.25x with nuclease free water.

Real time quantitative PCR

SYBR Green Expression Assay System (Applied Biosystems) was used to measure relative, normalized, mRNA expression levels. We assessed four separate microglial-specific cell surface genes: *TREM2, DAP12, CX3CR1,* and *AIF1* (Ransohoff and Perry 2009). Two cell type specific intermediate filaments, glial fibrillary acidic protein (*GFAP*), which is astrocyte-specific (Baba et al. 1997), and the pan-neuronal cell marker *NEFL* (Lepinoux-Chambaud and Eyer 2013), were used to assess for astrocytes and neurons, respectively. Additionally, we assessed for GABAergic interneurons specifically with parvalbumin (*PVL)* (Conde et al. 1994). The intermediate filament housekeeping gene beta-actin (*ACTB)* was used as an endogenous control. Forward and reverse primer sequences were generated using Primer3 software and synthesized by Operon (**Table 3.2.2)**.

Quantitative reverse transcriptase polymerase reaction (qRT-PCR) was performed using an ABI Prism 7900 Sequence Detection System (Life Technologies) with a 96-well format. Each qRT-PCR reaction contained 6.5µl water, 12.5µl SYBR Green master mix (Applied Biosystems), 1µl forward primer (10µM), 1ul reverse primer (10µM), and 4µl of cDNA (0.25x). Data was collected using the SDS2.3 Program (Applied Biosystems) under the following run parameters: 48°C for 30 min, 95°C for 10 min, 40 cycles of 95°C for 15 sec, 60°C for 1 min, and a final dissociation stage.

Table 3.2.1. Primers used for qRT-PCR.

Primer Name	Primer Sequence (5' to 3')	OD	MW	% GC content	T_m (°C)
ActinB-F	AGAAAATCTGGCACCACACC	4.1	6064	50	60.4
ActinB-R	AGAGGCGTACAGGGATAGCA	4.3	6240.1	55	62.4
Trem2-F	CCGGCTGCTCATCTTACTCT	3.3	5995	55	62.4
Trem2-R	AGTCATAGGGGCAAGACACC	4.2	6160	55	62.4
Dap12-F	GAGACCGAGTCGCCTTATCA	3.8	6102	55	62.4
Dap12-R	GTCATGATTCGGGCTCATTT	3.7	6114.1	45	58.4
Cx3cr1-F	GCAGATCCAGAGGTTCCCTT	3.7	6093	55	62.4
Cx3cr1-R	TAACAGGCCTCAGCCAAATC	3.9	6055	50	60.4
Gfap-F	CTGCGGCTCGATCAACTCA	3.5	5748.8	57.9	62.3
Gfap-R	TCCAGCGACTCAATCTTCCTC	3.6	6277.1	52.4	62.7
Nefl-F	AGCTGGAGGACAAGCAGAAC	4.4	6209.1	55	62.4
Nefl-R	TGCCATTTCACTCTTTGTGG	3.5	6065	45	58.4
Parvalbumin-F	CTGGAGACAAAGATGGGGAC	4.3	6240.1	55	62.4
Parvalbumin-R	CAGAGAGGTGGAAGACCAGG	4.4	6265.1	60	64.5
Aif1-F	AGCAGTGATGAGGATCTGCC	4.0	6182.1	55	62.4
Aif1-R	AGCATTCGTTTCAGGGACAT	3.9	6132.1	45	58.4

F: forward; R: reverse; OD: optical density; MW: molecular weight; T_m: melting temperature.

Data Analysis

The target genes and the endogenous control were measured with technical triplicates in each qRT-PCR run, and all genes were assessed in three separate, independent qRT-PCR runs. The cycle threshold number (C_t) was calculated using RQ Manager 1.2 Software (Applied Biosystems). Relative expression of each target gene was normalized to *ACTB* using the $\Delta\Delta C_t$ method. All p-values reported are based on a two-tailed Student's t-test. Only results with a p-value less than 0.05 were considered significant.

3.2.4 Results

In the pre-frontal cortex, relative quantification of microglial markers demonstrated significantly increased expression in autistic samples of *TREM2, DAP12, CX3CR1*, and *CD11b,* but not *AIF1* (**Figure 3.2.1**). The expression of *TREM2* was highest of all microglial markers, approximately 1.75-fold higher in autism brain tissue than controls (p=0.0016). The levels of *CX3CR1*, *CD11b,* and *DAP12* were 1.50-fold (p=0.0092), 1.39-fold (P=0.0017), and 1.34-fold (p=0.0086) higher in autistic samples relative to controls, respectively. Similarly, the relative expression of astrocyte marker *GFAP* was significantly higher in autism brain tissue (1.70-fold, p=0.0049). Conversely, however, both the pan-neuronal marker *NEFL,* and the GABAergic interneuron-specific marker *PVA,* were significantly lower in autistic samples compared to controls (0.68-fold, p=0.0034; and 0.52-fold, p=0.0020, respectively).

In post-mortem cerebellum, the relative expression of astrocyte marker *GFAP* was also significant higher in autism samples than in healthy controls (2.63-fold, p=0.0022; **Figure 3.2.2**). In contrast, the expression of microglial markers *TREM2, DAP12, CX3CR1,* and *AIF1* were *lower* in autism tissue than in control tissue, with fold changes of 0.780 (p=0.0056), 0.797 (p=0.0083), 0.659 (p=0.0029), and 0.808 (p=0.0052), respectively. Relative expression of neuronal markers *PVA* and *NEFL* were also lower in autism samples than in control samples (0.862-fold, p=0.033; and 0.798-fold, p=0.013, respectively), as was found in the prefrontal cortex.

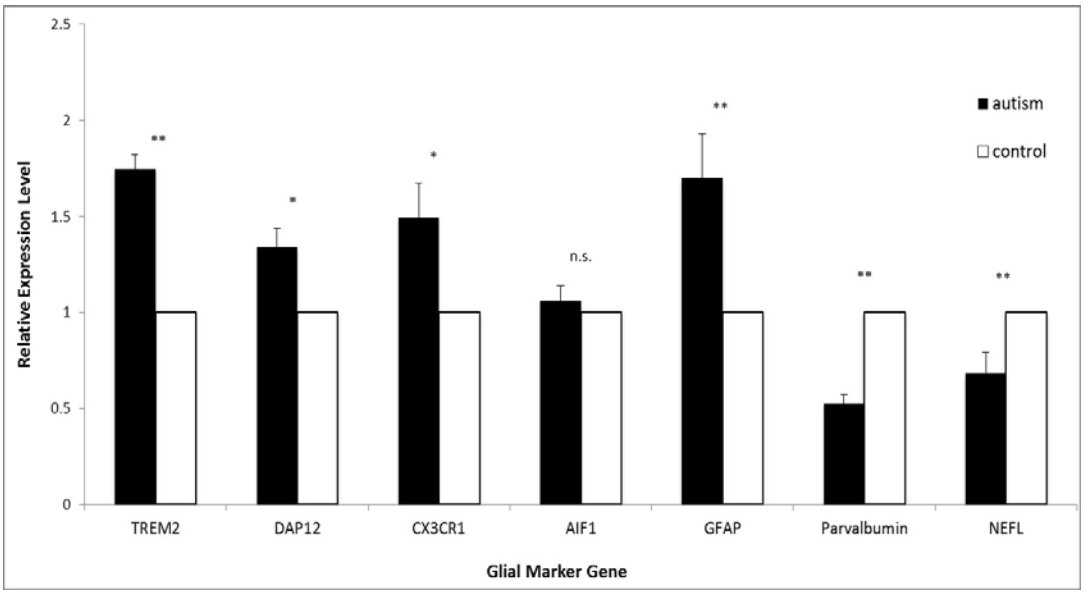


Figure 3.2.1. Expression of cell-type specific markers in pre-frontal cortex samples of autistic cases relative to controls. *p < 0.05, **p < 0.005, n.s. = not significant.

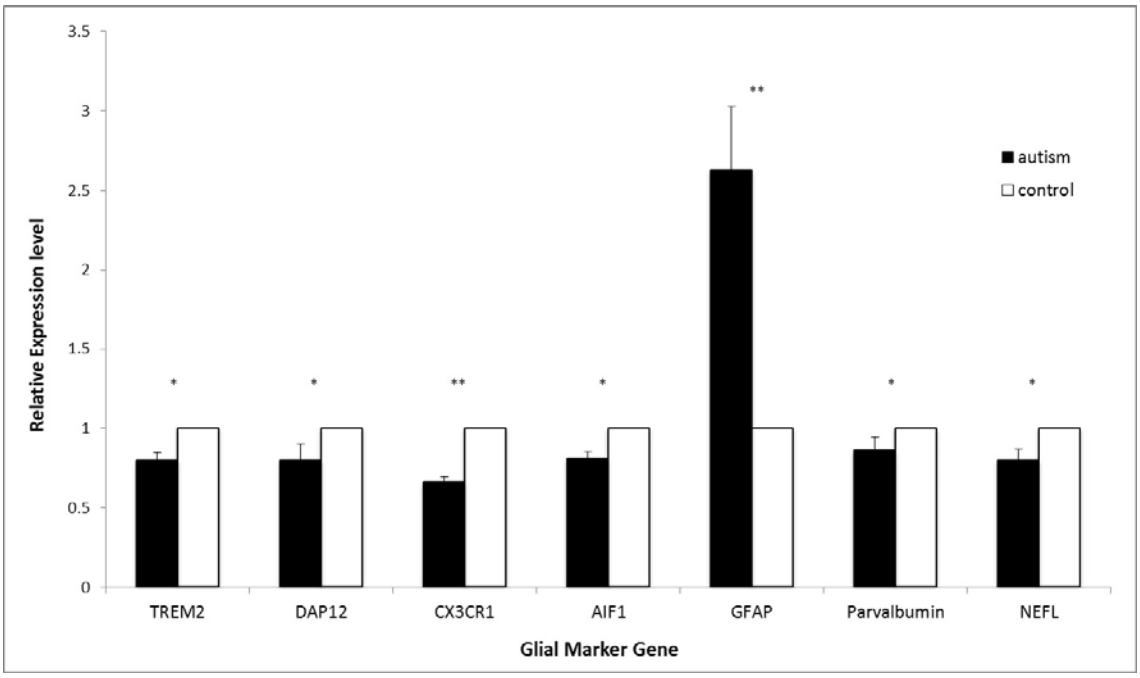


Figure 3.2.2. Expression of cell-type specific markers in cerebellum samples of autistic cases relative to controls. *$p < 0.05$, **$p < 0.005$.

3.2.5 Discussion

While there have been multiple studies assessing RNA expression levels in autistic tissue, here we report the specific assessment of microglial, astrocyte, and neuron-specific cell markers concurrently in two regions of autistic brains. Our results provide further evidence for the role of glia in the cellular pathophysiology of ASD. Moreover, we show that relative gene expression levels of cell-type specific markers may be a useful technique to screen for activation and/or altered numbers of glia and neurons in post-mortem brain tissue.

Microglia cell-marker research is still a relatively new area, and thus the markers used to quantify microglial cell number and activation are still debated. To address this issue, we used four different markers that are putatively microglial-specific. Our results demonstrate that in the PFC, there is increased expression of all microglial markers assessed, although *AIF1* did not reach statistical significance. However, previous reports have shown that *AIF1* expression in the brain is low (Imai et al. 1996), potentially contributing to this result. The finding of increased microglial cell markers in autistic PFC is in agreement with a number of studies that have found increased numbers and activation of microglia in autistic brains. For instance, Morgan, *et al*. showed increased microglial density in dorsolateral prefrontal cortex

(DLPFC) grey matter of ASD brains via IHC and stereology (Morgan et al. 2010), and they also demonstrated that microglia are more closely associated with neurons in autistic DLPFC than in controls (Morgan et al. 2012). Similarly, Tetreault, *et al*. also demonstrated increased microglial density in the frontoinsular and visual cortex of autistic brains as compared to controls (Tetreault et al. 2012). Additionally, a number of studies have specifically identified microglial activation in autistic frontal cortex, through both PET radiotracer imaging (Suzuki et al. 2013) and IHC/cytokine profiling approaches (Vargas et al. 2005). Our results further support these findings, and suggest that microglial cell-marker expression patterns in the frontal cortex may be able to serve as an accurate proxy for more technically challenging studies at the cellular level.

In contrast, our cerebellar results show significantly lower expression of all four microglial cell specific markers in autistic brains. While other studies have identified microglial activation in the cerebellum of autistic tissue (Casanova 2007), no study has attempted to specifically quantify microglial cells in the cerebellum using the markers assessed here, and therefore histopathologic studies in the cerebellum are needed to confirm these findings. One report described increased microglial cell activation in the cerebellum, assessed via HLA-DR staining in the white matter and granular cell layer of the cerebellum (Vargas et al. 2005), and another showed increased microglial activation throughout the brain (although most prominently in the cerebellum) using an *in vivo* PET metabolic radiotracer (Suzuki et al. 2013). However, HLA-DR expression in human microglial cells has been shown to be highly variable between individuals, and its expression actually *decreases* upon cytokine stimulation (Smith et al. 2013). Moreover, as discussed in the Appendix, the cerebellum is anatomically and physiologically unique; thus metabolic and pathological findings in the cerebellum must be interpreted with caution. Furthermore, our tissue samples from cerebellum contained all three layers of the cerebellar cortex, as opposed to the molecular layer only. Until direct histologic assessment of these microglial markers are performed in post-mortem autistic cerebellum, our results must be interpreted cautiously. Yet they suggest that while microglia may be activated in autistic cerebellum, they may not be as numerous as they are in autistic prefrontal cortex.

In both the PFC and the cerebellum, there was significantly increased expression of the astrocyte-specific marker *GFAP* in autistic brains. This trend was most prominent in the

cerebellum, where *GFAP* expression was over two-fold higher in ASD brains than in healthy controls. Our findings parallel those of previous studies, which have showed increased expression of GFAP protein in the cerebellum and cortex of patients of autism through IHC staining, western blotting, and mRNA expression (Vargas et al. 2005; Laurence and Fatemi 2005; Bailey et al. 1998; Purcell et al. 2001). Previous work in rats has demonstrated that GFAP displays a distinctive expression profile over developmental time in the brain. For instance, via Northern blotting and in situ hybridization histochemistry Landry et al. and found a caudal to rostral gradient of *GFAP* expression, consistent with overall brain maturation (Landry et al. 1990). Moreover, both the Landry et al. study and others have shown a transient increase in GFAP expression in the early postnatal period, which is most pronounced in the cerebellum (Sancho-Tello et al. 1995). Again, the distinct developmental expression profile suggests important regulatory roles for GFAP and astrocytes in brain development. While studies have not been done to quantify astrocyte numbers in the autistic cerebellum, our results and those of previous studies provide evidence for astroglial reaction in autism.

Interestingly, we also found significantly decreased expression of the pan-neuronal marker *NEFL* in both the PFC and the cerebellum of autistic brains. This result is also supported by previous studies, which have shown decreased *NEFL* mRNA expression in the anterior cingulate gyrus, motor cortex, and thalamus of autistic brains (Anitha et al. 2012). However, cell-level studies in autistic brain have produced conflicting results about neuron numbers. While a large body of evidence has suggested there is a loss of neurons in many areas of autistic brains (Kern et al. 2013), other studies have shown that young autistic brains may have upward of 70% more neurons in the PFC (Courchesne et al. 2011a). Importantly, though, is the age of the patient at time of death, as longitudinal studies have suggested that early brain overgrowth in ASD quickly reverses to a phenotype of neuronal loss (Courchesne et al. 2011b). Consequently, the older age of patients in this study may bias our findings towards the neuronal loss spectrum of the disease.

Similarly, we found significant decreases in the GABAergic interneuron-specific marker *PVA* in both the prefrontal cortex and cerebellum of autistic samples. Despite many studies demonstrating decreased GABAergic components across different areas of the autistic brain (Fatemi and Blatt 2011), the one pathological analysis of parvalbumin-positive interneurons

in ASD did not identify differences in the autistic cerebellum (Whitney et al. 2009). However, this study only assessed the molecular layer of the cerebellar cortex, whereas our tissue samples contained all three layers. Additionally, while parvalbumin interneurons have been shown to be unchanged in autistic posterior cingulate cortex and fusiform gyrus (Oblak et al. 2011), and increased in autistic hippocampus (Lawrence et al. 2010), they have not been directly assessed in autistic prefrontal cortex.

Overall, our findings suggest that the autistic brain by mid-childhood has molecular changes consistent with astrogliosis and neuronal loss in both the prefrontal cortex and cerebellum, and prefrontal cortex-specific microgliosis and/or activation. These findings support the notion that a complex interplay between glial dysfunction and/or reaction and neurogenesis may underlie the clinical manifestations of autism spectrum disorders.

One potential explanation of these findings is that there is an exaggeration in autistic brains of the normal processes that occur during the completion of brain development. It has been shown that microglial cells in the developing cerebral cortex of prenatal and postnatal macaques and rats limit the production of cortical neurons by phagocytizing neural precursor cells as neurogenesis nears completion (Cunningham et al. 2013). Furthermore, studies of mice with abnormal numbers of microglia have shown that alterations in microglial number perturbs neural development by directly affecting embryonic neural precursors, and induces astrogliosis (Antony et al. 2011). This process appears to be regulated to a large degree by the CX3CL1-CX3CR1 signaling axis, which our results also demonstrate to be abnormal in these autistic brain samples. CX3CL1 is expressed on neurons in the CNS, and CX3CR1 is expressed exclusively on microglia in the brain parenchyma. CX3CR1-deficient mice showed increased microglial cell-autonomous neurotoxicity in three different models of inflammation (Cardona et al. 2006). Perhaps this complex interplay is aberrant in the autistic brains studied, whereby abnormal microglial numbers and/or function result in increased phagocytosis of neural precursors, and consequently the astrogliosis and lower number of neurons described here and in other previous work.

This study has a number of limitations of note. Foremost is the relatively modest sample size. Unfortunately, post-mortem human brain research in general is hampered by the lack of accessibility to tissue samples, and pediatric samples in particular are scarce (Abbott 2011).

Therefore, replication with a large number of samples will be important. However, we chose qRT-PCR techniques in this pilot study because of the increased sensitive compared to whole-genome microarray or sequencing approaches, and therefore some aspects of the small sample size limitation is addressed. Secondly, due to the inter-individual heterogeneity of the brain, and in the brain-banking methodologies used in distinguishing areas of post-mortem brain tissue, it cannot be assumed that all samples will derive from the exact same anatomic site within the prefrontal cortex or cerebellum. This limitation is largely unavoidable. Lastly, the approach of using cell-type specific marker expression as a proxy for cell number and/or activation still needs verification, by assessing them concurrently with traditional histopathologic/ stereology analysis. However, the concordance of our results with previously published studies, and the scarcity of appropriate ASD brain tissue and technical expertise, suggest this may be a valuable and simple alternative 'screening' approach.

3.2.6 Conclusion

In summary, assessment of glial numbers and activation in autistic post-mortem brain research is hampered by the scarcity of appropriately-preserved tissue, and the technical challenge of traditional stereotactic methods. We show that glial and neuron cell-type specific markers have mRNA expression patterns that parallel known cellular aberrations in ASD. Our results provide further evidence that glial cells may play a role in the pathogenesis of ASD, and suggest that assessing for glial cell-type specific marker expression may represent a viable approach to relatively quantify glial cell patterns in ASD post-mortem research.

--

3.3 Altered Expression of the Mitochondrial Genome in Autism

3.3.1 Aim

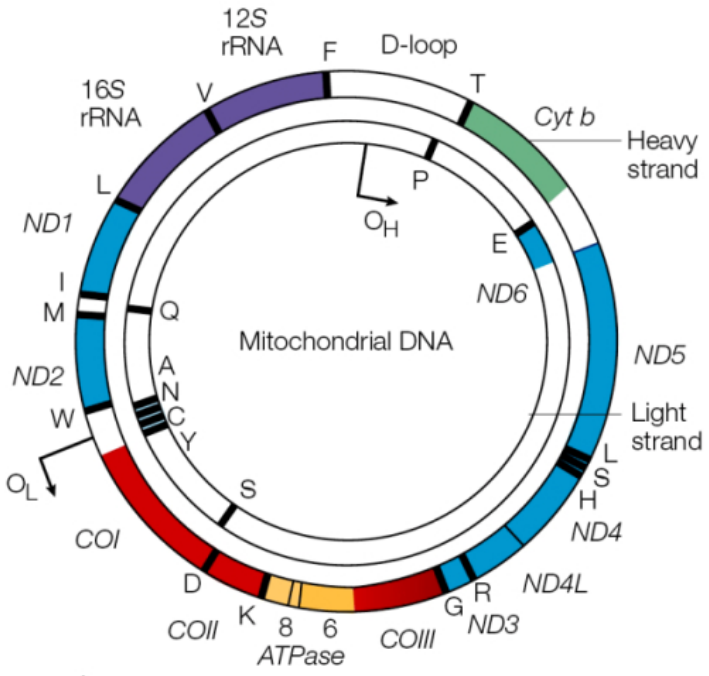
There is a long history of mitochondrial dysfunction in patients with neurologic and psychiatric symptoms, especially children. While recent efforts, including work presented in Chapter 2.2 of this thesis, have implicated mitochondrial dysfunction in ASD, no work has directly assessed the functional genomics of the unique mitochondrial genome in autistic brain tissue. Therefore, the purpose of this work was to specifically assess for changes in gene expression among the 37 mitochondrially-encoded genes, which have not been previously studied in ASD. Furthermore, we also investigated nuclear-encoded genes that function specifically in the mitochondria, in order to comprehensively survey for possible transcriptional dysregulation of the mitochondrial genome in autism.

3.3.2 Introduction

Mitochondria are distinct cellular organelles that are believed to have originated from a symbiotic event between eukaryotes and bacteria, such that in mammalian cells they retain a unique mitochondrial genome (**Figure 3.3.1**) in addition to having co-opted nuclear genes for use in mitochondrial function (Anderson et al. 2003). The main function of mitochondria is to generate adenosine triphosphate (ATP), the energy currency of human cells, from adenosine diphosphate by oxidizing glucose and fatty acids (Nunnari and Suomalainen 2012). Through a series of metabolic biochemical reactions that occur in the mitochondrial matrix (the TCA cycle), the reduced intermediates flavin adenine dinucleotide (FAD) and nicotinamide adenine dinucleotide (NAD) are generated. FAD and NAD then donate protons to the mitochondrial electron transport chain (ETC) in a series of reactions known as oxidative phosphorylation, ultimately resulting in the production of ATP by the flow of protons down their electrochemical gradient (**Figure 3.3.2**).

Mitochondria are the only organelle in mammalian cells with their own genome. The human mitochondrial DNA (mtDNA) contains 37 genes. These genes encode for 13 proteins that are subunits of complexes I, III, IV and V of the electron transport chain, in addition to two mitochondrial- specific ribosomal RNAs (rRNAs) and 22 mitochondrial-specific transfer

Figure 3.3.1. Map of the human mitochondrial genome. The genes that encode the subunits of the electron transport chain are shown in colors. The two ribosomal RNAs (rRNAs, 12S and 16S) are shown in purple. The 22 mitochondrial tRNAs are indicated by black lines and denoted by their single letter code. The displacement loop (D-loop) is a non-coding control region that initiates mitochondrial replication. The origin of light-strand replication is shown as OL. Adapted from: Taylor et al. Nature Reviews Genetics. 6, 390 (2005).



RNAs (tRNAs) that are required to transcribe and translate, respectively, the unique mitochondrial genome into ETC complex subunits. In addition to the 13 mitochondrially-encoded ETC genes, the remainder of the ETC complex subunits are coded by over 850 nuclear encoded genes (Cotter et al. 2004). The nuclear DNA also encodes hundreds of enzymes and other proteins that participate in carbohydrate and fatty acid oxidation exclusively in the mitochondrial matrix. Thus, the full complement of proteins that function specifically in the mitochondria consists of 37 mitochondrially-encoded genes, and over 1,000 genes encoded on the nuclear DNA. Consequently, mutations or altered expression in genes residing in either genome can impair mitochondrial function (DiMauro et al. 2003).

In addition to their function as cellular energy producers, the mitochondria are also intimately involved in initiating cell death in response to oxidative stress. The mitochondrial membranes contain a number of proteins that are capable of activating cellar apoptotic pathways if released into the cytoplasm (in particular Cytochrome C). Increased permeability of the mitochondrial membrane results in leakage of these proteins in the cytosol, activating

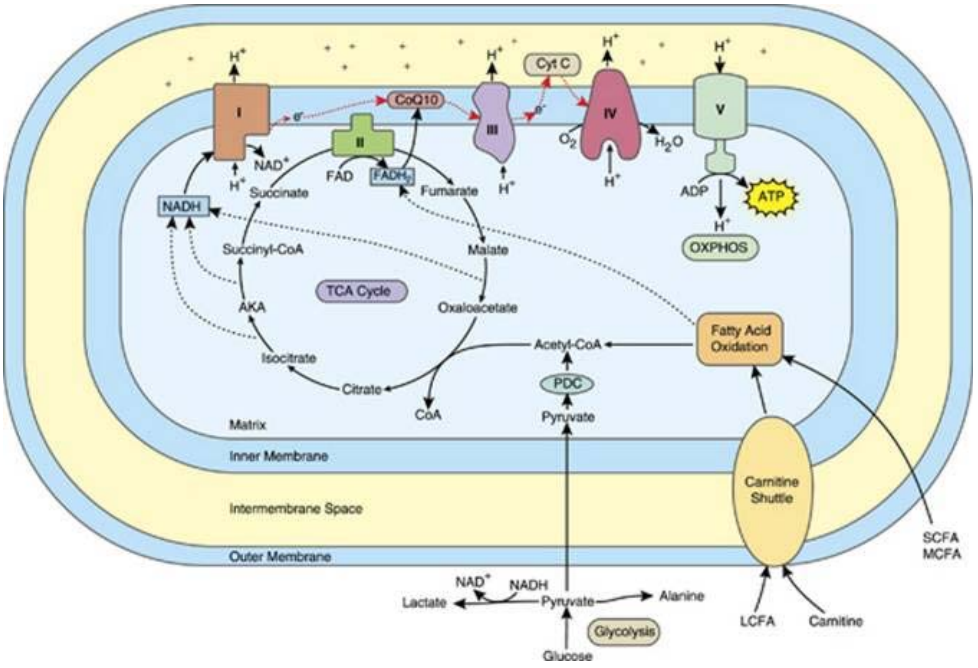


Figure 3.3.2. Schematic of mitochondrial ATP generation. The red arrows denote the flow of electrons in the ETC. Adapted from: DA Rossignol, RE Frye. Mol Psychiatry. 2012 March; 17(3): 290–314.

cell death mechanisms. The release of these proteins is tightly regulated by a family of pro- and anti-apoptotic molecules belonging to the Bcl family. The Bcl family of proteins comprise over 20 molecules that are in a delicate balance between initiating and inhibiting mitochondrially-mediated apoptosis. In response to sensing cellular insults, particularly in the form of DNA damage or protein mis-folding (both the end results of free radical oxidation), the Bcl sensors activate effector molecules that collectively cause the release of mitochondrial proteins into the cytosol. Once in the cytosol, these mitochondrial proteins (Cytochrome C in particular) activate the Caspase family of cell death enzymes, which represent a convergence point between mitochondrial and non-mitochondrial mediated cell death pathways. Capsases are able to auto-amplify the apoptotic signal, culminating in cell death through cytoskeletal breakdown and endonuclease activation (Kumar et al. 2010).

The number of mitochondria in each cell depends on the particular energy demands of that cell type. For example, skin cells, which have lower metabolic rates, have fewer mitochondria than more metabolically-demanding tissue types such as muscle, liver, and brain (Robin and Wong 1988). Neurons are cells with particularly high levels of metabolic activity and are therefore especially dependent on mitochondrial function (Ames 2000; Mattson and Liu 2002). In neurons, mitochondria are concentrated in the dendritic and axonal termini (Li et al. 2004). Their dysfunction has been shown to be involved with a

number of fundamental cellular defects, including reduced neurotransmitter release (Li et al. 2004). Consequently, the role of mitochondrially-mediated apoptosis and autophagy is becoming increasingly recognized as central to the pathogenesis of neurologic disease, in particular neurodegenerative diseases such as Alzheimer's and Parkinson's (Knott et al. 2008).

The role of mitochondrial dysfunction in autism is also becoming increasingly recognized. For instance, numerous studies have demonstrated a high proportion of children diagnosed with ASD have concurrent mitochondrial abnormalities as assessed via biochemical metabolites (Oliveira et al. 2005; Poling et al. 2006; Pastural et al. 2009). However, genetic studies of the mitochondrial DNA of autistic individuals have failed to identify significant increases in the mutation rate in ASD mtDNA (Kent et al. 2006; Kent et al. 2008; Serajee et al. 2006). Similarly, no conclusive studies have identified mutations in nuclear-encoded genes with mitochondrial function in ASD (Ramoz et al. 2004; Blasi et al. 2006).

Therefore, it is important to consider whether functional genomic aberrations in the mtDNA and nuclear-encoded mitochondrial genes may be contributing to the observed biochemical metabolic alterations in ASD, as opposed to overt DNA mutations. However, when the work in this thesis was begun, only one study had attempted to characterize mtDNA expression in ASD brain (Lepagnol-Bestel et al. 2008), and one other report assessed blood lymphocytes from individuals from ASD (Taurines et al. 2010). In both studies, only one gene was assessed (SLC25A12 and a component of complex I of the ETC, respectively). Recently, Anitha and colleagues measured the expression of 84 nuclear encoded mitochondrial genes that function in the ETC from multiple regions of autistic post-mortem brain tissue. They discovered a number of mitochondrial genes with alerted expression, many with brain-region specific expression patterns in ASD, suggesting that altered processing of mitochondrial genes may play a role in ASD (Anitha et al. 2013a; Anitha et al. 2013b).

Despite this important work linking biochemical metabolic alterations with altered mitochondrial gene expression, a comprehensive analysis of mitochondrial gene expression in ASD has not been undertaken, and assessment of the unique mitochondrially-encoded genes in particular is lacking. As mitochondrial function is very tissue-specific, it is imperative that functional genomics studies of mtDNA be carried out in brain tissue.

Therefore, the purpose of this study was to comprehensively assess the expression of all nuclear-encoded and mitochondrial-encoded genes in post-mortem autistic brain tissue, and to also assess for possible brain-region specific patterns. To do so, I performed microarray analysis of all genes with known mitochondrial function in post-mortem autistic prefrontal cortex and cerebellum samples.

3.3.3 Methods

Brain Samples and RNA Extraction

The ASD and control donor brain samples assessed in this study were full dataset described in **Table 3.0.1**. In brief, this cohort consisted of nine samples from autistic brain (5 prefrontal cortex and 4 cerebellum) and nine age and gender matched controls from the same brain regions. Total RNA was extracted using Trizol Reagent and purified using RNeasy kit (Qiagen), as previously described in Chapter 3.2.

hMitChip3 Microarray and Data Analysis

A human mitochondria-focused cDNA microarray (hMitChip3) was designed for this experiment and run by the service provider as previously described (Bai et al. 2007). The hMitChip3 contained all 37 mitochondrial–encoded genes, 1,098 nuclear–encoded mitochondrial genes, and 225 controls. Each gene was printed in triplicate. A total of 1,135 mitochondria-related genes are assessed by the array.

A total of 5ug RNA per sample was used for microarray labeling and hybridization. Slides were scanned using the ScanArray Express Microarray Scanner (Perkin-Elmer). Every gene on the hMitChip3 gene chip was printed in triplicate, and triplicate microarray experiments were done for every RNA sample, resulting in at least nine data points for each gene analyzed from each brain tissue sample.

The gene expression database was constructed using FileMaker software (FileMaker Pro, Inc.). Database construction, data filtering, and selection were done as described previously (Bai et al. 2007). The quantile normalization method (Bolstad et al. 2003) was used to normalize the microarray data. The normalized expression data were clustered and visualized

using Cluster version 3.0 (Eisen et al. 1998), and heat maps were generated by using MapleTree 6 software.

Statistics

Differentially expressed genes were identified by assessing for fold change differences in the average expression of the background-subtracted mean intensity ratios of a gene between autism and control samples. Using the LIMMA package in R/Bioconductor software suite (version 2.7.1; The R Foundation for Statistical Computing), I calculated the moderated t-statistic, raw P values, and False Discovery Rate (FDR). Genes were considered to be significantly differentially expressed if the absolute value of the log2 fold change was greater than 1.20 and the FDR q-value was less than 0.05. Sets of nuclear-encoded mitochondrial genes found to be differentially expressed in autistic brain were further analyzed for their functional ontologies using the Database for Annotation, Visualization and Integrated Discovery v6.7 (Huang et al. 2009). Only gene ontologies with a FDR multiple testing correction value less than 0.05 were considered significant.

3.3.4 Results

I discovered 7 mtDNA genes that were significantly differentially expressed in the prefrontal cortex of ASD samples as compared to controls, representing 19% of the mitochondrial genome (**Table 3.3.1**). Interestingly, all mtDNA genes that were up-regulated in ASD prefrontal cortex encode for tRNAs, while the only down-regulated mitochondrially-encoded gene in ASD prefrontal cortex was Cytochrome B. In the cerebellum, I discovered 21 differentially expressed mtDNA genes in ASD as compared to controls, representing 57% of the mitochondrial genome (**Table 3.3.2**). Five of the six tRNAs that were found to be up-regulated in ASD prefrontal cortex were also found to be up-regulated in ASD cerebellum, in addition to ten other tRNAs that showed increased expression in ASD cerebellum. Furthermore, Cytochrome B was also identified as differentially expressed in ASD cerebellum, although the direction of change was opposite that of the prefrontal cortex. In addition, a number of the components of the NADH dehydrogenase complex were up-regulated in ASD cerebellum, which was not observed in ASD prefrontal cortex.

Table 3.3.1. Differentially expressed mtDNA genes in the prefrontal cortex of ASD. Red denotes decreased expression in ASD and green denotes increased expression in ASD. FC, fold change; FDR, false discovery rate.

Mitochondrially Encoded					
OfficialGeneFullName	*GeneID*	*OfficialGeneSymbol*	*Ttest*	*FDR*	*FC*
mitochondrially encoded cytochrome b	4519	MT-CYB	7.57E-04	3.56E-03	0.75
mitochondrially encoded tRNA asparagine	4570	mt-Tn	1.72E-04	1.34E-03	1.99
mitochondrially encoded tRNA tyrosine	4579	mt-Ty	7.11E-04	3.36E-03	1.61
mitochondrially encoded tRNA cysteine	4511	mt-Tc	8.39E-04	3.84E-03	1.70
mitochondrially encoded tRNA methionine	4569	mt-Tm	9.88E-04	4.18E-03	1.24
mitochondrially encoded tRNA glutamic acid	4556	mt-Te	1.36E-03	5.11E-03	1.78
mitochondrially encoded tRNA isoleucine	4565	mt-Ti	2.16E-02	3.46E-02	1.21

Table 3.3.2. Differentially expressed mtDNA genes in the cerebellum of ASD. Green denotes increased expression in ASD. FC, fold change; FDR, false discovery rate.

Mitochondrially Encoded					
OfficialGeneFullName	*GeneID*	*OfficialGeneSymbol*	*Ttest*	*FDR*	*FC*
mitochondrially encoded tRNA tryptophan	4578	mt-Tw	3.07E-02	4.21E-02	1.25
mitochondrially encoded tRNA methionine	4569	mt-Tm	3.38E-02	4.53E-02	1.29
mitochondrially encoded NADH dehydrogenase 1	4535	ND1	2.83E-03	9.07E-03	1.31
mitochondrially encoded cytochrome b	4519	MT-CYB	1.27E-02	2.34E-02	1.31
mitochondrially encoded NADH dehydrogenase 4	4538	mt-Nd4	5.86E-03	1.40E-02	1.32
mitochondrially encoded tRNA lysine	4566	mt-Tk	4.51E-04	2.52E-03	1.40
mitochondrially encoded tRNA arginine	4573	mt-Tr	1.99E-02	3.22E-02	1.42
mitochondrially encoded NADH dehydrogenase 2	4536	mt-Nd2	1.41E-03	5.55E-03	1.43
mitochondrially encoded tRNA leucine 2 (CUN)	4568	TRNL2	4.66E-03	1.24E-02	1.43
mitochondrially encoded tRNA serine 2 (AGU/C)	4575	mt-Ts2	1.74E-03	6.48E-03	1.45
mitochondrially encoded tRNA isoleucine	4565	mt-Ti	1.49E-03	5.74E-03	1.52
mitochondrially encoded NADH dehydrogenase 5	4540	ND5	1.12E-03	4.72E-03	1.58
mitochondrially encoded NADH 4L	4539	mt-Nd4l	8.20E-05	8.20E-04	1.59
mitochondrially encoded tRNA threonine	4576	mt-Tt	5.17E-09	3.67E-06	1.72
mitochondrially encoded tRNA proline	4571	mt-Tp	3.65E-03	1.07E-02	1.73
mitochondrially encoded tRNA histidine	4564	mt-Th	1.29E-06	6.54E-05	1.89
mitochondrially encoded tRNA tyrosine	4579	mt-Ty	1.11E-05	2.38E-04	1.95
mitochondrially encoded tRNA aspartic acid	4555	mt-Td	5.56E-07	3.95E-05	1.96
mitochondrially encoded tRNA glutamic acid	4556	mt-Te	1.48E-05	2.77E-04	2.19
mitochondrially encoded tRNA glutamine	4572	mt-Tq	5.40E-06	1.82E-04	2.21
mitochondrially encoded tRNA asparagine	4570	mt-Tn	5.47E-05	6.93E-04	2.55

I concurrently assessed nuclear-encoded genes with mitochondrial function in both brain regions. I discovered 506 nuclear-encoded mitochondrial genes that were differentially expressed in ASD prefrontal cortex (232 up-regulated, 274 down-regulated), and 475 that were differentially expressed in ASD cerebellum (236 up-regulated, 239 down-regulated). I then performed gene ontology analysis on the nuclear encoded genes that were up- and down-

regulated in each ASD brain region separately. The results demonstrated brain-region specific alterations in mitochondrial gene expression in ASD (**Table 3.3.3**). Nuclear-encoded mitochondrial genes that showed increased expression in ASD prefrontal cortex were highly significantly related to the process of apoptosis, whereas genes that were down-regulated in ASD prefrontal cortex were significantly related to oxidative metabolism. In contrast, genes up-regulated in ASD cerebellum were related to oxidative metabolism but not apoptosis. Nuclear-encoded mitochondrial genes with decreased expression in ASD cerebellum were highly related to fatty acid oxidation.

Table 3.3.3. Significant gene ontologies of differentially expressed nuclear-encoded mitochondrial genes.

	↑ in ASD	↓ in ASD
Prefrontal Cortex	Apoptosis FAS/TNF signaling	Oxidative Phosphorylation ETC/Respiratory Chain
Cerebellum	Oxidative Phosphorylation ETC/Respiratory Chain	Fatty Acid Oxidation Tyrosine Phosphatase

3.3.5 Discussion

These results demonstrate substantial, brain-region specific changes in both mitochondrially-encoded and nuclear-encoded genes with mitochondrial function in ASD brain. These findings suggest that altered transcription of the mitochondrial genome may partly reconcile the biochemical alterations seen in patients with ASD with the known heritable nature of autism. Furthermore, as mitochondrial DNA is inherited only through the maternal lineage, altered mtDNA transcription in ASD may represent a mechanism that contributes to the significant sex difference in prevalence of the disorder.

Of particular interest is the up-regulation of many mitochondrially-encoded tRNAs in both ASD prefrontal cortex and cerebellum. Moreover, all tRNAs identified as differentially expressed in ASD samples are significantly up-regulated, and tRNAs were overwhelming the main mtDNA gene type that was altered in ASD.

Transfer RNAs are adaptor molecules, typically 73 to 94 nucleotides in length, which mediate the transfer of information encoded in the RNA sequence to direct the amino acid sequence of proteins. They do this by carrying their specified amino acid to the ribosome based on the codon specified in the messenger RNA. Therefore, tRNAs are a critical component of protein translation. Interestingly, the work described earlier in Chapter 2.2 identified both mitochondrial function and protein synthesis as key functions disrupted by ASD candidate genes. Furthermore, although not studied in mitochondria specifically, protein translation has been demonstrated to be globally abnormal in a number of autism spectrum disorders as well as ASD mouse models (Kelleher and Bear 2008). It is interesting to speculate that abnormal protein synthesis may cause a feedback loop whereby the cell (or mitochondria) up-regulates protein synthesis machinery, such as tRNAs, in an attempt to compensate for the production of abnormal protein. The unique nature of mitochondrial genome processing—that the 37 mtDNA genes require their own tRNAs—may highlight this defect. Future studies should assess the nuclear-encoded tRNAs in ASD brain to explore this hypothesis further.

In addition, these results show up-regulation of mitochondrially-encoded NADH dehydrogenase complex genes (Complex I) and up-regulation of nuclear-encoded genes related to oxidative phosphorylation in the cerebellum of ASD samples, and an opposite trend in the prefrontal cortex of ASD samples. These results are in line with previous studies of ASD patients, which have shown marked reductions in Complex I activity in ASD. However, those studies were conducted in peripheral blood, and our results suggest that oxidative defects may be brain-region specific in ASD. Up-regulation of oxidative phosphorylation in autistic cerebellum is an unexpected finding, but may relate to the known hyperactivity and repetitive mannerisms exhibited by ASD patients.

In contrast, apoptotic mitochondrial functions were significantly up-regulated in ASD prefrontal cortex but not in ASD cerebellum. This findings is particularly intriguing in light of previously discussed studies that have demonstrated microglial/synaptic pruning abnormalities in ASD, as well as the work described in Chapter 3.2 of this thesis where I demonstrated pre-frontal cortex specific increased microglial markers in ASD. These results further support previous studies suggesting that neuronal loss ultimately contributes to the alteration of higher cognitive functions in ASD patients, and potentially serves as a common

mechanism to reconcile the metabolic findings in ASD with known synaptic and microglial abnormalities.

3.3.6 Conclusions

In summary, I report here the first assessment of expression changes in ASD brain tissue from genes encoded by the mitochondrial DNA as well as all known mitochondrially-functioning nuclear encoded genes. The results reported in this chapter demonstrate increased expression of apoptotic mitochondrial genes in the prefrontal cortex of ASD samples with concurrent decreased expression of oxidative metabolism genes. These findings appear to be brain-region specific, as they were not observed in cerebellum samples from the same ASD patients. Furthermore, these results demonstrated a global up-regulation in mitochondrial tRNA genes, perhaps reflecting known problems in protein translation among ASD patients. Future studies will need to replicate these findings in larger cohorts, but these results suggest that alterations in mitochondrial genome processing may potentially represent a point of convergence among known deficits in protein translation, metabolic alterations, and synaptic apoptosis in patients with autism spectrum disorders.

--

Chapter 4. Conclusions

4.1 Summary

I set out to address four main questions in this thesis:

(i) Are there common gene expression properties/patterns among the genes implicated in autism that may be informative of their role in ASD pathogenesis?

(ii) Do these patterns provide insight into how so many genes with different functions can all relate to the same clinical phenotype?

(iii) Are there inherent gene expression differences between the developing male and female brain that may be informative of the significant bias in ASD seen in males?

(iv) Can studies of non-coding regions of the genome in ASD help explain some of the 'missing heritability' by their regulation of ASD genes?

The work described in Chapters 2.1 and 2.2 provide the first description of ASD candidate gene expression and co-expression patterns across normal human neurodevelopment. From these two studies, my results suggest that the heterogeneous ASD candidate genes relate to each other through shared transcriptional networks and gene expression patterns to converge upon a number of mechanisms and pathways that have been independently implicated in ASD pathogenesis. Namely, my results suggest that cytokine signaling, glia, mitochondrial function, regulation of transcription/translation, and synaptogenesis are all shared pathways during normal human neurodevelopment among the heterogeneous ASD candidate genes. Moreover, I performed two analyses specifically in autistic post-mortem brain tissue to confirm that two of these less well studied pathways are indeed aberrant in ASD brain—glia (Chapter 3.2) and mitochondrial function (Chapter 3.3). Together, these results provide important insight into potential common molecular mechanisms underlying ASD, which have the potential to be informative for future diagnostic and therapeutic development studies.

Additionally, the work presented in Chapter 2.3 describes unique gene expression differences in male and female brains, and shows how these relate to these implicated ASD pathways. While most research into the sex bias in ASD prevalence has focused on either behavioral or inherited causes, the results of this chapter provide a potentially new avenue of research that

has implications not just for autism, but for other neurodevelopmental disorders with sex-biased prevalence.

Finally, Chapters 2.4 and 3.1 provide strong support for the role of miRNAs and lncRNAs, respectively, in the functional genomics of ASD. Chapter 2.4 is the largest study to date to identify differentially expressed miRNAs in human brain development, and the miRNAs identified can now be more explicitly studied in cell systems. Chapter 3.1 provided the first evidence that long non-coding RNAs may contribute to the global dysregulation of gene expression in autistic brains, and more recent studies support my initial discovery. The work of these two chapters will play an important role in bringing to light the role of ncRNAs in the functional genomics of ASD, allowing future studies to explore specific mechanistic properties of the identified ASD ncRNAs.

The work described in this thesis has a number of important limitations of note. As discussed in the introduction, the small sample sizes assessed per time period group in the studies conducted in Chapter 2 hamper the power of these studies to detect all instances of differential expression. Furthermore, while the BrainSpan database used in the Chapter 2 studies is the largest available dataset of human brain gene expression, the samples were not available to perform confirmatory PCR or other follow up studies. Consequently, it will be important for future studies to replicate these findings in other for confirmation.

The studies of human post-mortem brain tissue are also limited in their size, and in particular, the work in Chapter 3.1 is to be considered very preliminary until larger studies can confirm the results in larger sample sizes. Additionally, it is difficult to control for other clinical factors completely, such as post-mortem interval and RNA integrity, although the samples assessed were not significantly different at a group in regard to these values. Despite these limitations, however, which are likely to remain in all of post-mortem brain research, the work described above presents some of the first global assessment of ncRNA and mitochondrial gene expression in autistic brain, potentially serving as important first steps for the field of autism genetics research.

4.2 Future Perspectives

Overall, the work of this thesis, in the context of previous and recent complementary studies, serves as an important first step in moving from ASD gene discovery to ASD pathway and mechanism understanding. However, as more genes are implicated in ASD it will be important for future work to replicate and expand upon these approaches to ensure that the most information about convergent ASD gene function is captured. Furthermore, novel analytical techniques to probe the relationship among ASD candidate genes should be undertaken. For instance, by integrating genome-wide methylation or other epigenetic marks with gene expression data both from normal and autistic post-mortem brains, a multi-level, systems biology approach can potentially provide novel insight into other functional genomics layers underlying ASD. Newer technologies and large consortia should allow such studies to be completed in the near future.

The results presented here suggest that through integration of the many layers of genetic, functional genomic, and epigenetic information that is now possible to obtain, the seemingly heterogeneous and complex nature of ASD genetics will be exposed to consist of a few broad pathological patterns, allowing for the development of targeted diagnostic and therapeutics for individuals with ASD. The work in this thesis provides some of the first evidence that supports this notion, and represents an important first step in understanding the etiology of ASD at this level.

--

Appendix

Theoretical hypothesis on the role of the cerebellum in autism

In the September 2012 issue of *The Cerebellum*, Fatemi et al. presented a comprehensive literature analysis of the putative role of the cerebellum in autism pathogenesis (Fatemi et al. 2012). While this is an important work, which synthesizes the main findings of cerebellar research in autism spectrum disorders, I believe there is an alternative hypothesis to the role of the cerebellum in autism that is more parsimonious.

The conclusion that the cerebellum is pathogenic in ASD is predicated on the notion that the cerebellum functions in the cognitive processes disrupted in autism, although such pathways remain undiscovered. While the cerebellar contribution to higher cognition has been debated for decades (Frings et al. 2007), a clear mechanistic understanding of how the cerebellum may integrate with processes affected in autism, such as theory of mind, is not well established—as Fatemi et al. noted. Human studies that have consistently implicated the cerebellum in ASD do so mostly on the basis of volumetric imaging studies, or postmortem histologic and molecular changes, including my own work (Ziats and Rennert 2013). However, as opposed to the notion that these changes are pathogenic—which would require an as yet undiscovered mechanism for the cerebellum in the higher cognitive functions affected in ASD—I propose instead that the unique anatomy, physiology, and development of the cerebellum may result in an exaggerated manifestation of the brain-wide pathologic changes that underlie autism, without being causal for the clinical phenotype. In this sense, then, the cerebellum in autism may be acting as an "anatomical beacon" of more subtle changes in other brain regions where the functional pathology actually rests.

The unique anatomy, physiology, and development of the cerebellum make it a distinct part of the human brain. The cerebellum has the highest cell density of any brain area, approximately four times that of the neocortex (Herculano-Houzel 2009, 2012), and cerebellar Purkinje cells have more synapses than any other cell type by orders of magnitude (Kandel et al. 2012). As building synapses requires the appropriate molecular "toolkit," the cerebellum's molecular complexity of transcripts (Mazin et al. 2013; Kang et al. 2011) and proteins (Fountoulakis et al. 2002; Martins-de-Souza et al. 2012) rivals that of the cerebral

cortex. Underlying the heightened synaptogenesis of the cerebellum is the need for energy to carry out this process, resulting in oxidative metabolic demand that is similar to the cerebral cortex as well (Howarth et al. 2012). The implications of these well-recognized cerebellar properties to autism are profound. The ASD phenotype is considered to ultimately result from synaptic dysfunction (Zoghbi 2003), which derives from underlying genetic changes that manifest in aberrant RNA and protein production (Voineagu 2012; Maurer 2012). Additionally, autism has a strong and growing association with related problems in oxidative metabolism (Rossignol and Frye 2012). Is it possible that cerebellar pathology in ASD is more evident than other brain areas purely because the cerebellum contains more of the components that are disrupted in autism?

If the molecular and cellular processes that are abnormal in ASD are dysfunctional throughout the brain, then these observations suggest that the cerebellum may have properties that result in an exaggerated manifestation of ASD pathology compared to other brain regions. Therefore, I hypothesize that the cerebellum may not be etiological in the pathogenesis of autism spectrum disorders; rather its unique anatomic and physiologic properties may accentuate the mechanisms that are aberrant throughout the autistic brain. Consequently, investigations into autism pathology may be more readily observed in the cerebellum because the changes are more obvious than the concomitant changes in other brain areas responsible for the clinical phenotype.

This hypothesis does not diminish the potential importance of the cerebellum to autism research. Harnessing this unique property has serious implications in diagnostic testing, for example with neuroimaging. Diagnostic tests may be able to identify biological changes in ASD patients earlier in life, which is known to correlate with improved patient outcomes (Howlin et al. 2009; Levy et al. 2009) by focusing on the cerebellum. While cerebellar changes may not directly cause the cognitive deficits of ASD, they could serve as an "internal biomarker" for the more subtle alterations that must therefore be ongoing in other brain areas but would require more sensitive techniques to detect.

Until it is understood how the cerebellum functions in the higher cognitive processes that are abnormal in autism, the field must consider the alternative hypothesis that changes found in the cerebellum of autistic patients are not pathogenic, but rather are collateral manifestations

of the cellular and molecular deficits that are present throughout the autistic brain. The distinctive nature of the cerebellum may exaggerate changes that are more subtle in other brain areas, without being causal of the ASD phenotype. However, such an interpretation does not diminish the importance of cerebellar research in autism, as this unique characteristic may make the cerebellum an ideal diagnostic target.

--

Detailed Description of Work Performed in Collaboration

All writing and work described in this thesis is the result of my own independent effort, under the guidance of my thesis mentors, with the following two exceptions:

- Chapter 2.2 was performed in direct collaboration with Ahmed Mahfouz, a PhD student at Delft University of Technology, The Netherlands. Ahmed and I met at a short course run by the Allen Institute for Brain Science. We collaborated to develop all aspects of the work described in Chapter 2.2 from the conception of the project to writing of the manuscript, with equal contribution from each of us. Specifically, the general conception for the project was a joint effort. I was responsible for processing the gene expression information, compiling the autism candidate gene list, and constructing a database of ASD candidate gene expression. Ahmed wrote and executed the Matlab script that performed correlation analysis on this dataset. I then performed the gene ontology analysis and manually curated the results for biological relevance to autism. We both were responsible for creating the figures in this Chapter (I created Figures 1, 2, and 9 while Ahmed created Figures 3-8 as part of the Matlab script), and we worked in collaboration to interpret the results. The writing in Chapter 2.2 I did myself with input from Ahmed on the methods section.

- Chapter 3.3 was performed in collaboration with Catherine Edmonson, a MD student at University of Florida School of Medicine. Catherine worked in the laboratory of one of my PhD thesis mentors, Dr. Owen Rennert, for a period of time. The work described in this Chapter was a result of equal contributions between Catherine and myself. Specifically, I was responsible for the conception of the experiment and the experimental design. Catherine performed the qRT-PCR experiments. We jointly analyzed and interpreted the results. I was responsible for creating the figures and writing up the manuscript.

Additional Tables and Figures

Table A1. Demographic information of donor brains in the Brainspan atlas used in this analysis. All information that was from: www.brainspan.org is presented.

Donor # for this Study	Allen Institute Donor ID	Allen Institute Donor Description	Donor Age at Death	Donor Gender	Donor Ethnicity	Brain pH	PMI (hours)
1	12287	H376.IIIB.50	16 pcw	M	H	unknown	2
2	12837	H376.IIIB.51	16 pcw	M	E	6.84	1
3	12879	H376.IIIB.52	16 pcw	M	A/E	6.44	1
4	12880	H376.IIIB.53	17 pcw	F	E	unknown	2
5	12885	H376.IV.53	19 pcw	F	H	unknown	2
6	12365	H376.IV.51	21 pcw	F	E	6.61	20
7	12886	H376.IV.54	21 pcw	M	A	6.65	4
8	12288	H376.IV.50	24 pcw	M	E	6.58	2
9	12296	H376.VI.50	4 mos	M	E	6.60	22
10	12889	H376.VI.51	4 mos	M	A	6.70	20
11	12890	H376.VI.52	6 mos	M	E	6.26	26
12	12977	H376.VII.51	10 mos	M	A	5.96	18
13	12979	H376.VIII.53	2 yrs	F	E	6.30	12
14	12836	H376.VIII.52	3 yrs	F	E	6.03	8
15	12980	H376.VIII.54	3 yrs	M	H	5.69	16
16	12298	H376.VIII.50	4 yrs	M	A	6.52	20
17	12841	H376.IX.51	8 yrs	M	A	6.54	30
18	12981	H376.IX.52	8 yrs	M	A	6.25	16
19	12289	H376.IX.50	11 yrs	F	A	6.70	22
20	12831	H376.X.51	13 yrs	F	A	6.34	19.5
21	12299	H376.X.50	15 yrs	M	A	6.93	14.5
22	12984	H376.X.53	18 yrs	M	E	6.21	28
23	12832	H376.X.52	19 yrs	F	E	5.91	9.5
24	13057	H376.XI.60	21 yrs	F	E	6.81	18
25	12300	H376.XI.50	23 yrs	M	A	6.36	10.5
26	12290	H376.XI.52	30 yrs	F	E	6.92	9.5
27	12302	H376.XI.53	36 yrs	M	E	6.42	18
28	12303	H376.XI.54	37 yrs	M	A	6.37	13
29	12304	H376.XI.56	40 yrs	F	A	6.82	30.5
30	12305	H376.XI.55	40 yrs	M	E	6.27	28

E = European, As = Asian, A= African American, H = Hispanic, A/E = Afircan American/European, PMI = post-mortem interval

Table A2. List of Autism Candidate Genes Used in Chapter 2 Analyses.
** indicates the gene was in the AutDB database when the analysis in Chapter 2.1 was performed as described in the methods section of that Chapter..

Gene Symbol	Entrez ID	Gene Name
A2BP1**	54715	ataxin 2-binding protein 1
ABAT**	18	4-aminobutyrate aminotransferase
ABCC1	4363	ATP-binding cassette, sub-family C (CFTR/MRP), member 1
ACCN1	40	amiloride-sensitive cation channel 1, neuronal
ACSL4	2182	acyl-CoA synthetase long-chain family member 4
ADA**	100	adenosine deaminase
ADH5	128	alcohol dehydrogenase 5 (class III), chi polypeptide, pseudogene 4; alcohol dehydrogenase 5 (class III), chi polypeptide
ADORA2A**	135	adenosine A2a receptor
ADSL**	158	adenylosuccinate lyase
AFF2	2334	AF4/FMR2 family, member 2
AGAP1	116987	ArfGAP with GTPase domain, ankyrin repeat and PH domain 1
AHCYL2	23382	adenosylhomocysteinase-like 2
AHI1**	54806	Abelson helper integration site 1
ALDH5A1**	7915	aldehyde dehydrogenase 5 family, member A1
ALDH7A1	501	aldehyde dehydrogenase 7 family, member A1
ALOX5AP**	241	arachidonate 5-lipoxygenase-activating protein
ANKRD11**	29123	ankyrin repeat domain 11; hypothetical protein LOC100128265
AP1S2	653653	adaptor-related protein complex 1, sigma 2 subunit pseudogene; adaptor-related protein complex 1, sigma 2 subunit
APBA2	321	amyloid beta (A4) precursor protein-binding, family A, member 2
APC**	324	adenomatous polyposis coli
APCDD1	147495	adenomatosis polyposis coli down-regulated 1
APOE	100129500	hypothetical LOC100129500; apolipoprotein E
ARHGEF6	9459	Rac/Cdc42 guanine nucleotide exchange factor (GEF) 6
ARID1B	57492	AT rich interactive domain 1B (SWI1-like)
ARNT2**	9915	aryl-hydrocarbon receptor nuclear translocator 2
ARX**	170302	aristaless related homeobox
ASS1**	445	argininosuccinate synthetase 1
ASTN2**	23245	astrotactin 2
ATP10A**	57194	ATPase, class V, type 10A
ATP1A2	477	ATPase, Na+/K+ transporting, alpha 2 (+) polypeptide
ATP2B2	491	ATPase, Ca++ transporting, plasma membrane 2
ATP6AP1	537	ATPase, H+ transporting, lysosomal accessory protein 1
ATP6AP2	10159	ATPase, H+ transporting, lysosomal accessory protein 2
ATRX	546	alpha thalassemia/mental retardation syndrome X-linked (RAD54 homolog, S. cerevisiae)
AUTS2**	26053	autism susceptibility candidate 2

AVPR1A**	552	arginine vasopressin receptor 1A
BAIAP2	10458	BAI1-associated protein 2
BCL2	596	B-cell CLL/lymphoma 2
BCL6	604	B-cell CLL/lymphoma 6
BDNF	627	brain-derived neurotrophic factor
BLMH	642	bleomycin hydrolase
BRAF	673	v-raf murine sarcoma viral oncogene homolog B1
BZRAP1**	9256	benzodiazapine receptor (peripheral) associated protein 1
C16orf68	79091	chromosome 16 open reading frame 68
C3orf58**	205428	chromosome 3 open reading frame 58
C7orf68	29923	chromosome 7 open reading frame 68
CACNA1C**	100131098	hypothetical protein LOC100131098; calcium channel, voltage-dependent, L type, alpha 1C subunit
CACNA1G**	8913	calcium channel, voltage-dependent, T type, alpha 1G subunit
CACNA1H**	8912	calcium channel, voltage-dependent, T type, alpha 1H subunit
CACNA1I	8911	calcium channel, voltage-dependent, T type, alpha 1I subunit
CADM1**	23705	cell adhesion molecule 1
CADPS2**	93664	Ca++-dependent secretion activator 2
CALCA	796	calcitonin-related polypeptide alpha
CARHSP1	23589	calcium regulated heat stable protein 1, 24kDa
CASK	8573	calcium/calmodulin-dependent serine protein kinase (MAGUK family)
CASP3	836	caspase 3, apoptosis-related cysteine peptidase
CBS**	875	cystathionine-beta-synthase
CCDC64**	92558	coiled-coil domain containing 64
CD44**	960	CD44 molecule (Indian blood group)
CDH10**	1008	cadherin 10, type 2 (T2-cadherin)
CDH22**	64405	cadherin-like 22
CDH8	1006	cadherin 8, type 2
CDH9**	1007	cadherin 9, type 2 (T1-cadherin)
CDK14	5218	cyclin-dependent kinase 14
CDKL5**	6792	cyclin-dependent kinase-like 5
CEP290	80184	centrosomal protein 290kDa
CHD7	55636	chromodomain helicase DNA binding protein 7
CHI3L1	1116	chitinase 3-like 1 (cartilage glycoprotein-39)
CHN1	1123	chimerin (chimaerin) 1
CHRM5	1133	cholinergic receptor, muscarinic 5
CHRNA4	1137	cholinergic receptor, nicotinic, alpha 4
CHRNB2	1141	cholinergic receptor, nicotinic, beta 2 (neuronal)
CLCN4	1183	chloride channel 4
CLDN5	7122	claudin 5
CNKSR2	22866	connector enhancer of kinase suppressor of Ras 2
CNR1**	1268	cannabinoid receptor 1 (brain)
CNTN3	5067	contactin 3 (plasmacytoma associated)

CNTN4**	152330	contactin 4
CNTNAP2**	26047	contactin associated protein-like 2
CNTNAP5**	129684	contactin associated protein-like 5
COMT	1312	catechol-O-methyltransferase
CREBBP	1387	CREB binding protein
CTNNA2**	1496	catenin (cadherin-associated protein), alpha 2
CTNNA3	29119	catenin (cadherin-associated protein), alpha 3
CTSD	1509	cathepsin D
CTTNBP2**	83992	cortactin binding protein 2
CUX1	1523	cut-like homeobox 1
CYFIP1	23191	cytoplasmic FMR1 interacting protein 1
CYR61	3491	cysteine-rich, angiogenic inducer, 61
DAB1**	1600	disabled homolog 1 (Drosophila)
DAO	1610	D-amino-acid oxidase
DAPK1**	1612	death-associated protein kinase 1
DCC	1630	deleted in colorectal carcinoma
DCUN1D1**	54165	DCN1, defective in cullin neddylation 1, domain containing 1
DCX**	1641	doublecortin
DDX3X**	1654	DEAD (Asp-Glu-Ala-Asp) box polypeptide 3, X-linked
DGCR2	9993	DiGeorge syndrome critical region gene 2
DGCR6	8214	DiGeorge syndrome critical region gene 6
DHCR7**	1717	7-dehydrocholesterol reductase
DHRS3	9249	dehydrogenase/reductase (SDR family) member 3
DLG4	1742	discs, large homolog 4 (Drosophila)
DLGAP2**	9228	discs, large (Drosophila) homolog-associated protein 2
DLX1**	1745	distal-less homeobox 1
DLX2**	1746	distal-less homeobox 2
DLX5	1749	distal-less homeobox 5
DLX6	1750	distal-less homeobox 6
DMPK**	1760	dystrophia myotonica-protein kinase
DNAJC10	54431	DnaJ (Hsp40) homolog, subfamily C, member 10
DNER	92737	delta/notch-like EGF repeat containing
DOC2A	8448	double C2-like domains, alpha
DOCK4	9732	dedicator of cytokinesis 4
DPP10**	57628	dipeptidyl-peptidase 10
DPP6**	1804	dipeptidyl-peptidase 6
DRD1	1812	dopamine receptor D1
DRD3**	1814	dopamine receptor D3
DRD4	1815	dopamine receptor D4
DRD5	1816	dopamine receptor D5
DYNLT3	6990	dynein, light chain, Tctex-type 3
EGR2**	1959	early growth response 2
EHMT1	79813	euchromatic histone-lysine N-methyltransferase 1

EIF4E**	100131565	eukaryotic translation initiation factor 4E; similar to hCG1777996
ELN	2006	elastin
EN2**	2020	engrailed homeobox 2
EPHA5	2044	EPH receptor A5
EPHB6	2051	EPH receptor B6
ERBB4	2066	v-erb-a erythroblastic leukemia viral oncogene homolog 4 (avian)
ESRRB**	2103	estrogen-related receptor beta
EXOC4	60412	exocyst complex component 4
EXT1	2131	exostoses (multiple) 1
F13A1**	2162	coagulation factor XIII, A1 polypeptide
FABP5**	728729	fatty acid binding protein 5-like 2; fatty acid binding protein 5
FABP7**	2173	fatty acid binding protein 7, brain
FAM84A	653602	hypothetical LOC653602; family with sequence similarity 84, member A
FBXO33**	254170	F-box protein 33
FEZF1	389549	FEZ family zinc finger 1
FEZF2**	55079	FEZ family zinc finger 2
FGD1	2245	FYVE, RhoGEF and PH domain containing 1
FHIT**	2272	fragile histidine triad gene
FLT1**	2321	fms-related tyrosine kinase 1 (vascular endothelial growth factor/vascular permeability factor receptor)
FMR1**	2332	fragile X mental retardation 1
FOXG1	2290	forkhead box G1
FOXP1	27086	forkhead box P1
FOXP2**	93986	forkhead box P2
FRMPD4**	9758	FERM and PDZ domain containing 4
FTSJ1	24140	FtsJ homolog 1 (E. coli)
GABRA2	2555	gamma-aminobutyric acid (GABA) A receptor, alpha 2
GABRA4**	2557	gamma-aminobutyric acid (GABA) A receptor, alpha 4
GABRA5	2558	gamma-aminobutyric acid (GABA) A receptor, alpha 5
GABRB1**	2560	gamma-aminobutyric acid (GABA) A receptor, beta 1
GABRB3**	2562	gamma-aminobutyric acid (GABA) A receptor, beta 3
GABRG1	2565	gamma-aminobutyric acid (GABA) A receptor, gamma 1
GABRG3	2567	gamma-aminobutyric acid (GABA) A receptor, gamma 3
GAD1	2571	glutamate decarboxylase 1 (brain, 67kDa)
GAD2	2572	glutamate decarboxylase 2 (pancreatic islets and brain, 65kDa)
GADD45B	4616	growth arrest and DNA-damage-inducible, beta
GALNT13**	2589	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 13 (GalNAc-T13);
GALNTL4	374378	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase-like 4
GAMT	2593	guanidinoacetate N-methyltransferase

GATA3	2625	GATA binding protein 3
GCLC	2729	glutamate-cysteine ligase, catalytic subunit
GDI1	2664	GDP dissociation inhibitor 1
GFAP	2670	glial fibrillary acidic protein
GLO1**	2739	glyoxalase I
GLRA2**	2742	glycine receptor, alpha 2
GLRA3	8001	glycine receptor, alpha 3
GLRB	2743	glycine receptor, beta
GNAS**	2778	GNAS complex locus
GNPTAB	79158	N-acetylglucosamine-1-phosphate transferase, alpha and beta subunits
GPD2	2820	glycerol-3-phosphate dehydrogenase 2 (mitochondrial)
GPM6B	2824	glycoprotein M6B
GPR139	124274	G protein-coupled receptor 139
GPR173	54328	G protein-coupled receptor 173
GPX1**	2876	glutathione peroxidase 1
GRIA2	2891	glutamate receptor, ionotropic, AMPA 2
GRIA3	2892	glutamate receptor, ionotrophic, AMPA 3
GRIK2**	2898	glutamate receptor, ionotropic, kainate 2
GRIN2A**	2903	glutamate receptor, ionotropic, N-methyl D-aspartate 2A
GRIN2B	2904	glutamate receptor, ionotropic, N-methyl D-aspartate 2B
GRIN3B	116444	glutamate receptor, ionotropic, N-methyl-D-aspartate 3B
GRIP1	23426	glutamate receptor interacting protein 1
GRIPAP1	56850	GRIP1 associated protein 1
GRM8**	2918	glutamate receptor, metabotropic 8
GSG1L	146395	GSG1-like
GSTM1**	2944	glutathione S-transferase mu 1
GSTP1	2950	glutathione S-transferase pi 1
GTF2I	100093631	general transcription factor II, i; general transcription factor II, i, pseudogene
HAPLN4	404037	hyaluronan and proteoglycan link protein 4
HAT1	8520	histone acetyltransferase 1
HCFC1	3054	host cell factor C1 (VP16-accessory protein)
HDAC6	10013	histone deacetylase 6
HERC6	55008	hect domain and RLD 6
HIRIP3	8479	HIRA interacting protein 3
HLA-A**	3105	major histocompatibility complex, class I, A
HLA-DRB1**	3126	major histocompatibility complex, class II, DR beta 4; major histocompatibility complex, class II, DR beta 1
HMGB1	100130561	high-mobility group box 1; high-mobility group box 1-like 10
HNRNPH2**	6173	ribosomal protein L36a pseudogene 51; ribosomal protein L36a pseudogene 37; ribosomal protein L36a pseudogene 49; heterogeneous nuclear ribonucleoprotein H2 (H'); ribosomal protein L36a

HRAS**	3265	v-Ha-ras Harvey rat sarcoma viral oncogene homolog
HS3ST5**	222537	heparan sulfate (glucosamine) 3-O-sulfotransferase 5
HSD11B1**	3290	hydroxysteroid (11-beta) dehydrogenase 1
HSPA6	3311	heat shock 70kDa protein 7 (HSP70B); heat shock 70kDa protein 6 (HSP70B')
HSPB8	26353	heat shock 22kDa protein 8
HTR1B**	3351	5-hydroxytryptamine (serotonin) receptor 1B
HTR2A**	3356	5-hydroxytryptamine (serotonin) receptor 2A
HTR2C**	3358	5-hydroxytryptamine (serotonin) receptor 2C
HTR7**	3363	5-hydroxytryptamine (serotonin) receptor 7 (adenylate cyclase-coupled)
ID3	3399	inhibitor of DNA binding 3, dominant negative helix-loop-helix protein
IFI16	3428	interferon, gamma-inducible protein 16
IFITM3	10410	interferon induced transmembrane protein 3 (1-8U)
IGF2	3481	insulin-like growth factor 2 (somatomedin A); insulin; INS-IGF2 readthrough transcript
IMMP2L**	83943	IMP2 inner mitochondrial membrane peptidase-like (S. cerevisiae)
ING3	54556	inhibitor of growth family, member 3
INPP1**	3628	inositol polyphosphate-1-phosphatase
IQSEC2	23096	IQ motif and Sec7 domain 2
IRAK1	3654	interleukin-1 receptor-associated kinase 1
JARID2	3720	jumonji, AT rich interactive domain 2
JMJD1C**	221037	jumonji domain containing 1C
JMJD6	23210	jumonji domain containing 6
KCND1	3750	potassium voltage-gated channel, Shal-related subfamily, member 1
KCND2	3751	potassium voltage-gated channel, Shal-related subfamily, member 2
KCNMA1**	3778	potassium large conductance calcium-activated channel, subfamily M, alpha member 1
KDM4C	23081	lysine (K)-specific demethylase 4C
KDM5C	8242	lysine (K)-specific demethylase 5C
KIAA1586**	57691	KIAA1586
KLHL21	9903	kelch-like 21 (Drosophila)
KRAS	3845	v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog
L1CAM	3897	L1 cell adhesion molecule
LAMB1**	3912	laminin, beta 1
LAMC3	10319	laminin, gamma 3
LEMD3	23592	LEM domain containing 3
LHFPL3	375612	lipoma HMGIC fusion partner-like 3
LRFN5**	145581	leucine rich repeat and fibronectin type III domain containing 5
LRRC1**	55227	leucine rich repeat containing 1
LRRN3	54674	leucine rich repeat neuronal 3
LRRTM3	347731	leucine rich repeat transmembrane neuronal 3

LRRTM4	80059	leucine rich repeat transmembrane neuronal 4
LZTS2**	84445	leucine zipper, putative tumor suppressor 2
MACROD2**	140733	MACRO domain containing 2
MAOA**	4128	monoamine oxidase A
MAOB	4129	monoamine oxidase B
MAP1A	4130	microtubule-associated protein 1A
MAP2**	4133	microtubule-associated protein 2
MAP2K1	5604	mitogen-activated protein kinase kinase 1
MAP2K3	5606	mitogen-activated protein kinase kinase 3
MAPK1	5594	mitogen-activated protein kinase 1
MAPK3	5595	hypothetical LOC100271831; mitogen-activated protein kinase 3
MARK1**	4139	MAP/microtubule affinity-regulating kinase 1
MBD1**	4152	methyl-CpG binding domain protein 1
MBD2	8932	methyl-CpG binding domain protein 2
MBD3**	53615	methyl-CpG binding domain protein 3
MBD4**	8930	methyl-CpG binding domain protein 4
MBD5	55777	methyl-CpG binding domain protein 5
MCF2	4168	MCF.2 cell line derived transforming sequence
MCM7	4176	minichromosome maintenance complex component 7
MCPH1**	79648	microcephalin 1
MDGA2**	161357	MAM domain containing glycosylphosphatidylinositol anchor 2
MECP2**	4204	methyl CpG binding protein 2 (Rett syndrome)
MED12**	9968	mediator complex subunit 12
MEF2C**	4208	myocyte enhancer factor 2C
MET**	4233	met proto-oncogene (hepatocyte growth factor receptor)
MFSD6	54842	major facilitator superfamily domain containing 6
MID1	4281	midline 1 (Opitz/BBB syndrome)
MIF	4282	macrophage migration inhibitory factor (glycosylation-inhibiting factor)
MKKS	8195	McKusick-Kaufman syndrome
MTHFR	4524	5,10-methylenetetrahydrofolate reductase (NADPH)
MTNR1A	4543	melatonin receptor 1A
MYH11	4629	myosin, heavy chain 11, smooth muscle
MYO16**	23026	myosin XVI
MYO1D	4642	myosin ID
NBEA**	26960	neurobeachin
NCAM1	4684	neural cell adhesion molecule 1
NDN	4692	necdin homolog (mouse)
NDNL2**	56160	necdin-like 2
NDP	4693	Norrie disease (pseudoglioma)
NDUFA5	4698	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 5, 13kDa
NF1**	4763	neurofibromin 1
NFIX	4784	nuclear factor I/X (CCAAT-binding transcription factor)
NGF	4803	nerve growth factor (beta polypeptide)

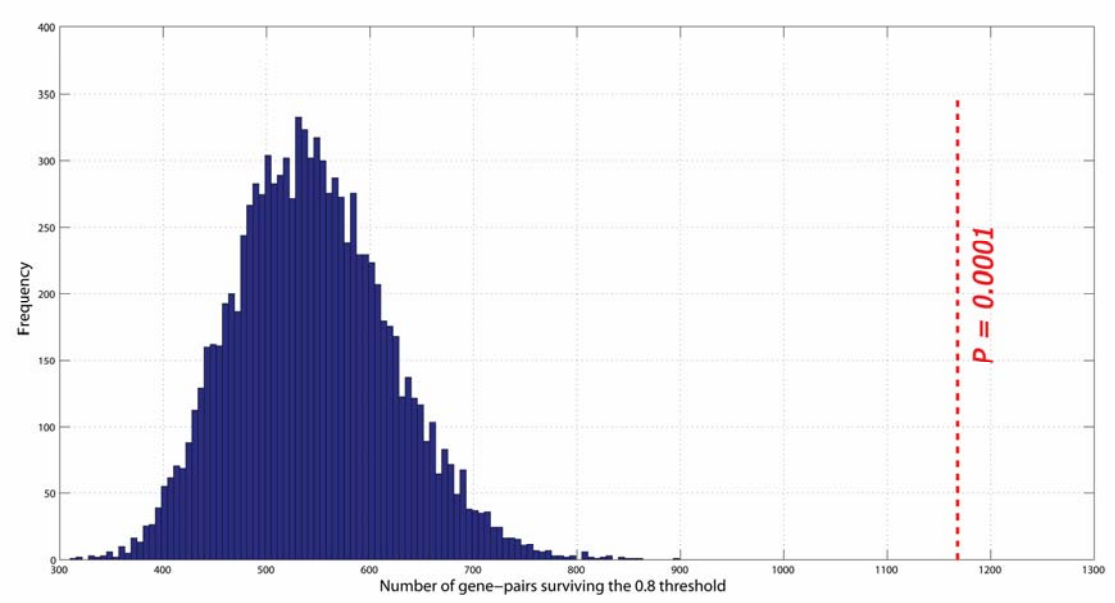
NIPA1	123606	non imprinted in Prader-Willi/Angelman syndrome 1
NIPA2	81614	non imprinted in Prader-Willi/Angelman syndrome 2
NIPBL	25836	Nipped-B homolog (Drosophila)
NLGN1**	22871	neuroligin 1
NLGN3**	54413	neuroligin 3
NLGN4X**	57502	neuroligin 4, X-linked
NLGN4Y**	22829	neuroligin 4, Y-linked
NOS1	4842	nitric oxide synthase 1 (neuronal)
NOS1AP**	9722	nitric oxide synthase 1 (neuronal) adaptor protein
NOTCH2	4853	Notch homolog 2 (Drosophila)
NPAS2**	4862	neuronal PAS domain protein 2
NPTX2	4885	neuronal pentraxin II
NPY1R	4886	neuropeptide Y receptor Y1
NPY5R	4889	neuropeptide Y receptor Y5
NRCAM**	4897	neuronal cell adhesion molecule
NRP2**	8828	neuropilin 2
NRXN1**	9378	neurexin 1
NRXN2	9379	neurexin 2
NSD1	64324	nuclear receptor binding SET domain protein 1
NTF3	4908	neurotrophin 3
NTNG1**	22854	netrin G1
NTRK1**	4914	neurotrophic tyrosine kinase, receptor, type 1
NTRK3**	4916	neurotrophic tyrosine kinase, receptor, type 3
OCRL	4952	oculocerebrorenal syndrome of Lowe
OMG	4974	oligodendrocyte myelin glycoprotein
OPCML	4978	opioid binding protein/cell adhesion molecule-like
OPHN1**	4983	oligophrenin 1
OPRM1**	4988	opioid receptor, mu 1
OSBPL6	114880	oxysterol binding protein-like 6
OXT**	5020	oxytocin, prepropeptide
PAFAH1B1	5048	platelet-activating factor acetylhydrolase, isoform Ib, subunit 1 (45kDa)
PARK2**	5071	Parkinson disease (autosomal recessive, juvenile) 2, parkin
PAX3	5077	paired box 3
PAX6	5080	paired box 6
PCDH10**	57575	protocadherin 10
PCDH11X	27328	protocadherin 11 X-linked
PCDH19**	57526	protocadherin 19
PCDH9**	5101	protocadherin 9
PCYT1B	9468	phosphate cytidylyltransferase 1, choline, beta
PDE4A	5141	phosphodiesterase 4A, cAMP-specific
PDE4B	5142	phosphodiesterase 4B, cAMP-specific
PDZD4**	57595	PDZ domain containing 4
PER1**	5187	period homolog 1 (Drosophila)

PLAUR	5329	plasminogen activator, urokinase receptor
PLD5	200150	phospholipase D family, member 5
PLN**	5350	phospholamban
PLXNA3	55558	plexin A3
PLXNA4	91584	plexin A4
POMGNT1	55624	protein O-linked mannose beta1,2-N-acetylglucosaminyltransferase
PPP1R3F**	89801	protein phosphatase 1, regulatory (inhibitor) subunit 3F
PQBP1	10084	polyglutamine binding protein 1
PRKCB**	5579	protein kinase C, beta
PRKX	5613	protein kinase, X-linked
PRL	5617	prolactin
PRLR	5618	prolactin receptor
PRODH	5625	proline dehydrogenase (oxidase) 1
PRPF31	26121	pre-mRNA processing factor 31
PRRT2	112476	proline-rich transmembrane protein 2
PSD3	23362	pleckstrin and Sec7 domain containing 3
PSMD10**	5716	proteasome (prosome, macropain) 26S subunit, non-ATPase, 10
PTCHD1**	139411	patched domain containing 1
PTEN**	5728	phosphatase and tensin homolog; phosphatase and tensin homolog pseudogene 1
PTGS2**	5743	prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase)
PTPN11	442113	protein tyrosine phosphatase, non-receptor type 11; similar to protein tyrosine phosphatase, non-receptor type 11
PTPRD	5789	protein tyrosine phosphatase, receptor type, D
PTPRN2	5799	protein tyrosine phosphatase, receptor type, N polypeptide 2
PTS	5805	6-pyruvoyltetrahydropterin synthase
PUM2	23369	pumilio homolog 2 (Drosophila)
RAB11FIP5	26056	RAB11 family interacting protein 5 (class I)
RAB39B**	116442	RAB39B, member RAS oncogene family
RAI1**	10743	retinoic acid induced 1
RAPGEF4**	11069	Rap guanine nucleotide exchange factor (GEF) 4
RB1CC1**	9821	RB1-inducible coiled-coil 1
REEP3**	221035	receptor accessory protein 3
RELN**	5649	reelin
RFWD2**	64326	ring finger and WD repeat domain 2
RFX4	5992	regulatory factor X, 4 (influences HLA class II expression)
RIMS3**	9783	regulating synaptic membrane exocytosis 3
RIT1	6016	Ras-like without CAAX 1
RNF216L	441191	ring finger protein 216-like
RNF8	9025	ring finger protein 8
ROBO1**	642132	roundabout, axon guidance receptor, homolog 1 (Drosophila)

ROBO2	6092	roundabout, axon guidance receptor, homolog 2 (Drosophila)
ROBO3	64221	roundabout, axon guidance receptor, homolog 3 (Drosophila)
ROBO4	54538	roundabout homolog 4, magic roundabout (Drosophila)
RORA	6095	RAR-related orphan receptor A
RPL10**	285176	ribosomal protein L10; ribosomal protein L10 pseudogene 15
RPP25	54913	ribonuclease P/MRP 25kDa subunit
RPS6KA2**	6196	ribosomal protein S6 kinase, 90kDa, polypeptide 2; hypothetical LOC100127984
RPS6KA3	6197	ribosomal protein S6 kinase, 90kDa, polypeptide 3
RYR2	6262	ryanodine receptor 2 (cardiac)
SATB2	23314	SATB homeobox 2
SCN1A**	6323	sodium channel, voltage-gated, type I, alpha subunit
SCN2A**	6326	sodium channel, voltage-gated, type II, alpha subunit
SCT	6343	secretin
SDC2**	6383	syndecan 2
SDF2L1	23753	stromal cell-derived factor 2-like 1
SEMA4D	10507	sema domain, immunoglobulin domain (Ig), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 4D
SEMA5A**	9037	sema domain, seven thrombospondin repeats (type 1 and type 1-like), transmembrane domain and short cytoplasmic domain, (semaphorin) 5A
SERPIND1	3053	serpin peptidase inhibitor, clade D (heparin cofactor), member 1
SERPINH1	871	serpin peptidase inhibitor, clade H (heat shock protein 47), member 1, (collagen binding protein 1)
SEZ6L2**	26470	seizure related 6 homolog (mouse)-like 2
SFRS3	6428	splicing factor, arginine/serine-rich 3
SGSH	6448	N-sulfoglucosamine sulfohydrolase
SH3KBP1**	30011	SH3-domain kinase binding protein 1
SHANK1	50944	SH3 and multiple ankyrin repeat domains 1
SHANK2**	22941	SH3 and multiple ankyrin repeat domains 2
SHANK3**	85358	SH3 and multiple ankyrin repeat domains 3
SHISA9	729993	shisa family member 9
SHROOM2	357	shroom family member 2
SLC16A3	9123	solute carrier family 16, member 3 (monocarboxylic acid transporter 4)
SLC1A1**	6505	solute carrier family 1 (neuronal/epithelial high affinity glutamate transporter, system Xag), member 1
SLC25A12**	8604	solute carrier family 25 (mitochondrial carrier, Aralar), member 12
SLC25A13	10165	solute carrier family 25, member 13 (citrin)
SLC38A5	92745	solute carrier family 38, member 5
SLC4A10**	57282	solute carrier family 4, sodium bicarbonate transporter, member 10
SLC6A8**	6535	solute carrier family 6 (neurotransmitter transporter, creatine), member 8

SLC7A5	8140	solute carrier family 7, member 5
SLC9A6**	10479	solute carrier family 9 (sodium/hydrogen exchanger), member 6
SLC9A9**	285195	solute carrier family 9 (sodium/hydrogen exchanger), member 9
SLTM	79811	SAFB-like, transcription modulator
SMO	6608	smoothened homolog (Drosophila)
SNAP25	6616	synaptosomal-associated protein, 25kDa
SNRPN	6638	small nuclear ribonucleoprotein polypeptide N
SPON2	10417	spondin 2, extracellular matrix protein
SPP1	6696	secreted phosphoprotein 1
SSBP1	6742	single-stranded DNA binding protein 1
ST7**	7982	suppression of tumorigenicity 7
ST8SIA2	8128	ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 2
STK33	65975	serine/threonine kinase 33
STK39**	27347	serine threonine kinase 39 (STE20/SPS1 homolog, yeast)
STS	412	steroid sulfatase (microsomal), isozyme S
STX1A	6804	syntaxin 1A (brain)
SUCLG2**	8801	similar to sucb; succinate-CoA ligase, GDP-forming, beta subunit
SYN1	6853	synapsin I
SYNGAP1	8831	synaptic Ras GTPase activating protein 1 homolog (rat)
SYPL1	6856	synaptophysin-like 1
SYT17**	51760	synaptotagmin XVII; synaptotagmin VII
TAC1	6863	tachykinin, precursor 1
TAF1C	9013	TATA box binding protein (TBP)-associated factor, RNA polymerase I
TAOK2	9344	TAO kinase 2
TBR1	10716	T-box, brain, 1
TCN2	6948	transcobalamin II; macrocytic anemia
TFPI	7035	tissue factor pathway inhibitor (lipoprotein-associated coagulation inhibitor)
TGM3	7053	transglutaminase 3 (E polypeptide, protein-glutamine-gamma glutamyltransferase)
TH**	7054	tyrosine hydroxylase
TIMP1	7076	TIMP metallopeptidase inhibitor 1
TLE2	7089	transducin-like enhancer of split 2 (E(sp1) homolog, Drosophila)
TLK1	9874	tousled-like kinase 1
TLK2	11011	tousled-like kinase 2
TMEM130**	222865	transmembrane protein 130
TMEM47	83604	transmembrane protein 47
TMEM98	440181	similar to transmembrane protein 98; transmembrane protein 98
TRO	7216	trophinin
TSC1**	7248	tuberous sclerosis 1
TSC2**	7249	tuberous sclerosis 2
TSGA14	95681	testis specific, 14
TSN**	7247	translin

TSPAN12	23554	tetraspanin 12
TSPAN7**	7102	tetraspanin 7
TUBGCP5	114791	tubulin, gamma complex associated protein 5
UBE2H**	7328	ubiquitin-conjugating enzyme E2H (UBC8 homolog, yeast)
UBE2L3	7332	ubiquitin-conjugating enzyme E2L 3
UBE3A**	7337	ubiquitin protein ligase E3A
UPF3B	65109	UPF3 regulator of nonsense transcripts homolog B (yeast)
USP9X	8239	ubiquitin specific peptidase 9, X-linked
VASH1**	22846	vasohibin 1
VIP	7432	vasoactive intestinal peptide
VIPR2	7434	vasoactive intestinal peptide receptor 2
WNK3**	65267	WNK lysine deficient protein kinase 3
WNT2**	7472	wingless-type MMTV integration site family member 2
XPC**	7508	xeroderma pigmentosum, complementation group C
YWHAB	7529	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein
YWHAE	440917	similar to 14-3-3 protein epsilon (14-3-3E)
ZNF214	7761	zinc finger protein 214
ZNF385B	151126	zinc finger protein 385B
ZNF498	221785	zinc finger protein 498
ZNF622	90441	zinc finger protein 622



Supplementary Figure A3. Distribution plot of the number of strongly correlated gene-pairs per gene set from Chapter 2.2. The distribution of the number of gene-pairs remaining after applying the threshold (absolute correlation > 0.8 at any developmental stage) shows that that the number of strongly correlated gene-pairs from the the ASD list (dashed red line) is highly significant (p = 10-4). Blue bars correspond to the 10,000 random gene sets analyzed.

References

Abbott A: Tissue-bank shortage: Brain child. *Nature*. 2011, 478(7370):442-3.

Abdallah MW, Larsen N, Grove J, Nørgaard-Pedersen B, Thorsen P, Mortensen EL, Hougaard DM. Amniotic fluid chemokines and autism spectrum disorders: an exploratory study utilizing a Danish Historic Birth Cohort. *Brain Behav Immun*. 2012; 26(1):170-6.

Abel KM, Drake R, Goldstein JM. Sex differences in schizophrenia. *Int Rev Psychiatry* 2010; 22: 417–428.

Abelson JF, Kwan KY, O'Roak BJ, Baek DY, Stillman AA, Morgan TM, *et al*. Sequence variants in SLITRK1 are associated with Tourette's syndrome. *Science*. 2005; 310(5746):317-20.

Abrahams BS, Arking DE, Campbell DB, Mefford HC, Morrow EM, Weiss LA, et al. SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol Autism*. 2013; 3;4(1):36.
Abrahams BS, Geschwind DH. Advances in autism genetics: on the threshold of a new neurobiology. *Nat Rev Genet*. 2008; 9:341–355.

Abrahams BS, Geschwind DH. Connecting genes to brain in the autism spectrum disorders. *Arch Neurol*. 2010; 67:395–399.

Abu-Elneel K, Liu T, Gazzaniga FS, *et al*. Heterogeneous dysregulation of microRNAs across the autism spectrum. *Neurogenetics*. 2008; 3:153–161.

Akula N, Barb J, Jiang X, Wendland JR, Choi KH, Sen SK, et al. RNA-sequencing of the brain transcriptome implicates dysregulation of neuroplasticity, circadian rhythms and GTPase binding in bipolar disorder. *Mol Psychiatry*. 2014 Jan 7. [Epub ahead of print]

Alkan C, Coe BP, Eichler EE. Genome structural variation discovery and genotyping. *Nat. Rev. Genet.* 2011; 12:363–76.

Allen NC, Bagade S, McQueen MB, Ioannidis JP, Kavvoura FK, *et al*. Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: the SzGene database. *Nat Genet*. 2008; 40:827–834.

Amaral PP, Mattick JS. Noncoding RNA in development. *Mamm Genome*. 2008; 19(7-8):454-92

American Psychiatric Association. *Diagnostic and statistical manual of mental disorders, 3rd edition*. Washington, DC: American Psychiatric Publishing, 1980.

American Psychiatric Association. *Diagnostic and statistical manual of mental disorders, 4th edition.* Arlington, VA: American Psychiatric Publishing, 2000.

American Psychiatric Association. *Diagnostic and statistical manual of mental disorders, 5th ed*. Arlington, VA: American Psychiatric Publishing, 2013.

Ames A. CNS energy metabolism as related to function. *Brain Res Brain Res Rev*. 2000; 34:42–68.
Amir RE, Van den Veyver IB, Wan M, Tran CQ, Francke U, Zoghbi HY. Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat. Genet*. 1999; 23:185–88.

Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol* 2010; 11: R106.

Anderson GM, Jacobs-Stannard A, Chawarska K, Volkmar FR, Kliman HJ. Placental trophoblast inclusions in autism spectrum disorder. *Biol Psychiatry*. 2007; 61(4):487-91.

Andersson SG, Karlberg O, Canbäck B, Kurland CG. On the origin of mitochondria: a genomics perspective. *Philos Trans R Soc Lond B Biol Sci*. 2003; 358(1429):165-77.

Anitha A, Nakamura K, Thanseem I, Matsuzaki H, Miyachi T, Tsujii M, *et al*. Downregulation of the expression of mitochondrial electron transport complex genes in autism brains. *Brain Pathol.* 2013; 23(3):294-302.

Anitha A, Nakamura K, Thanseem I, Yamada K, Iwayama Y, Toyota T, *et al*. Brain region-specific altered expression and association of mitochondria-related genes in autism. *Mol Autism* 2012, 3(1):12.

Anney R, Klei L, Pinto D, Regan R, Conroy J, *et al.* A genome-wide scan for common alleles affecting risk for autism. *Hum Mol Genet*. 2010; 19:4072–4082.

Anthony TE, Mason HA, Gridley T, Fishell G, Heintz N. Brain lipid-binding protein is a direct target of Notch signaling in radial glial cells. *Genes Dev*. 2005; 19:1028–1033.

Antony JM, Paquin A, Nutt SL, Kaplan DR, Miller FD. Endogenous microglia regulate development of embryonic cortical precursor cells. *J Neurosci Res*. 2011; 89(3):286-98.

Ashwood P, Enstrom A, Krakowiak P, Hertz-Picciotto I, Hansen RL, et al. Decreased transforming growth factor beta1 in autism: a potential link between immune dysregulation and impairment in clinical behavioral outcomes. *J Neuroimmunol*. 2008; 204:149–153.

Awasaki T, Huang Y, O'Connor MB, Lee T. Glia instruct developmental neuronal remodeling through TGF-beta signaling. *Nat Neurosci*. 2011; 14:821–823.

Baba H, Nakahira K, Morita N, Tanaka F, Akita H, Ikenaka K. GFAP gene expression during development of astrocyte. *Dev Neurosci* 1997, 19(1):49-57.

Bai X, Wu J, Zhang Q, *et al*. Third-generation human mitochondria-focused cDNA microarray and its bioinformatic tools for analysis of gene expression. *Biotechniques* 2007; 42:365–75.

Bailey A, Le Couteur A, Gottesman I, Bolton P, Simonoff E, *et al*. Autism as a strongly genetic disorder: evidence from a British twin study. *Psychol Med*. 1995; 25:63–77.

Bailey A, Luthert P, Dean A, Harding B, Janota I, Montgomery M, *et al*. Clinicopathological study of autism. *Brain* 1998, 121(5):889-905.

Bailey A., Le Couteur A., Gottesman I., Bolton P, Simonoff E, Yuzda E, Rutter M. Autism as a strongly genetic disorder: evidence from a British twin study. *Psychological medicine*. 1995; 25:63-77.

Balaj L, *et al*. Tumour microvesicles contain retrotransposon elements and amplified oncogene sequences. *Nature Commun*. 2011; 2:180.

Barabási, A.-L., Gulbahce, N. and Loscalzo, J. Network medicine: a network-based approach to human disease. *Nature reviews. Genetics*. 2011; 12, 56-68.

Baron-Cohen S, Knickmeyer RC, Belmonte MK. Sex differences in the brain: implications for explaining autism. *Science*. 2005; 4:819–823.

Baron-Cohen S, Lombardo MV, Auyeung B, Ashwin E, Chakrabarti B, Knickmeyer R. Why are autism spectrum conditions more prevalent in males? *PLoS Biol*. 2011; 4:e1001081.

Baron-Cohen S. The extreme male brain theory of autism. *Trends Cogn Sci.* 2002; 4:248–254.

Baryshnikova A, Costanzo M, Kim Y, Ding H, Koh J, Toufighi K, *et al*. Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. *Nat Methods*. 2010; 7(12):1017-24.

Basu SN, Kollu R, Banerjee-Basu S. AutDB: a gene reference resource for autism research. *Nucleic Acids Res.* 2009; 37:D832–836.

Belgard TG, Marques AC, Oliver PL, Abaan HO, Sirey TM, Hoerder-Suabedissen A, *et al*. A transcriptomic atlas of mouse neocortical layers. *Neuron*. 2011; 71(4):605-16.

Ben-David E, Shifman S. Combined analysis of exome sequencing points toward a major role for transcription regulation during brain development in autism. *Mol Psychiatry*. 2013; 18(10):1054-6.

Ben-David E, Shifman S. Networks of neuronal genes affected by common and rare variants in autism spectrum disorders. *PLoS Genet.* 2012; 8(3):e1002556.

Benson PR, Kersh J. Marital quality and psychological adjustment among mothers of children with ASD: cross-sectional and longitudinal relationships. *J Autism Dev Disord*. 2011; 41(12):1675-85.

Berg J, Geschwind D. Autism genetics: searching for specificity and convergence. *Genome Biology*. 2012; 13:247.

Berglund L, Bjorling E, Oksvold P, Fagerberg L, Asplund A, *et al*. A genecentric Human Protein Atlas for expression profiles based on antibodies. *Mol Cell Proteomics.* 2008; 7:2019–2027.

Bernard D, Prasanth KV, Tripathi V, Colasse S, Nakamura T, Xuan Z, *et al*. A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression. *EMBO J*. 2010; 29(18):3082-93.

Betancur C, Sakurai T, Buxbaum JD. The emerging role of synaptic cell-adhesion pathways in the pathogenesis of autism spectrum disorders. *Trends Neurosci.* 2009; 4:402–412.

Betancur C: Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. *Brain Res* 2011, 1380:42-77.

Bielle F, Griveau A, Narboux-Nême N, Vigneau S, Sigrist M, Arber S, Wassef M, Pierani A. Multiple origins of Cajal-Retzius cells at the borders of the developing pallium. *Nature Neuroscience*. 2005; 8(8): 1002-1012.

Bishop KM, Ruvenstein JL, O'Leary DD. Distinct actions of Emx1, EMX2, and Pax6 in regulating the specification of areas in the developing neocortex. *The Journal of Neuroscience*. 2002; 22(17): 7627-7638.

Blasi F, Bacchelli E, Carone S, Toma C, Monaco AP, Bailey AJ, Maestrini E. SLC25A12 and CMYA3 gene variants are not associated with autism in the IMGSAC multiplex family sample. *Eur J Hum Genet*. 2006; 14(1):123-6.

Blumberg SJ, Bramlett MD, Kogan MD, *et al*. Changes in prevalence of parent-reported autism spectrum disorder in school-aged U.S. children: 2007 to 2011–2012. *National health statistics reports No 65*. Hyattsville, MD: National Center for Health Statistics. 2013.

Boksa P. Effects of prenatal infection on brain development and behavior: a review of findings from animal models. *Brain Behav Immun*. 2010; 24(6):881-97.

Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 2003; 19:185–93.

Bolton MM, Eroglu C. Look who is weaving the neural web: glial control of synapse formation. *Curr Opin Neurobiol*. 2009; 19:491–497.

Bonati MT, Russo S, Finelli P, *et al*. Evaluation of autism traits in Angelman syndrome: A resource to unfold autism genes. *Neurogenetics*. 2007; 8:169–178.

Bond AM, Vangompel MJ, Sametsky EA, Clark MF, Savage JC, Disterhoft JF, Kohtz JD. Balanced gene regulation by an embryonic brain ncRNA is critical for adult hippocampal GABA circuitry. *Nat Neurosci*. 2009; 12(8):1020-7.

Borrell V, Götz M. Role of radial glial cells in cerebral cortex folding. *Curr Opin Neurobiol*. 2014; 10;27C:39-46.

Bourgeron, T. A synaptic trek to autism. *Current opinion in neurobiology*. 2009; 19, 231-234.

Braunschweig D, Ashwood P, Krakowiak P, Hertz-Picciotto I, Hansen R, *et al*. Autism: maternally derived antibodies specific for fetal brain proteins. *Neurotoxicology*. 2008; 29:226–231.

Brown M, Keynes R, Lumsden A. The developing brain. Oxford: Oxford University Press, 2002.

Bullmore E, Sporns O. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci.* 2009; 10(3):186-98.

Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES *et al*. Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci* 2013; 14: 365–376.

Buxbaum JD, Silverman JM, Smith CJ, Greenberg DA, Kilifarski M, et al. Association between a GABRB3 polymorphism and autism. *Mol Psychiatry*. 2002; 7:311–316.

Bystron I, Blakemore C, Rakic P. Development of the human cerebral cortex: boulder committee revised. *Nat Rev Neurosci*. 2008; 9(2):110-112.

Caceres M., Lachuer J, Zapala MA, Redmond JC, Kudo L, Geschwind DH, *et al*. Elevated gene expression levels distinguish human from non-human primate brains. *PNAS*. 2003; 100:13030–13035.

Cahoy J, Emery B, Kaushal A, Foo L, Zamanian J, Christopherson K, *et al.* A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J Neurosci.* 2008; 28, 264-278.

Calin GA, Liu CG, Ferracin M, Hyslop T, Spizzo R, Sevignani C, *et al*. Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas. *Cancer Cell.* 2007; 3:215-229.

Cameron JS, Alexopoulou L, Sloane JA, DiBernardo AB, Ma Y, Kosaras B, et al. Toll-like receptor 3 is a potent negative regulator of axonal growth in mammals. *J Neurosci*. 2007; 27(47):13033-41.

Cao F, Yin A, Wen G, Sheikh AM, Tauqeer Z, Malik M, *et al*. Alteration of astrocytes and Wnt/β-catenin signaling in the frontal cortex of autistic subjects. *J Neuroinflammation*. 2012; 9(1):223.

Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, *et al*. The transcriptional landscape of the mammalian genome. *Science*. 2005; 309(5740):1559-63.

Cardona AE, Pioro EP, Sasse ME, Kostenko V, Cardona SM, Dijkstra IM, et al. Control of microglial neurotoxicity by the fractalkine receptor. *Nat Neurosci*. 2006; 9(7):917-24.

Carper, R.A. and Courchesne, E. Localized enlargement of the frontal cortex in early autism. *Biological psychiatry*. 2005; 57, 126-133.

Casanova MF. The neuropathology of autism. *Brain Pathol*. 2007; 17(4):422-33.

Cayre M, Canoll P, Goldman JE. Cell migration in the normal and pathological postnatal mammalian brain. *Progress in Neurobiology*. 2009; 88(1): 41-63.

Centers for Disease Control and Prevention: Prevalence of Autism Spectrum Disorders — Autism and Developmental Disabilities Monitoring Network, 14 Sites, United States, 2008. *MMWR* 2012, 61(3): 1-24.

Chahrour M, Jung SY, Shaw C, Zhou X, Wong ST, *et al*. MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science*. 2008; 320:1224–1229.

Champagne FA. Epigenetic mechanisms and the transgenerational effects of maternal care. Front Neuroendocrinol. 2008; 29:386–397.

Chen K, Rajewsky N. The evolution of gene regulation by transcription factors and microRNAs. *Nat Rev Genet* 2007; 8: 93–103.

Chi JG, Dooling EC, Gilles FH. Gyral development of the human brain. Annals of Neurology. 1977; 1(1): 86-93.

Ching MS, Shen Y, Tan WH, Jeste SS, Morrow EM, *et al*. Deletions of NRXN1 (neurexin-1) predispose to a wide spectrum of developmental disorders. *Am J Med Genet B Neuropsychiatr Genet*. 2010; 153B:937–947.

Chodroff RA, Goodstadt L, Sirey TM, Oliver PL, Davies KE, Green ED, *et al*. Long noncoding RNA genes: Conservation of sequence and brain expression among diverse amniotes. *Genome Biology*. 2010; 11:R72.

Chow ML, Pramparo T, Winn ME, Barnes CC, Li HR, Weiss L, *et al*. Age-dependent brain gene expression and copy number anomalies in autism suggest distinct pathological processes at young versus mature ages. *PLoS Genet*. 2012; 8(3):e1002592.

Christian SL, Brune CW, Sudi J, Kumar RA, Liu S, *et al*. Novel submicroscopic chromosomal abnormalities detected in autism spectrum disorder. *Biol. Psychiatry*. 2008; 63:1111–17.

Chu Y, Corey DR. RNA sequencing: platform selection, experimental design, and data interpretation. Nucleic Acid Ther. 2012; 22(4):271-4.

Colantuoni C, Lipska BK, Ye T, Hyde TM, Tao R, Leek JT, *et al*. Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature*. 2011; 478:519–524.

Conde F, Lund JS, Jacobowitz DM, Baimbridge KG, Lewis DA. Local circuit neurons immunoreactive for calretinin, calbindin D-28k or parvalbumin in monkey prefrontal cortex: distribution and morphology. *J Comp Neurol* 1994, 341(1):95–116.

Cookson W, Liang L, Abecasis G, Moffatt M, Lathrop M. Mapping complex disease traits with global gene expression. *Nat Rev Genet*. 2009; 10(3):184-94.
Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, *et al*. A copy number variation morbidity map of developmental delay. *Nat. Genet.* 2011; 43:838–46.

Cooper JA. A mechanism for inside-out lamination in the neocortex. *Trends Neurosci* 2008; 31(3): 113-119.

Cotter D, Guda P, Fahy E, Subramaniam S. MitoProteome: mitochondrial protein sequence database and annotation system. *Nucleic Acids Res* 2004; 32D463–D467.D467.

Courchesne E, Campbell K, Solso S. Brain growth across the life span in autism: age-specific changes in anatomical pathology. *Brain Res* 2011, 1380:138-45.

Courchesne E, Mouton PR, Calhoun ME, Semendeferi K, Ahrens-Barbeau C, Hallet MJ, *et al*. Neuron number and size in prefrontal cortex of children with autism. *JAMA* 2011, 306(18):2001-10.

Courchesne E, Pierce K, Schumann CM, Redcay E, Buckwalter JA, Kennedy DP, Morgan J. Mapping early brain development in autism. *Neuron.* 2007; 56(2):399-413.

Courchesne, E. and Pierce, K. Why the frontal cortex in autism might be talking only to itself: local over-connectivity but long-distance disconnection. *Current opinion in neurobiology*. 2005; 15, 225-230.

Croft DP, Krause J, James R. Social networks in the guppy (Poecilia reticulata). *Proc Biol Sci.* 2004; 271 Suppl 6:S516-9.

Croonenberghs J, Bosmans E, Deboutte D, Kenis G, Maes M. Activation of the inflammatory response system in autism. *Neuropsychobiology*. 2002; 45:1–6.

Cunningham CL, Martínez-Cerdeño V, Noctor SC. Microglia regulate the number of neural precursor cells in the developing cerebral cortex. *J Neurosci*. 2013; 33(10):4216-33.

Davis TH, Cuellar TL, Koch SM, Barker AJ, Harfe BD, McManus MT, Ullian EM. Conditional loss of Dicer disrupts cellular and tissue morphogenesis in the cortex and hippocampus. *J Neurosci*. 2008; 28(17):4322-30.

de la Grange P, Gratadou L, Delord M, Dutertre M, Auboeuf D. Splicing factor and exon profiling across human tissues. *Nucleic Acids Research*. 2010; 38:2825–2838.

De Pietri TD, Pulvers JN, Haffner C, Murchison EP, Hannon GJ, Huttner WB. miRNAs are essential for survival and differentiation of newborn neurons but not for expansion of neural progenitors during early neurogenesis in the mouse embryonic neocortex. *Development*. 2008; 135:3911-3921.

de Winter JCF. Using the Student's t-test with extremely small sample sizes. *Practical Assessment, Research & Evaluation*. 2013; 18: 10.

Dekaban AS, Sadowsky D. Changes in brain weights during the span of human life: Relation of brain weights to body heights and body weights. *Annals of Neurology*. 1978; 4:345–356.

Delaloy C, Liu L, Lee JA, Su H, Shen F, Yang GY, et al. MicroRNA-9 coordinates proliferation and migration of human embryonic stem cell-derived neural progenitors. *Cell Stem Cell*. 2010; 6(4):323-35.

Delaval K, Feil R. Epigenetic regulation of mammalian genomic imprinting. *Curr Opin Genet Dev*. 2004; 14(2):188-95.

Derecki NC, Cronk JC, Lu Z, Xu E, Abbott SB, Guyenet PG, Kipnis J. Wild type microglia arrest pathology in a mouse model of Rett syndrome. *Nature.* 2012; 484(7392):105-9.
Derecki NC, Privman E, Kipnis J. Rett syndrome and other autism spectrum disorders–brain diseases of immune malfunction? *Mol Psychiatry*. 2010; 15:355–363.

Desai AR, McConnel SK. Progressive restriction in fate potential by neural progenitors during cerebral cortical development. *Development*. 2000; 127(13): 2863-2872.

Devlin B, Scherer SW. Genetic architecture in autism spectrum disorder. *Curr Opin Genet Dev*. 2012; 22(3):229-37.

DiMauro S, Schon EA. Mitochondrial respiratory-chain diseases. *N Engl J Med*. 2003; 348:2656–2668.

Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, *et al*. Landscape of transcription in human cells. *Nature*. 2012; 489(7414):101-8.

Dugas JC, Cuellar TL, Scholze A, Ason B, Ibrahim A, Emery B, *et al*. Dicer1 and miR-219 are required for normal oligodendrocyte differentiation and myelination. *Neuron*. 2010; 65(5):597-611.

Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F *et al*. An integrated encyclopedia of DNA elements in the human genome. *Nature*2012; 489: 57–74.

Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics*. 2009; 10:48.

Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* 1998; 95:14863–8.

Eisenberg E, Levanon EY. Human housekeeping genes are compact. *Trends Genet.* 2003; 19:362–365.

Eisenberg L, Kanner L. Childhood schizophrenia; symposium, 1955. VI. Early infantile autism, 1943-1955. *Am J Orthopsychiatry* 1956; 26:556-66.

Enard W, Khaitovich P, Klose J, Zöllner S, Heissig F, Giavalisco P, *et al*. Intra- and interspecific variation in primate gene expression patterns. *Science*. 2002; 296:340–343.

Enstrom AM, Lit L, Onore CE, Gregg JP, Hansen RL, et al. Altered gene expression and function of peripheral blood natural killer cells in children with autism. *Brain Behav Immun*. 2009; 23:124–133.

Enstrom AM, Onore CE, Van de Water JA, Ashwood P. Differential monocyte responses to TLR ligands in children with autism spectrum disorders. *Brain Behav Immun.* 2010; 24:64–71.

Eroglu C, Barres BA. Regulation of synaptic connectivity by glia. *Nature* 2010, 468:223–231.

Faghihi MA, Modarresi F, Khalil AM, Wood DE, Sahagan BG, Morgan TE, *et al.* Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase. *Nat Med*. 2008; 14(7):723-30.

Fagiolini M, Jensen CL, Champagne FA. Epigenetic influences on brain development and plasticity. *Curr Opin Neurobiol*. 2009; 19(2):207-12.

Fatemi SH, Aldinger KA, Ashwood P, Bauman ML, Blaha CD, Blatt GJ, *et al*. Consensus paper: pathological role of the cerebellum in autism. Cerebellum. *2012*; 11(3):777–807.

Fatemi SH, Blatt GJ. Alterations in GABAergic biomarkers in the autism brain: research findings and clinical implications. *Anat Rec (Hoboken)* 2011, 294(10): 1646-52.

Fatemi SH, Folsom TD, Reutiman TJ, Lee S. Expression of astrocytic markers aquaporin 4 and connexin 43 is altered in brains of subjects with autism. *Synapse.* 2008; 62(7):501-7.

Fischer A, Sananbenesi F, Wang X, Dobbin M, Tsai LH. Recovery of learning and memory is associated with chromatin remodelling. *Nature*. 2007;447:178–182.

Folstein S, Rutter M. Infantile autism: a genetic study of 21 twin pairs. *J. Child Psychol. Psychiatry.* 197*;* 18:297–321.

Fombonne E. Epidemiology of pervasive developmental disorders. *Pediatr Res.* 2009; 65:591–598.
Fountoulakis M, Juranville JF, Dierssen M, Lubec G. Proteomic analysis of the fetal brain. *Proteomics.* 2002; 2(11):1547–76.

Freese JL, Pino D, Pleasure SJ. Wnt signaling in development and disease.*Neurobiol Dis* 2010; 38: 148–153.

Freitag CM. The genetics of autistic disorders and its clinical relevance: a review of the literature. *Mol Psychiatry*. 2007; 12:2–22.

Frings M, Maschke M, Timmann D. Cerebellum and cognition: viewed from philosophy of mind. Cerebellum. *2007*; 6(4):328–34.

Galperin MY. The Molecular Biology Database Collection: 2005 update. *Nucleic Acids Res*. 2005; 33:D5–24.

Garbett K, Ebert PJ, Mitchell A, Lintas C, Manzi B, Mirnics K, Persico AM. Immune transcriptome alterations in the temporal cortex of subjects with autism. *Neurobiol Dis*. 2008; 30(3):303-11.

Gavalas A, Ruhrberg C, Livet J, Henderson CE, Krumlauf R. Neuronal defects in the hindbrain of Hoxa1, Hob2 and Hoxb2 mutatants reflect regulatory interactions among these hox genes. *Development*. 2003; 130(23): 5663-5679.

Geschwind D. Genetics of autism spectrum disorders. *Trends in cognitive sciences*. 2011; 15: 409-416.

Geschwind DH, Levitt P. Autism spectrum disorders: developmental disconnection syndromes. *Curr Opin Neurobiol*. 2007; 17:103–111.

Geschwind DH. Autism: many genes, common pathways? *Cell.* 2008; 135:391–395.

Ghahramani Seno MM, Hu P, Gwadry FG, Pinto D, Marshall CR, Casallo G, Scherer SW. Gene and miRNA expression profiles in autism spectrum disorders. *Brain Res*. 2011; 1380:85-97.

Gillberg C, Cederlund M, Lamberg K, Zeijlon L. Brief report: 'the autism epidemic'. The registered prevalence of autism in a Swedish urban area. *J Autism Dev Disord*. 2006; 4:429–435.

Gilman SR, Iossifov I, Levy D, Ronemus M, Wigler M, Vitkup D. Rare de novo variants associated with autism implicate a large functional network of genes involved in formation and function of synapses. *Neuron*. 2011; 70:898–907.

Ginhoux F, Greter M, Leboeuf M, Nandi S, See P, *et al*. Fate mapping analysis reveals that adult microglia derive from primitive macrophages. *Science*. 2010; 330:841–845.

Ginsberg MR, Rubin RA, Falcone T, Ting AH, Natowicz MR. Brain transcriptional and epigenetic associations with autism. *PLoS One*. 2012; 7(9):e44736.

Giraldez AJ, Cinalli RM, Glasner ME, Enright AJ, Thomson JM, Baskerville S, *et al*. MicroRNAs regulate brain morphogenesis in zebrafish. *Science*. 2005; 308(5723):833-8.

Girirajan S, Rosenfeld JA, Coe BP, Parikh S, Friedman N, *et al*. Phenotypic heterogeneity of genomic disorders and rare copy-number variants. *N. Engl. J. Med*. 2012; 367:1321–31.

Girirajan S, Rosenfeld JA, Cooper GM, Antonacci F, Siswara P, *et al.* A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. *Nat. Genet.* 2010; 42:203–9.

Gkogkas C, Khoutorsky A, Ran I, Rampakakis E, Nevarko T, Weatherill D. *et al.* Autism-related deficits via dysregulated eIF4E-dependent translational control. *Nature*. 2013; 493:371-377.

Glessner JT, Wang K, Cai G, Korvatska O, Kim CE, *et al*. Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature*. 2009; 459:569–73.

Goines P, Van de Water J. The immune system's role in the biology of autism. *Curr Opin Neurol*. 2010; 23:111–117.

Goines PE, Croen LA, Braunschweig D, Yoshida CK, Grether J, Hansen R, et al. Increased midgestational IFN-γ, IL-4 and IL-5 in women bearing a child with autism: A case-control study. *Mol Autism*. 2011; 2:13.

Guilmatre A, Dubourg C, Mosca AL, Legallic S, Goldenberg A, *et al*. Recurrent rearrangements in synaptic and neurodevelopmental genes and shared biologic pathways in schizophrenia, autism, and mental retardation. *Arch Gen Psychiatry*. 2009; 66:947–956.

Hajeri VA, Amatruda JF. Studying synthetic lethal interactions in the zebrafish system: insight into disease genes and mechanisms. *Dis Model Mech*. 2012; 5(1):33-7.

Hallmayer J, Cleveland S, Torres A, Phillips J, Cohen B, Torigoe T, *et al.* Genetic heritability and shared environmental factors among twin pairs with autism. *Archives of general psychiatry*. 2011; 68:1095-1102.

Hansen KF, Karelina K, Sakamoto K, Wayman GA, Impey S, Obrietan K. miRNA-132: a dynamic regulator of cognitive capacity. *Brain Struct Funct*. 2013; 218(3):817-31.

Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, Kjems J. Natural RNA circles function as efficient microRNA sponges. *Nature*. 2013; 495(7441):384-8.

Hansen TB, Wiklund ED, Bramsen JB, Villadsen SB, Statham AL, Clark SJ, Kjems J. miRNA-dependent gene silencing involving Ago2-mediated cleavage of a circular antisense RNA. *EMBO J*. 2011; 30(21):4414-22.

Hart SN, Therneau TM, Zhang Y, Poland GA, Kocher JP. Calculating sample size estimates for RNA sequencing data. *J Comput Biol*. 2013; 20(12):970-8.

Hashimoto Y, Akiyama Y, Yuasa Y. Multiple-to-multiple relationships between microRNAs and target genes in gastric cancer. *PLoS One*. 2013; 8(5):e62589.
He X, Sanders SJ, Liu L, De Rubeis S, Lim ET, Sutcliffe JS, *et al*. Integrated model of de novo and inherited genetic variants yields greater power to identify risk genes. *PLoS Genet*. 2013; 9(8):e1003671.

Herculano-Houzel S. The human brain in numbers: a linearly scaled-up primate brain. *Front Hum Neurosci.* 2009; 3:31.

Herculano-Houzel S. The remarkable, yet not extraordinary, human brain as a scaled-up primate brain and its associated cost. *Proc Natl Acad Sci USA*. 2012; 109 Suppl 1:10661–8.

Hobert O. Gene regulation by transcription factors and microRNAs. *Science*2008; 319: 1785–1786.

Hoerder-Suabedissen A, Oeschger FM, Krishnan ML, Belgard TG, Wang WZ, Lee S, *et al*. Expression profiling of mouse subplate reveals a dynamic gene network and disease association with autism and schizophrenia. *PNAS USA*. 2013; 110(9):3555-60.

Holt R, Barnby G, Maestrini E, Bacchelli E, Brocklebank D, *et al*. Linkage and candidate gene studies of autism spectrum disorders in European populations. *Eur J Hum Genet*. 2010; 18:1013–1019.

Holt R, Monaco AP. Links between genetics and pathophysiology in the autism spectrum disorders. *EMBO Mol Med*. 2011; 8:438–450.

Howarth C, Gleeson P, Attwell D. Updated energy budgets for neural computation in the neocortex and cerebellum. *J Cereb Blood Flow Metab.* 2012; 32(7):1222–32.

Howlin P, Magiati I, Charman T. Systematic review of early intensive behavioral interventions for children with autism. *Am J Intellect Dev Disabil*. 2009; 114(1):23–41.

Hsiao EY, Patterson PH. Placental regulation of maternal-fetal interactions and brain development. *Dev Neurobiol*. 2012; 72(10):1317-26.

Hsiao LL, Dangond F, Yoshida T, Hong R, Jensen RV, *et al*. A compendium of gene expression in normal human tissues. *Physiol Genomics*. 2001; 7:97–104.

Hu HY, Guo S, Xi J, Yan Z, Fu N, Zhang X *et al*. MicroRNA expression and regulation in human, chimpanzee, and macaque brains. *PLoS Genet* 2011; 7: e1002327.

Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009;4: 44–57.

Huguet G, Ey E, Bourgeron T. The genetic landscapes of autism spectrum disorders. *Annu Rev Genomics Hum Genet*. 2013; 14:191-213.

Huttenlocher PR, de Courten C. The development of synapses in striate cortex of man. *Human neurobiology*. 1987; 6(1): 1-9.

Hutton LC1, Castillo-Melendez M, Walker DW. Uteroplacental inflammation results in blood brain barrier breakdown, increased activated caspase 3 and lipid peroxidation in the late gestation ovine fetal cerebellum. *Dev Neurosci*. 2007; 29(4-5):341-54.

Hutvagner G, Zamore PD. A microRNA in a multiple-turnover RNAi enzyme complex. *Science*. 2002; 297:2056–2060.
Imai Y, Ibata I, Ito D, Ohsawa K, Kohsaka S. A novel gene Iba1 in the major histocompatibility complex class III region encoding an EF hand protein expressed in a monocytic lineage. *Biochem Biophys Res Commun* 1996, 224:855–862.

Insel TR. Rethinking schizophrenia. *Nature*. 2010; 468: 187-193.

Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, *et al*. De novo gene disruptions in children on the autistic spectrum. *Neuron.* 2012; 74:285–99.

Ip B, Wappler I, Peters H, Lindsay S, Clowry G, Bayatti N. Investigating gradients of gene expression involved in early human cortical development. *Journal of anatomy*. 2010; 217, 300-311.

Itsara A, Wu H, Smith JD, Nickerson DA, Romieu I, *et al*. De novo rates and selection of large copy number variation. *Genome Res.* 2010; 20:1469–81.

Jacquemont ML, Sanlaville D, Redon R, Raoul O, Cormier-Daire V, *et al*. Array-based comparative genomic hybridisation identifies high frequency of cryptic chromosomal rearrangements in patients with syndromic autism spectrum disorders. *J. Med. Genet.* 2006; 43:843–49.

Jacquemont S, Reymond A, Zufferey F, Harewood L, Walters RG, *et al*. Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus. *Nature.* 2011; 478:97–102.

Janusonis S. Comparing two small samples with an unstable, treatment-independent baseline. *Journal of Neuroscience Methods*. 2009; 179: 173-178.

Jazin E, Cahill L. Sex differences in molecular neuroscience: from fruit flies to humans. *Nat Rev Neurosci* 2010; 11: 9–17.

Johnson MB, Kawasawa YI, Mason CE, Krsnik Z, Coppola G, Bogdanovic D, *et al*. Functional and evolutionary insights into human brain development through global transcriptome analysis. *Neuron*. 2009; 62:494–509.

Jones AR, Overly CC, Sunkin SM. The Allen Brain Atlas: 5 years and beyond. *Nat Rev Neurosci*. 2009; 10:821–828.

Just MA, Cherkassky VL, Keller TA, Kana RK, Minshew NJ. Functional and anatomical cortical underconnectivity in autism: evidence from an FMRI study of an executive function task and corpus callosum morphometry. *Cereb Cortex.* 2007; 17(4):951-61.

Just MA, Cherkassky VL, Keller TA, Minshew NJ. Cortical activation and synchronization during sentence comprehension in high-functioning autism: evidence of underconnectivity. *Brain*. 2004;127(8):1811-21.

Kacimi R, Giffard RG, Yenari MA. Endotoxin-activated microglia injure brain derived endothelial cells via NF-kappaB, JAK-STAT and JNK stress kinase pathways. *J Inflamm (Lond)*. 2011; 8:7.

Kalkman HO. A review of the evidence for the canonical Wnt pathway in autism spectrum disorders. *Mol Autism*. 2012; 3(1):10.

Kandel ER, Schwartz J, Jessell T, Siegelbaum S, Hudspeth AJ. *Principles of Neural Science. 5th ed.* New York: McGraw-Hill Education; 2012. p. 966.

Kang HJ, Kawasawa YI, Cheng F, Zhu Y, Xu X, Li M *et al*. Spatio-temporal transcriptome of the human brain. *Nature.* 2011; 478: 483–489.

Kanner L. Autistic disturbances of affective contact. *Nervous Child*. 1943; 2:217-50.

Kastner DL, Aksentijevich I, Goldbach-Mansky R. Autoinflammatory disease reloaded: a clinical perspective. *Cell.* 2010; 140:784–790.

Kaul D, Habbel P, Derkow K, Krüger C, Franzoni E, Wulczyn FG, Bereswill S, et al. Expression of Toll-like receptors in the developing brain. *PLoS One*. 2012;7(5):e37767.

Kawase-Koga Y, Otaegi G, Sun T: Different timings of Dicer deletion affect neurogenesis and gliogenesis in the developing mouse central nervous system. *Dev Dyn*. 2009; 238:2800-2812.

Kelleher RJ, Bear MF. The autistic neuron: troubled translation? *Cell*. 2008; 135(3):401-6.

Kent L, Gallagher L, Elliott HR, Mowbray C, Chinnery PF. An investigation of mitochondrial haplogroups in autism. *Am J Med Genet B Neuropsychiatr Genet*. 2008; 147B(6):987-9.

Kent L, Lambert C, Pyle A, Elliott H, Wheelwright S, Baron-Cohen S, Chinnery PF. The mitochondrial DNA A3243A>G mutation must be an infrequent cause of Asperger syndrome. *J Pediatr*. 2006; 149(2):280-1.

Kern JK, Geier DA, Sykes LK, Geier MR. Evidence of neurodegeneration in autism spectrum disorder. *Transl Neurodegener* 2013, 2(1):17.

Kettenmann H, Kirchhoff F, Verkhratsky A. Microglia: new roles for the synaptic stripper. *Neuron*. 2013;77(1):10-8.

Khaitovich P, Muetzel B, She X, Lachmann M, Hellmann I, Dietzsch J. Regional patterns of gene expression in human and chimpanzee brains. *Genome Research*. 2004; 14:1462–1473.

Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, *et al*. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci USA* 2009; 28:11667-11672.

Kiecker C, Lumsden A. Hedgehog signaling from ZLI regulates diencephalic regional identity. *Nature Neuroscience*. 2004; 7(11): 1242-1249.

Kim MS, Pinto SM, Getnet D, Nirujogi RS, Manda SS, Chaerkady R, et al. A draft map of the human proteome. Nature. 2014; 509(7502):575-81.

Klei L, Sanders SJ, Murtha MT, Hus V, Lowe JK, *et al*. Common genetic variants, acting additively, are a major source of risk for autism. *Mol. Autism.* 2012; 3:9.

Knott AB, Perkins G, Schwarzenbacher R, Bossy-Wetzel E. Mitochondrial fragmentation in neurodegeneration. *Nat Rev Neurosci.* 2008; 9(7):505-18.

Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, *et al*. Rate of de novo mutations and the importance of father's age to disease risk. *Nature*. 2012; 488:471–75.

Konopka W, Kiryk A, Novak M, Herwerth M, Parkitna JR, Wawrzyniak M, *et al*. MicroRNA loss enhances learning and memory in mice. *J Neurosci*. 2010; 30(44):14835-42.

Krieglstein K, Zheng F, Unsicker K, Alzheimer C. More than being protective: functional roles for TGF-beta/activin signaling pathways at central synapses. *Trends Neurosci* 2011; 34: 421–429.
Kuhn A, Thu D, Waldvogel HJ, Faull RL, Luthi-Carter R. Population specific expression analysis (PSEA) reveals molecular changes in diseased brain. *Nat Methods* 2011, 8(11):945-7.

Kumar V, Abbas AK, Aster JC, Fausto N. *Robbins & Cotran Pathologic Basis of Disease, 8th edition*. Philadelphia: Elsevier Inc, 2010.

Kuss AW, Chen W. MicroRNAs in brain function and disease. *Current Neurology & Neuroscience Reports*. 2008; 8:190–197.

Lambert N, Lambot MA, Bilheu A, Albert V, Englert Y, Libert F, *et al*. Genes expressed in specific areas of the human fetal cerebral cortex display distinct patterns of evolution. *PLoS ONE*. 2011; 6:e17753.

Landers M, Calciano MA, Colosi D, Glatt-Deeley H, Wagstaff J, Lalande M. Maternal disruption of Ube3a leads to increased expression of Ube3a-ATS in trans. *Nucleic Acids Res*. 2005; 33(13):3976-84.

Landry CF, Ivy GO, Brown IR. Developmental expression of glial fibrillary acidic protein mRNA in the rat brain analyzed by in situ hybridization. *J Neurosci Res*. 1990; 25(2):194-203.

Lathia JD, Okun E, Tang SC, Griffioen K, Cheng A, Mughal MR, et al. Toll-like receptor 3 is a negative regulator of embryonic neural progenitor cell proliferation. *J Neurosci*. 2008; 28(51):13978-84.

Laurence JA, Fatemi SH. Glial fibrillary acidic protein is elevated in the superior, frontal, parietal and cerebellar cortices of patients with autism. *Cerebellum.* 2005; 4:206-210.

Lawrence YA, Kemper TL, Bauman ML, Blatt GJ. Parvalbumin-, calbindin-, and calretinin-immunoreactive hippocampal interneuron density in autism. *Acta Neurol Scand* 2010, 121(2):99-108.
Leblond CS, Heinrich J, Delorme R, Proepper C, Betancur C, *et al*. Genetic and functional analyses of *SHANK2* mutations suggest a multiple hit model of autism spectrum disorders. *PLoS Genet.* 2012; 8:e1002521.

Lehnardt S. Innate immunity and neuroinflammation in the CNS: the role of microglia in Toll-like receptor-mediated neuronal injury. *Glia*. 2010;58(3):253-63.

Lenroot RK and Giedd JN. Brain development in children and adolescents: Insights from anatomical magnetic resonance imaging. *Neuroscience and Biobehavioral Reviews*. 2006; 30(6):718-729.

Leone DP, Srinivasan K, Chen B, Alcamo E, McConnell SK. The determination of projection neuron identify in the developing cerebral cortex. *Current Opinion in Neurobiology*. 2008; 18(1): 28-35.

Lepagnol-Bestel AM, Maussion G, Boda B, Cardona A, Iwayama Y, Delezoide AL, *et al*. SLC25A12 expression is associated with neurite outgrowth and is upregulated in the prefrontal cortex of autistic subjects. *Mol Psychiatry*. 2008; 13(4):385-97.

Lépinoux-Chambaud C, Eyer J. Review on intermediate filaments of the nervous system and their pathological alterations. *Histochem Cell Biol.* 2013; 140(1):13-22.

Levitt P, Campbell DB. The genetic and neurobiologic compass points toward common signaling dysfunctions in autism spectrum disorders. *J Clin Invest.* 2009; 119:747–754.

Levitt P. Structural and functional maturation of the developing primate brain. *J Pediatr*. 2003; 143:S35–45.

Levy SE, Mandell DS, Schultz RT. Autism. *Lancet*. 2009; 374(9701):1627–38.

Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets.*Cell* 2005; 120: 15–20.

Li Z, Okamoto K, Hayashi Y, Sheng M. The importance of dendritic mitochondria in the morphogenesis and plasticity of spines and synapses. *Cell*. 2004; 119:873–887.

Liao Y, Lönnerdal B. Global microRNA characterization reveals that miR-103 is involved in IGF-1 stimulated mouse intestinal cell proliferation. *PLoS One*. 2010; 5(9):e12976.

Limperopoulos C, Bassan H, Sullivan NR, Soul JS, Robertson RL Jr, Moore M, *et al*. Positive screening for autism in ex-preterm infants: prevalence and risk factors. *Pediatrics*. 2008; 121(4):758-65.

Lintas C, Sacco R, Persico AM. Genome-wide expression studies in Autism spectrum disorder, Rett syndrome, and Down syndrome. *Neurobiol Dis.* 2010; 45(1):57-68.

Lioy DT, Garg SK, Monaghan CE, Raber J, Foust KD, Kaspar BK, *et al*. A role for glia in the progression of Rett's syndrome. *Nature.* 2011; 475(7357):497-500.

Lipovich L, Dachet F, Cai J, Bagla S, Balan K, Jia H, Loeb JA. Activity-dependent human brain coding/noncoding gene regulatory networks. *Genetics*. 2012; 192(3):1133-48.

Liu C, Zhang F, Li T, Lu M, Wang L, Yue W, Zhang D. MirSNP, a database of polymorphisms altering miRNA target sites, identifies miRNA-related SNPs in GWAS SNPs and eQTLs. *BMC Genomics*. 2012; 13:661.

Liu Q, Paroo Z. Biochemical principles of small RNA pathways. *Annu Rev Biochem*. 2010; 79:295-319.

Lockhart DJ, Barlow C. DNA arrays and gene expression analysis in the brain. In H. R. Chin & S. O. Moldin (Eds.) *Methods in genomic neuroscience*. New York, NY: CRC Press, 2001. pp. 109–140.

Lumsden A, Keynes R. Segmental patterns of neuronal development in the chick hindbrain. *Nature*. 1989; 337(6206): 424-428.

Ma Y, Li J, Chiu I, Wang Y, Sloane JA, Lü J, Kosaras B, et al. Toll-like receptor 8 functions as a negative regulator of neurite outgrowth and inducer of neuronal apoptosis. *J Cell Biol*. 2006; 175(2):209-15.

Mabb AM, Judson MC, Zylka MJ, Philpot BD. Angelman syndrome: insights into genomic imprinting and neurodevelopmental phenotypes. *Trends Neurosci*. 2011; 34:293–303.

Maezawa I, Jin LW. Rett syndrome microglia damage dendrites and synapses by the elevated release of glutamate. *J Neurosci*. 2010; 30(15):5346-56.

Maezawa I, Swanberg S, Harvey D, LaSalle JM, Jin LW. Rett syndrome astrocytes are abnormal and spread MeCP2 deficiency through gap junctions. *J Neurosci*. 2009; 29(16):5051-61.

Maezawa, I., Calafiore, M., Wulff, H. and Jin, L.-W. Does microglial dysfunction play a role in autism and Rett syndrome? *Neuron glia biology*. 2011; 7, 85-97.

Mansfield KD, Keene JD. The ribonome: a dominant force in co-ordinating gene expression. *Biol. Cell*. 2009; 101:169–181.

Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, *et al*. Structural variation of chromosomes in autism spectrum disorder.*Am. J. Hum. Genet.* 2008; 82:477–88.

Martins M, Rosa A, Guedes LC, Fonseca BV, Gotovac K, Violante S, *et al*. Convergence of miRNA expression profiling, α-synuclein interacton and GWAS in Parkinson's disease. *PLoS One*. 2011; 6(10):e25443.

Martins-de-Souza D, Guest PC, Guest FL, Bauder C, Rahmoune H, Pietsch S, *et al*. Characterization of the human primary visual cortex and cerebellum proteomes using shotgun mass spectrometry-data-independent analyses. *Proteomics.* 2012; 12(3):500–4.

Mattick JS, Makunin IV. Non-coding RNA. *Hum Mol Genet*. 2006; 15 Spec No 1:R17-29.

Mattson MP, Liu D. Energetics and oxidative stress in synaptic plasticity and neurodegenerative disorders. *Neuromolecular Med*. 2002; 2:215–231.

Mattson MP. NF-kappaB in the survival and plasticity of neurons. *Neurochem Res*. 2005; 30:883–893.

Matuszek, G, Talebizadeh, Z. Autism Genetic Database (AGD): a comprehensive database including autism susceptibility gene-CNVs integrated with known noncoding RNAs and fragile sites. *BMC medical genetics*. 2009; 10, 102.

Maurer MH. Genomic and proteomic advances in autism research. *Electrophoresis.* 2012;33(24): 3653–8.

Mazin P, Xiong J, Liu X, Yan Z, Zhang X, Li M, *et al*. Widespread splicing changes in human brain development and aging. *Mol Syst Biol*. 2013; 22(9):633.

Mclean CY, Bristor D, Hiller M, *et al*. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol*. 2010; 28:495–501.

Mellios N, Sur M. The Emerging Role of microRNAs in Schizophrenia and Autism Spectrum Disorders. *Front Psychiatry*. 2012 Apr 25;3:39.

Menashe I, Grange P, Larsen EC, Banerjee-Basu S, Mitra PP. Co-expression profiling of autism genes in the mouse brain. *PLoS Comput Biol*. 2013;9(7):e1003128.

Meng L, Person RE, Beaudet AL. Ube3a-ATS is an atypical RNA polymerase II transcript that represses the paternal expression of Ube3a. *Hum Mol Genet*. 2012; 21(13):3001-12.

Mercer TR, Dinger ME, Sunkin SM, Mehler MF, Mattick JS. Specific expression of long noncoding RNAs in the mouse brain. *Proc. Natl. Acad. Sci. USA*. 2008; 105:716–721.

Miles JH. Autism spectrum disorders—a genetics review. *Genet. Med.* 2011; 4:278–294.

Minshew NJ, Keller TA. The nature of brain dysfunction in autism: Functional brain imaging studies. *Curr Opin Neurol*. 2010; 2:124–130.

Mitchell KJ. The genetics of neurodevelopmental disease. *Curr Opin Neurobiol*. 2011; 21:197–203.

Modarresi F, Faghihi MA, Patel NS, Sahagan BG, Wahlestedt C, Lopez-Toledano MA. Knockdown of BACE1-AS Nonprotein-Coding Transcript Modulates Beta-Amyloid-Related Hippocampal Neurogenesis. *Int J Alzheimers Dis*. 2011; 2011:929042.

Moldin, SO, Rubenstein, JLR. *Understanding autism: From basic neuroscience to treatment*. Boca Raton, FL: CRC/Taylor & Frances, 2006. pp 475-502.

Molloy CA, Morrow AL, Meinzen-Derr J, Schleifer K, Dienger K, *et al*. Elevated cytokine levels in children with autism spectrum disorder. *J Neuroimmunol*. 2006; 172:198–205.

Molnár Z, Métin C, Stoykova A, Tarabykin V, Price DJ, Francis F, *et al*. Comparative aspects of cerebral cortical development. *Eur J Neurosci*. 2006; ;23(4):921-34.

Molyneaux BJ, Arlotta P, Menezes JR, Macklis JD. Neuronal subtype specification in the cerebral cortex. *Nature Reviews Neuroscience*. 2007; 8:427–437.

Moncini S, Salvi A, Zuccotti P, *et al*. The role of miR-103 and miR-107 in regulation of CDK5R1 expression and in cellular migration. *PLoS One.* 2011; 6:e20038.

Morgan JT, Chana G, Abramson I, Semendeferi K, Courchesne E, Everall IP. Abnormal microglial-neuronal spatial organization in the dorsolateral prefrontal cortex in autism. *Brain Res*. 2012; 1456:72-81.

Morgan JT, Chana G, Pardo CA, Achim C, Semendeferi K, *et al.* Microglial activation and increased microglial density observed in the dorsolateral prefrontal cortex in autism. *Biol Psychiatry*. 2010; 68:368–376.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods.* 2008; 5:621–628.

Nadler JJ, Zou F, Huang H, Moy SS, Lauder J, Crawley JN, *et al.* Large-scale gene expression differences across brain regions and inbred strains correlate with a behavioral phenotype. *Genetics*. 2006; 174:1229–1336.

Nakamura H, Katahira T, Matsunaga E, Sato T. Isthmus organizer for midbrain and hindbrain development. *Brain Research Reviews*. 2005; 49(2): 120-126.

Natera-Naranjo O, Aschrafi A, Gioio AE, Kaplan BB. Identification and quantitative analyses of microRNAs located in the distal axons of sympathetic neurons. *RNA*. 2010; 16(8):1516-29.

Naumova OY, Palejev D, Vlasova NV, Lee M, Rychkov SY, Babich ON, *et al*. Age-related changes of gene expression in the neocortex: Preliminary data on RNA-seq of the transcriptome in three functionally distinct cortical areas. *Development and Psychopathology*. 2008; 24:1427–1442.

Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A, *et al*. Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature*. 2012; 4:242–245.

Nelson PT, Wang WX. MiR-107 is reduced in Alzheimer's disease brain neocortex: Validation study. *J. Alzheimer's Dis*. 2010; 21:75–79.

Neves-Pereira M, Müller, B., Massie, D., Williams, J., O'Brien, P., Hughes, A., *et a*l. Deregulation of EIF4E: a novel mechanism for autism. *Journal of medical genetics*, 2009; 46, 759-765.

Nunnari J, Suomalainen A. Mitochondria: in sickness and in health. *Cell*. 2012; 148(6):1145-59.

Oblak AL, Rosene DL, Kemper TL, Bauman ML, Blatt GJ. Altered posterior cingulate cortical cyctoarchitecture, but normal density of neurons and interneurons in the posterior cingulate cortex and fusiform gyrus in autism. *Autism Res* 2011, 4(3):200-11.

Okun E, Griffioen K, Barak B, Roberts NJ, Castro K, Pita MA, et al. Toll-like receptor 3 inhibits memory retention and constrains adult hippocampal neurogenesis. *PNAS USA*. 2010; 107(35):15625-30.

Oldham M, Konopka G, Iwamoto K, Langfelder P, Kato T, Horvath S, Geschwind D. Functional organization of the transcriptome in human brain. *Nature neuroscience*. 2008; 11:1271-1282.

Oliveira G, Diogo L, Grazina M, Garcia P, Ataíde A, Marques C, *et al*. Mitochondrial dysfunction in autism spectrum disorders: a population-based study. *Dev Med Child Neurol*. 2005; 47(3):185-9.

O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, *et al*. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nat. Genet.* 2011; 43:585–89.

O'Roak BJ, State MW. Autism genetics: strategies, challenges, and opportunities. *Autism Res*. 2008; 1(1):4-17.

O'Roak BJ, Vives L, Fu W, Egertson JD, Stanaway IB, *et al*. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science*. 2012; 338:1619–22.

O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, *et al*. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature*. 2012; 485(7397):246-50.

Orom UA, Derrien T, Beringer M, Gumireddy K, Gardini A, Bussotti G, *et al*. Long noncoding RNAs with enhancer-like function in human cells. *Cell.* 2010; 1:46-58.

Ozair MZ, Kintner C, Brivanlou AH. Neural induction and early patterning in vertebrates. *Wiley Interdiscip Rev Dev Biol*. 2013 Jul;2(4):479-98.

Ozonoff S, Young GS, Carter A, Messinger D, Yirmiya N, et al. Recurrence risk for autism spectrum disorders: a Baby Siblings Research Consortium study. *Pediatrics*. 2011; 128:e488–95.

Paolicelli RC, Bolasco G, Pagani F, Maggi L, Scianni M, Panzanelli P, Giustetto M, Ferreira TA, Guiducci E, Dumas L, Ragozzino D, Gross CT. Synaptic pruning by microglia is necessary for normal brain development. *Science* 2011, 6048:1456-8.

Pardo CA, Vargas DL, Zimmerman AW. Immunity, neuroglia and neuroinflammation in autism. *Int Rev Psychiatry*. 2005; 17:485–495.

Pastural E, Ritchie S, Lu Y, Jin W, Kavianpour A, Khine Su-Myat K, *et al*. Novel plasma phospholipid biomarkers of autism: mitochondrial dysfunction as a putative causative mechanism. *Prostaglandins Leukot Essent Fatty Acids.* 2009; 81(4):253-64.

Paul T, Keshavan M, Giedd JN. Why do many psychiatric disorders emerge during adolescence? *Nature Rev Neurosci*. 2008; 9:947-957.

Paus T, Collins DL, Evans AC, Leonard G, Pike B, Zijdenbos A. Maturation of white matter in the human brain: a review of magnetic resonance studies. *Brain Research Bulletin.* 2001; 54(3): 255-266.

Paz-Yaacov N, *et al*. Adenosine-to-inosine RNA editing shapes transcriptome diversity in primates. *Proc. Natl Acad. Sci. USA*. 2010; 107:12174–12179.

Pérez-Bercoff Å, Hudson CM, Conant GC. A conserved mammalian protein interaction network. *PLoS One*. 2013; 8(1):e52581.

Peters J, Williamson CM. Control of imprinting at the Gnas cluster. *Adv Exp Med Biol*. 2008; 626:16–26.

Pham NV, Nguyen MT, Hu JF, Vu TH, Hoffman AR. Dissociation of IGF2 and H19 imprinting in human brain. *Brain Res.* 1998; 810(1-2):1-8.

Phillips PC. Epistasis--the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet*. 2008; 9(11):855-67.

Pickett J, London E. The neuropathology of autism: a review. *J Neuropathol Exp Neurol*. 2005; 64:925–935.

Pickett J, London E. The neuropathology of autism: a review. *J Neuropathol Exp Neurol.* 2005; 64:925–935.

Pieretti M, Zhang FP, Fu YH, Warren ST, Oostra BA, *et al*. Absence of expression of the FMR-1 gene in fragile X syndrome. *Cell*. 1991; 66:817–22.

Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, *et al*. Functional impact of global rare copy number variation in autism spectrum disorders. *Nature.* 2010; 466:368–72.
Piton A, Jouan L, Rochefort D, Dobrzeniecka S, Lachapelle K, Dion P, *et al*. Analysis of the effects of rare variants on splicing identifies alterations in GABA(A) receptor genes in autism spectrum disorder individuals. *Eur J Hum Genet.* 2013 Jul;21(7):749-56.

Piven J, Palmer P. Cognitive deficits in parents from multiple-incidence autism families. *J Child Psychol Psychiatry*. 1997; 38:1011–1021.

Poling JS, Frye RE, Shoffner J, Zimmerman AW. Developmental regression and mitochondrial dysfunction in a child with autism. *J Child Neurol*. 2006; 21(2):170-2.

Pollard KS, *et al*. An RNA gene expressed during cortical development evolved rapidly in humans. *Nature*. 2006; 443:167–172.

Ponjavic J, Oliver PL, Lunter G, Ponting CP. Genomic and transcriptional co-localization of protein-coding and long non-coding RNA pairs in the developing brain. *PLoS Genet*. 2009; 5(8):e1000617.

Ponting CP, Belgard TG. Transcribed dark matter: meaning or myth? *Hum Mol Genet*. 2010; 19(R2):R162-8.

Ponting CP, Oliver PL, Reik W. Evolution and functions of long noncoding RNAs. *Cell*. 2009; 136(4):629-41.

Purcell AE, Jeon OH, Zimmerman AW, Blue ME, Pevsner J. Postmortem brain abnormalities of the glutamate neurotransmitter system in autism. *Neurology*. 2001; 57(9):1618-28.

Qureshi IA, Mehler MF. Emerging roles of non-coding RNAs in brain evolution, development, plasticity and disease. *Nat Rev Neurosci*. 2012; 13(8):528-41.

Qureshi IA, Mehler MF. Non-coding RNA networks underlying cognitive disorders across the lifespan. *Trends Mol Med*. 2011; 17(6):337-46.

Ramoz N, Reichert JG, Smith CJ, Silverman JM, Bespalova IN, Davis KL, Buxbaum JD. Linkage and association of the mitochondrial aspartate/glutamate carrier SLC25A12 gene with autism. *Am J Psychiatry*. 2004; 161(4):662-9.

Ramskold D, Wang ET, Burge CB, Sandberg, R. An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. *PLoS Computational Biology*. 2009; 5:e1000598.

Ransohoff R, Perry V. Microglial Physiology: Unique Stimuli, Stimuli, Specialized Responses. *Annu Rev Immunol* 2009, 27:119-145.

Rao PA, Beidel DC. The impact of children with high-functioning autism on parental stress, sibling adjustment, and family functioning. *Behav Modif*. 2009; 33(4):437-51.

Redies C, Hertel N, Hübner CA. Cadherins and neuropsychiatric disorders. *Brain Res*. 2012; 1470:130-44.

Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, *et al*. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell.* 2007; 7:1311-1323.

Ritvo E, Jorde L, Mason-Brothers A., Freeman B, Pingree C, Jones M., *et al*. The UCLA-University of Utah epidemiologic survey of autism: recurrence risk estimates and genetic counseling. *The American journal of psychiatry*. 1989; 146:1032-1036.
Robin ED, Wong R. Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. *J Cell Physiol*. 1988; 136(3):507-13.

Robinson EB, Lichtenstein P, Anckarsäter H, Happé F, Ronald A. Examining and interpreting the female protective effect against autistic behavior. *Proc Natl Acad Sci U S A*. 2013; 110(13):5258-62.

Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.*Bioinformatics* 2010; 26: 139–140.

Robles JA, Qureshi SE, Stephen SJ, Wilson SR, Burden CJ, Taylor JM. Efficient experimental design and analysis strategies for the detection of differential expression using RNA-Sequencing. *BMC Genomics* 2012; 13: 484.

Ronald A, Happe F, Bolton P, Butcher LM, Price TS, *et al*. Genetic heterogeneity between the three components of the autism spectrum: a twin study. *J Am Acad Child Adolesc Psychiatry*. 2006; 45:691–699.

Ronemus M, Iossifov I, Levy D, Wigler M. The role of de novo mutations in the genetics of autism spectrum disorders. *Nat Rev Genet*. 2014; 15(2):133-41.

Rossignol DA, Frye RE. Mitochondrial dysfunction in autism spectrum disorders: a systematic review and meta-analysis. *Mol Psychiatry*. 2012; 17(3): 290–314.

Roth RB, Hevezi P, Lee J, Willhite D, Lechner SM, Foster AC, *et al*. Gene expression analyses reveal molecular relationships among 20 regions of the human CNS. *Neurogenetics*. 2006; 7:67–80.

Rubenstein JL, Merzenich MM. Model of autism: increased ratio of excitation/inhibition in key neural systems. *Genes Brain Behav.* 2003; 2(5):255-67

Sakai Y, Shaw CA, Dawson BC, Dugas DV, Al-Mohtaseb Z, *et al*. Protein interactome reveals converging molecular pathways among autism disorders. *Sci Transl Med*. 2011; 3:86ra49.

Sala C, Piëch V, Wilson NR, Passafaro M, Liu G, Sheng M. Regulation of dendritic spine morphology and synaptic function by Shank and Homer. *Neuron*. 2001; 31(1):115-30.

Samaco RC, Hogart A, LaSalle JM. Epigenetic overlap in autism-spectrum neurodevelopmental disorders: MECP2 deficiency causes reduced expression of UBE3A and GABRB3. *Hum Mol Genet*. 2005; 14:483–492.

Sanchez-Vives MV, McCormick DA. Cellular and network mechanisms of rhythmic recurrent activity in neocortex. *Nat Neurosci*. 2000; 3(10):1027-34.

Sancho-Tello M, Vallés S, Montoliu C, Renau-Piqueras J, Guerri C. Developmental pattern of GFAP and vimentin gene expression in rat brain and in radial glial cultures. *Glia*. 1995; 15(2):157-66.

Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, *et al*. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron*. 2011; 70:863–85.

Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, *et al*. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature*. 2012; 485:237–41.

Santarelli DM, Beveridge NJ, Tooney PA, Cairns MJ. Upregulation of dicer and microRNA expression in the dorsolateral prefrontal cortex Brodmann area 46 in schizophrenia. *Biol Psychiatry*. 2011; 69:180–187

Santini E, Huynh T, MacAskill A, Carter A, Pierre P, Ruggero D, *et al*. Exaggerated translation causes synaptic and behavioural aberrations associated with autism. *Nature*. 2013; 493, 411-415.

Sarachana T, Zhou R, Chen G, Manji HK, Hu VW. Investigation of post-transcriptional gene regulatory networks associated with autism spectrum disorders by microRNA expression profiling of lymphoblastoid cell lines. *Genome Med*. 2010; 2(4):23.

Saurat N, Andersson T, Vasistha NA, Molnár Z, Livesey FJ. Dicer is required for neural stem cell multipotency and lineage progression during cerebral cortex development. *Neural Dev*. 2013; 29(8):14.

Schaefer GB, Mendelsohn NJ. Clinical genetics evaluation in identifying the etiology of autism spectrum disorders. *Genet Med.* 2008; 10:301–305.

Schofield CM, Hsu R, Barker AJ, Gertz CC, Blelloch R, Ullian EM. Monoallelic deletion of the

microRNA biogenesis gene Dgcr8 produces deficits in the development of excitatory synaptic transmission in the prefrontal cortex. *Neural Dev*. 2011; 6:11.

Schonrock N, Ke YD, Humphreys D, Staufenbiel M, Ittner LM, Preiss T, *et al*. Neuronal microRNA deregulation in response to Alzheimer's disease amyloid-beta. *PLoS ONE*. 2010; 5:e11070.

Schumann CM, Nordahl CW. Bridging the gap between MRI and postmortem research in autism. *Brain Res*. 2011; 1380:175–186.

Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, *et al*. Strong association of de novo copy number mutations with autism. *Science*. 2007; 316:445–49.

Selkirk CG, McCarthy Veach P, Lian F, Schimmenti L, LeRoy BS. Parents' perceptions of autism spectrum disorder etiology and recurrence risk and effects of their perceptions on family planning: Recommendations for genetic counselors. *J Genet Couns*. 2009; 18(5):507-19.

Seno MM, Hu P, Gwadry FG, *et al*. Gene and miRNA expression profiles in autism spectrum disorders. *Brain Res.* 2011; 1380:85–97.

Serajee FJ, Zhang H, Huq A. Prevalence of common mitochondrial point mutations in autism. *Neuropediatrics*. 2006; 37(1): S127.

Shao NY, Hu HY, Yan Z, Xu Y, Hu H, Menzel C *et al*. Comprehensive survey of human brain microRNA by deep sequencing. *BMC Genomics* 2010; 11: 409.

Sharan R, Suthram S, Kelley RM, Kuhn T, McCuine S, Uetz P, *et al.* Conserved patterns of protein interaction in multiple species. *Proc Natl Acad Sci USA*. 2005; 102(6):1974-9.

Sheikh A, Li X, Wen G, Tauqeer Z, Brown W, Malik M. Cathepsin D and apoptosis related proteins are elevated in the brain of autistic subjects. *Neuroscience*. 2010; 165, 363-370.

Sheng ZH, Cai Q. Mitochondrial transport in neurons: impact on synaptic homeostasis and neurodegeneration. *Nature Rev Neuroscience*. 2013; 13, 77-93.

Shibata M, Nakao H, Kiyonari H, Abe T, Aizawa S. MicroRNA-9 regulates neurogenesis in mouse telencephalon by targeting multiple transcription factors. *J Neurosci*. 2011; 31(9):3407-22.

Silber J, Lim DA, Petritsch C, Persson AI, Maunakea AK, Yu M, *et al*. miR-124 and miR-137 inhibit proliferation of glioblastoma multiforme cells and induce differentiation of brain tumor stem cells. *BMC Med*. 2008; 6:14.

Skog J, *et al*. Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers. *Nature Cell Biol*. 2008; 10:1470–1476.

Skuse DH. Imprinting, the X-chromosome, and the male brain: explaining sex differences in the liability to autism. *Pediatr Res*. 2000; 47(1):9-16.

Smalley S, Asarnow R, Spence M. Autism and genetics. A decade of research. *Archives of general psychiatry*. 1998; 45:953-961.

Smibert P, Bejarano F, Wang D, Garaulet DL, Yang JS, Martin R, *et al*. A Drosophila genetic screen yields allelic series of core microRNA biogenesis factors and reveals post-developmental roles for microRNAs. *RNA*. 2011; 17(11):1997-2010.

Smith AM, Gibbons HM, Oldfield RL, Bergin, PM, Mee EW, Curtis MA, *et al*. M-CSF increases proliferation and phagocytosis while modulating receptor and transcription factor expression in adult microglia. *J Neuroinflammation*, 2013, 10:85

Smith M, Flodman PL, Gargus JJ, Simon MT, Verrell K., Haas R, *et al.* Mitochondrial and ion channel gene alterations in autism. *Biochimica et Biophysica Acta (BBA) – Bioenergetics*. 2012; 1817, 1796-1802.

Smith R, Sadee W. Synaptic signaling and aberrant RNA splicing in autism spectrum disorders. *Frontiers in synaptic neuroscience*. 2011; 3, 1.

Smoot M, Ono K., Ruscheinski J, Wang PL, Ideker T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*. 2011; 27, 431-432.

Somel M, Guo S, Fu N, Yan Z, Hu HY, Xu Y *et al*. MicroRNA, mRNA, and protein expression link development and aging in human and macaque brain.*Genome Res* 2010; 20: 1207–1218.

Somel M, Liu X, Tang L, Yan Z, Hu H, Guo S *et al*. MicroRNA-driven developmental remodeling in the brain distinguishes humans from other primates. *PLoS Biol* 2011; 9: e1001214.

Somel M, Franz H, Yan Z, Lorenc A, Guo S, Giger T, *et al*. Transcriptional neoteny in the human brain. *PNAS*. 2009; 106:5743–5757.

Song HJ1, Stevens CF, Gage FH. Neural stem cells from adult hippocampus develop essential properties of functional CNS neurons. *Nat Neurosci.* 2002; 5(5):438-45.

Sowell, ER, Thompson PM, Toga AW. Mapping changes in the human cortex throughout the span of life. *Neuroscientist*. 2004; 10:372–392.

Spooren W, Lindemann L, Ghosh A, Santarelli L. Synapse dysfunction in autism: a molecular medicine approach to drug discovery in neurodevelopmental disorders. *Trends in pharmacological sciences*. 2012; 33, 669-684.

St Laurent G, Faghihi MA, Wahlestedt C. Non-coding RNA transcripts: Sensors of neuronal stress, modulators of synaptic plasticity, and agents of change in the onset of Alzheimer's disease. *Neuroscience Letters*. 2009; 466:81–88.
Steffenburg S, Gillberg C, Hellgren L, Andersson L, Gillberg I, Jakobsson G, Bohman M. A twin study of autism in Denmark, Finland, Iceland, Norway and Sweden. *Journal of child psychology and psychiatry, and allied disciplines*. 1989; 30:405-416.

Stiles J, Jernigan TL. The basics of brain development. *Neuropsychol Rev*. 2010; 20:327-348.

Strand AD, Aragaki AK, Baquet ZC, Hodges A. Cunningham P, Holmans P, *et al*. Conservation of regional gene expression in mouse and human brain. *PLoS Genetics*. 2007; 3:e59.

Stuart J, Segal E, Koller D, Kim S. A gene-coexpression network for global discovery of conserved genetic modules. *Science.* 2003; 302:249-255.

Student. The probable error of a mean. *Biometrika*. 1908; 6: 1-25.

Sun T, Patoine C, Abu-Khalil A, Visvader J, Sum E, Cherry T, *et al.* Early asymmetry of gene transcription in embryonic human left and right cerebral cortex. *Science. 2005;* 308, 1794-1798.

Suzuki K, Sugihara G, Ouchi Y, Nakamura K, Futatsubashi M, Takebayashi K, *et al*. Microglial activation in young adults with autism spectrum disorder. *JAMA Psychiatry*. 2013; 70(1):49-58.

Szafranski P, Schaaf CP, Person RE, Gibson IB, Xia Z, *et al*. Structures and molecular mechanisms for common 15q13.3 microduplications involving CHRNA7: benign or pathological? *Hum. Mutat*. 2010; 31:840–50.

Szatmari P, Paterson AD, Zwaigenbaum L, Roberts W, Brian J, *et al*. Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nat. Genet.* 2007; 39:319–28.

Talebizadeh Z, Butler MG, Theodoro MF. Feasibility and relevance of examining lymphoblastoid cell lines to study role of microRNAs in autism. *Autism Res*. 2008; 1(4):240-50

Tao J, Wu H, Lin Q, Wei W, Lu XH, Cantle JP, *et al*. Deletion of astroglial Dicer causes non-cell-autonomous neuronal dysfunction and degeneration. *J Neurosci*. 2011; 31(22):8306-19.

Taurines R, Thome J, Duvigneau JC, Forbes-Robertson S, Yang L, Klampfl K, *et al*. Expression analyses of the mitochondrial complex I 75-kDa subunit in early onset schizophrenia and autism spectrum disorder: increased levels as a potential biomarker for early onset schizophrenia. *Eur Child Adolesc Psychiatry*. 2010; 19(5):441-8.

Tebbenkamp AT, Borchelt DR. Analysis of chaperone mRNA expression in the adult mouse brain by meta analysis of the Allen Brain Atlas. *PLoS One.* 2010; 5:e13675.

Tetreault NA, Hakeem AY, Jiang S, Williams BA, Allman E, Wold BJ, Allman JM. Microglia in the cerebral cortex in autism. *J Autism Dev Disord.* 2012; 42(12):2569-84.

Tornow S, Mewes HW. Functional modules by relating protein interaction networks and gene expression. *Nucleic Acids Res*. 2003; 31(21):6283-9.

Uhlmann EJ, Apicelli AJ, Baldwin RL, Burke SP, Bajenaru ML, Onda H, *et al*. Heterozygosity for the tuberous sclerosis complex (TSC) gene products results in increased astrocyte numbers and decreased p27-Kip1 expression in TSC2+/- cells. *Oncogene.* 2002; 21(25):4050-9.

Vargas DL, Nascimbene C, Krishnan C, Zimmerman AW, Pardo CA. Neuroglial activation and neuroinflammation in the brain of patients with autism. *Ann Neurol.* 2005; 57(1):67-81.

Velmeshev D, Magistri M, Faghihi MA. Expression of non-protein-coding antisense RNAs in genomic regions related to autism spectrum disorders. *Mol Autism*. 2013; 4(1):32.

Visser L, Melief MJ, van Riel D, van Meurs M, Sick EA, Inamura S, et al. Phagocytes containing a disease-promoting Toll-like receptor/Nod ligand are present in the brain during demyelinating disease in primates. *Am J Pathol*. 2006; 169(5):1671-85.

Vogel AC, Power JD, Petersen SE, Schlaggar BL. Development of the brain's functional network architecture. *Neuropsychol Rev*. 2010; 20:362–375.

Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, et al. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature.* 2011; 474:380–384.

Voineagu I. Gene expression studies in autism: moving from the genome to the transcriptome and beyond. *Neurobiol Dis*. 2012; 45(1):69-75.

Volpe JJ. Overview: normal and abnormal human brain development. *Ment Retard Dev Disabil Res Rev*. 2000. 6;1:1-5.

Vorstman JA, Staal WG, van Daalen E, van Engeland H, Hochstenbach PF, Franke L. Identification of novel autism candidate regions through analysis of reported cytogenetic abnormalities associated with autism. *Mol. Psychiatry*. 2006; 11:18–28.

Wall DP, Pivovarov R, Tong M, Jung JY, Fusaro VA, DeLuca TF *et al*. Genotator: a disease-agnostic tool for genetic annotation of disease. *BMC Med Genomics* 2010; 3: 50.

Wang ET, Sandberg R., Luo S, Khrebtukova I, Zhang L, Mayr C, *et al*. Alternative isoform regulation in human tissue transcriptomes. *Nature*. 2008; 456:470–476.

Wang K, Zhang H, Ma D, Bucan M, Glessner JT, Abrahams BS, et al. Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature*. 2009; 459(7246):528-33.

Wang K, Zhang H, Ma D, Bucan M, Glessner JT, *et al*. Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature*. 2009; 459:528–33.

Wang X. miRDB: a microRNA target prediction and functional annotation database with a wiki interface. *RNA* 2008; 14: 1012–1017.

Weiss LA, Arking DE, Daly MJ, Chakravarti A. A genome-wide linkage and association scan reveals novel loci for autism. *Nature*. 2009; 461:802–8.

Werling DM, Geschwind DH. Sex differences in autism spectrum disorders.*Curr Opin Neurol* 2013; 26: 146–153.

Whitney ER, Kemper TL, Rosene DL, Bauman ML, Blatt GJ. Density of cerebellar basket and stellate cells in autism: evidence for a late developmental loss of Purkinje cells. *J Neurosci Res* 2009, 87(10):2245-54.

Winden K, Oldham M, Mirnics K, Ebert P, Swan C, Levitt P, *et al.* The organization of the transcriptional network in specific neuronal classes. *Molecular systems biology*. 2009; 5:291.

Wolfe C, Kohane I, Butte A. Systematic survey reveals general applicability of "guilt-by-association" within gene coexpression networks. *BMC bioinformatics*. 2005; 6:227.

Wu J, Xie X. Comparative sequence analysis reveals an intricate network among REST, CREB and miRNA in mediating neuronal gene expression. *Genome Biol*. 2006; 7(9):R85.

Xu LM, Li JR, Huang Y, Zhao M, Tang X, Wei L. AutismKB: an evidence-based knowledgebase of autism genetics. *Nucleic acids research*. 2012; 40, 22.

Yang C, Iyer RR, Yu AC, Yong RL, Park DM, Weil RJ, et al. β-Catenin signaling initiates the activation of astrocytes and its dysregulation contributes to the pathogenesis of astrocytomas. *PNAS USA*. 2012; 109(18):6963-8.

Yasui DH, Xu H, Dunaway KW, Lasalle JM, Jin LW, Maezawa I. MeCP2 modulates gene expression pathways in astrocytes. *Mol Autism*. 2013; 4(1):3.

Yeo G, Holste D, Kreiman G, Burge CB. Variation in alternative splicing across human tissues. *Genome Biology*. 2004; 5:R74.

Yoo AS, Sun AX, Li L, Shcheglovitov A, Portmann T, Li Y, *et al*. MicroRNA-mediated conversion of human fibroblasts to neurons. *Nature*. 2011; 476(7359):228-31.

Yuskaitis CJ, Beurel E, Jope RS. Evidence of reactive astrocytes but not peripheral immune system activation in a mouse model of Fragile X syndrome. *Biochim Biophys Acta.* 2010; 1802(11):1006-12.

Zabell SL. On Student's 1908 article "The Probable Error of a Mean." *Journal of the American Statistical Association*. 2008; 103: 1-7.

Zalc B, Goujet D, Colman D. The origin of the myelination program in vertebrates. **Curr Biol**. 2008; 18(12):R511-2.

Zhao C, Sun G, Li S, Shi Y. A feedback regulatory loop involving microRNA-9 and nuclear receptor TLX in neural stem cell fate determination. *Nat Struct Mol Biol*. 2009; 16(4):365-71.

Zhao X, He X, Han X, Yu Y, Ye F, Chen Y, *et al*. MicroRNA-mediated control of oligodendrocyte differentiation. *Neuron*. 2010; 65(5):612-26.

Zoghbi HY. Postnatal neurodevelopmental disorders: meeting at the synapse? *Science*. 2003; 302:826–830.

Supplementary Information

Most supplementary information referred to in this thesis consists of very large data tables containing raw gene expression or gene ontology results. In no instances are the supplementary data referred to necessary to interpret the results presented herein. Therefore, in an attempt to conserve space, the Supplementary Tables described in the text and listed below are included on the accompanying CD attached to this work. Additionally, the same dataset that is found on the accompanying CD can be accessed for free at:

http:// mziats.wix.com/cambridgethesis

Chapter 2.1

Supplementary Table S1. List of AutDB Genes.

Supplementary Document S2. Allen Brain Institute BrainSpan Detailed Methods.

Supplementary Table S3. SZGene Gene List.

Supplementary Table S4. CarpeDB Gene List.

Supplementary Table S5. AutDB Expression Heatmap.

Supplementary Table S6. SZGene Expression Heatmap.

Supplementary Table S7. CarpeDB Expression Heatmap.

Supplementary Table S8. Validation of approach via known constantly expressed genes.

Supplementary Table S9. Validation of approach via housekeeping genes.

Supplementary Table S10. Validation of approach via Intermediate Filaments.

Supplementary Table S11. Expression heatmap of CNR1.

Supplementary Table S12. Highly Expressed ASD Genes by Region.

Supplementary Figure S13. Reelin pathway depicting ASD-associated genes in yellow.

Chapter 2.2

Supplementary Table S14. GO enrichment of ASD modules

Chapter 2.3

Supplementary Table S15. Known sex biased genes analyzed.

Supplementary Tables S16 – 19. Confirmatory GO analysis using IPA and GeneGO.

Supplementary Table 20. Male sex biased pathways are enriched among autism GO terms.

Supplementary Table S21. 'Expression' versus 'Inherited' gene sets.

Supplementary Table S22. 'Expression' set GO enrichment results.

Supplementary Table S23. 'Inherited' set GO enrichment results.

Chapter 2.4

Supplementary Table S24. Full miRNA dataset analyzed after removing lowly-expressed miRNAs.

Supplementary Table S25 – S27. DE miRNAs within brain regions (S25), between brain regions (S26), and between sexes in the prefrontal cortex (S27).

Supplementary Tables S28-33. Gene targets of differentially expressed miRNAs by region (S28) and gene ontology terms of gene targets by region (S29-S33).

Chapter 3.1

Supplementary Table S34. PCR primers used for RT-PCR confirmation.

--