Title
The Function and Regulation of AUTS2

Author
Oksenberg, Nir

Publication Date
2014

Peer reviewed|Thesis/dissertation

The Function and Regulation of AUTS2

by

Nir Oksenberg

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biomedical Sciences

in the

[redacted]

Acknowledgements

The work presented below and my success and happiness as a graduate student would not have been possible without the help of many people. First and foremost, Nadav Ahituv, my graduate advisor has been a fantastic mentor and role model. He will always have my gratitude for his dedication to helping me succeed as a graduate student and beyond. His guidance has not let me down and his enthusiasm and intellect has made me become a better scientist. It has been a lot of fun, and a great pleasure to work with Nadav these last 6 years and I am confident our relationship will remain strong in the future.

I owe a huge part of my success and happiness to the members of the Ahituv lab, who I saw on a daily basis and made this lab a great place to work. Their fantastic personalities and selfless attitudes truly made my time here enjoyable and efficient. These people include Sierra Nishizaki, Betty Booker, Julia Vandermeer, Walter Eckalbar, Meej party fun Kim, Hane Ryu, Robin Smith, Karl Murphy, Ramon Birnbaum, Fumitaka Inoue, Marcelo Luizon, Navneet Matharu, Stacy Jones, Jingjing Zhao, Saziya Latif, Loan Nguyen, Natalie Feliciano, Peter Lu, Mandy Gomez, Natalie Geeter, Angel Ku, and others who have passed by our lab. I would also like to thank my amazing collaborators. Genevieve Haliburton from Katherine Pollard's lab and Laurie Stevison from Jeff Wall's lab have been instrumental to my publications. Gen and Laurie have been a lot of fun to work with and have shown me how great collaborations can work.

My committee, Nadav Ahituv, John Rubenstein, Laurie Weiss, and Steve Hamilton have been incredibly helpful, and a joy to work with. I thank them all very much for their commitment to my success. Previous mentors including Michelle Hermiston and Art Weiss were critical in shaping me into a successful graduate student and I thank them all. I would also like to thank all

the members of the BMS office here at UCSF, in particular Lisa Magargal, Monique Piazza, Nathan Jew and Caroline Rutland. Without their help, I would be lost.

My friends and family, both in and out of the sciences, have made the last 6 years here in San Francisco amazing. I can't begin to list them all and describe their contribution to my success and happiness but I would like to specifically thank my parents Jorge and Donna Oksenberg, my sister Yael Markley, her husband John and my nephew Oliver for their constant support and countless dinners and fantastic bottles of wine. I can't thank Brittany Anderton enough for her endless love and support. Yien Kuo and Ramon Birnbaum have also helped me immensely throughout my graduate career and I look forward to keeping in touch with them. All of my friends at UCSF, including members of the Policy Club, and throughout California have made this university and city a fantastic place to be. You know who you are. Thank you. Finally, thank you Moe for being a good dog.

Contributions of co-authors to the presented work

Chapter 1: Work from this chapter is adapted from the following published peer reviewed review article: [Oksenberg, N., & Ahituv, N. (2013). The role of AUTS2 in neurodevelopment and human evolution. *Trends in Genetics : TIG*, 1–9. doi:10.1016/j.tig.2013.08.001]. NO wrote this article with the help of NA.

Chapter 2: Work from this chapter is adapted from the following published peer reviewed article: [Oksenberg, N., Stevison, L., Wall, J. D., & Ahituv, N. (2013). Function and regulation of AUTS2, a gene implicated in autism and human evolution. *PLoS Genetics*, *9*(1), e1003221. doi:10.1371/journal.pgen.1003221]. NO performed all of the experiments and wrote the article. NA assisted with planning the experiments and writing the paper. LS, in the lab of JW used a bioinformatics approach to assist analyzing public data and planning experiments.

Chapter 3: Work from this chapter is adapted from the following submitted article: [Nir Oksenberg, Genevieve D. E. Haliburton, Walter Eckalbar, Sierra Nishizaki, Karl Murphy, Katherine S. Pollard, Ramon Y. Birnbaum, Nadav Ahituv. (2014) Genome-wide distribution of Auts2 binding localizes with active neurodevelopmental genes]. NO performed the wet lab experiments with the assistance of RB, SN and KM. GH and WE analyzed the ChIP-seq data and performed bioinformatics analyses. NA and KP assisted NO and GE in writing the manuscript and planning experiments. NO and GE are co-first authors.

The Function and Regulation of *AUTS2*

Nir Oksenberg

Abstract

The autism susceptibility candidate 2 (*AUTS2*) gene is associated with multiple neurological diseases, including autism, and has been implicated as an important gene in human-specific evolution. In this thesis, I begin (chapter 1) with an introduction reviewing the literature regarding *AUTS2*, including its discovery, expression, association with autism and other neurological and non-neurological traits, implication in human evolution, function, regulation, and genetic pathways. Part of the research included in this introduction was performed by me and co-authors, and described in detail in the following chapters.

In chapter 2, I investigate the expression, function, and regulation of *auts2*. *auts2* is expressed primarily in the zebrafish brain. Knockdown of this gene in zebrafish leads to a smaller head size, neuronal reduction and decreased mobility. I identified twenty-three functional zebrafish enhancers, ten of which are active in the brain. Mouse enhancer assays characterized three brain enhancers that overlap an ASD-associated deletion and four enhancers that reside in regions implicated in human evolution, two of which are active in the brain. Combined, I show that *AUTS2* is important for neurodevelopment and expose candidate enhancer sequences in which nucleotide variation could lead to neurological disease and human-specific traits.

In chapter 3, I investigate the regulatory role and targets of Auts2. Using ChIP-seq and RNA-seq on mouse embryonic day 16.5 forebrains, we I elucidated the gene regulatory networks of Auts2. It was found that the majority of promoters bound by Auts2 belong to genes highly expressed in the developing forebrain, suggesting that Auts2 is involved in transcriptional

activation. Auts2 non-promoter bound regions significantly overlap developing brain-associated enhancer marks and are located near genes involved in neurodevelopment. Auts2 marked sequences are enriched for binding site motifs of neurodevelopmental transcription factors, including Pitx3 and TCF3. I characterized ten non-coding Auts2 marked sites near critical ASD-related genes for enhancer activity in zebrafish, four of which showed positive enhancer activity. Additionally, I characterized two of the positive brain enhancers near *NRXN1* and *ATP2B2* in mice. The results implicate Auts2 as an active regulator of important neurodevelopmental genes and pathways and identify novel genomic regions which could be associated with ASD and other neurodevelopmental diseases. In summary, this thesis investigates the function of *AUTS2*. I conclude that this gene is critical for the proper development of neurons, and may act as a cofactor to positively regulate genes expressed in the forebrain involved in neurodevelopment.

Table of contents

List of tables

List of figures

Chapter 1 – Introduction

1.1 Overview

The autism susceptibility candidate 2 (*AUTS2*) gene is associated with multiple neurological diseases, including autism, and has been implicated as an important gene in human-specific evolution. Recent functional analysis of this gene has revealed a potential role in neuronal development. In this chapter, I review the literature regarding *AUTS2*, including its discovery, expression, association with autism and other neurological and non-neurological traits, implication in human evolution, and genetic pathways.

1.2 Neurodevelopmental disorders

Neurodevelopmental disorders are characterized by motor, speech, cognitive and behavioral dysfunctions caused by impairment in growth and development of the central nervous system (CNS). Neurodevelopmental disorders encompass, but are not limited, to intellectual disability (ID), developmental delay (DD), and autism spectrum disorders (ASDs) (Fleischhacker & Brooks, 2006). ASDs are known as pervasive developmental disorders that are common (1/88 in the United States) (Baio, 2012) and highly heritable (Risch et al., 1999). ASDs are characterized by variable deficits in social communication, language, and restrictive and repetitive behaviors and present as a wide spectrum of phenotypes (Geschwind, 2009). Other neurological abnormalities such as ID, DD, epilepsy, sensory and motor abnormalities, gastrointestinal phenotypes, developmental regression, sleep disturbance, mood disorders, conduct disorders, aggression, and attention deficit hyperactivity disorder (ADHD) are also frequently associated with ASD (Geschwind, 2009). ASDs are heritable, complex disorders.

Several different genes have been implicated in these disorders containing either common variants with small effects or rare variants with larger consequences (Abrahams & Geschwind, 2008; Stein, Parikshak, & Geschwind, 2013). Over the years, studies examining individual patients, along with advances in sequencing technologies that have allowed the examination of a large number of individuals, have produced a myriad of new ASD, ID, and DD candidate genes, including *AUTS2*.

1.3 The discovery of *AUTS2*

AUTS2 was first identified in 2002 when it was found to be disrupted due to a balanced translocation in a pair of monozygotic (MZ) twins with ASD (Sultana et al., 2002). *AUTS2* was mapped to 7q11, spans 1.2 megabases and is approximately 340 kilobases upstream from the Williams–Beuren syndrome (WBS) critical region, a region that when deleted, causes a neurodevelopmental disorder characterized by a distinctive "elfin" facial appearance, a cheerful demeanor, developmental delay, strong language skills, and cardiovascular problems (Martens, Wilson, & Reutens, 2008). The AUTS2 protein sequence is highly conserved, with 62% amino acid conservation between humans and zebrafish (Kalscheuer et al., 2007). It contains regions of homology to other proteins, such as the Dwarfin family consensus sequence, human topoisomerase, and Fibrosin (FBRS), a fibroblast growth factor (Sultana et al., 2002). Additionally, the drosophila gene *tay* has limited similarity to *AUTS2*. *tay* mutants have reduced walking speed and activity, thought to be associated with structural defects in the protocerebral bridge (Poeck, Triphan, Neuser, & Strauss, 2008). Sequence analysis of AUTS2 identified no membrane-spanning domains, but identified two proline-rich domains and a predicted PY motif (PPPY) at amino acids 515-519 (Figure 1) (Sultana et al., 2002). The PY motif is a potential

WW-domain-binding region that is involved in protein-protein interactions and is present in the activation domain of various transcription factors, suggesting that AUTS2 may be involved in transcriptional regulation (Kalscheuer et al., 2007). Other predicted protein motifs include several cAMP and cGMP-dependent protein kinase phosphorylation sites and putative N-glycosylation sites (Sultana et al., 2002). In addition, AUTS2 has eight cac (His) repeats (Figure 1) (Sultana et al., 2002), which has been shown to be associated with localization at nuclear speckles (Salichs, Ledda, Mularoni, Albà, & de la Luna, 2009), subnuclear structures where components of the RNA splicing machinery are stored and assembled (Lamond & Spector, 2003). Evidence of nuclear localization sequences as well as several predicted protein-protein interaction domains (SH2 and SH3) were also observed for this protein (Figure 1). No evidence was found for any signal peptide in AUTS2, suggesting that it is not secreted or exposed to the cellular membrane (Bedogni, Hodge, Nelson, et al., 2010). No DNA-binding domains have been identified. Taken together, sequence analysis has revealed limited insight into the function of this gene.

1.4 *AUTS2* is a nuclear protein that is expressed in the central nervous system

Multiple reports have characterized the expression of *AUTS2* in different organisms, concluding that it is primarily expressed in the brain. Northern blot shows strong *AUTS2* expression in human fetal brain in the frontal, parietal and temporal regions, but not in the occipital lobe. Expression was also identified in the skeletal muscle and kidney with lower expression in the placenta, lung and leukocytes (Sultana et al., 2002). In human post-mortem fetal brain, *AUTS2* mRNA expression was found in the telencephalon (uniformly), ganglionic eminence, cerebellum anlagen and, more weakly, in the medulla oblongata at 8 weeks. *AUTS2*

was also found to be strongly expressed in the cortical plate and ventricular zone. Twenty-three week-old human fetal brains showed *AUTS2* expression in the dentate gyrus, CA1 and CA3 pyramidal cell subregions, the ganglionic eminence, caudate nucleus and putamen nuclei (Lepagnol-Bestel et al., 2008). *AUTS2* was also shown to be expressed in the neocortex and prefrontal cortex up to the late mid-fetal stage (Y. E. Zhang, Landback, Vibranovski, & Long, 2011). Gene expression profiles from ten human ocular tissues found *AUTS2* to be the 20th highest expressed gene in the sclera (Wagner et al., 2013). Sequencing of total RNA from human brain and liver found a large fraction of reads (up to 40%) to be within introns (Ameur et al., 2011). The authors identified enrichment of intronic RNA in brain tissues, particularly for genes involved in axonal growth and synaptic transmission. *AUTS2* was among the ten genes with the highest intronic RNA score in fetal brain. Three of the top ten genes (neurexin 1 (*NRXN1*), protocadherin 9 (*PCDH9*), and methionine sulfoxide reductase A (*MSRA*) have also been implicated in autism. Additionally, for long introns, including the first half of *AUTS2*, there is a 5' to 3' slope in read coverage with significantly higher levels of RNA present in the 5' end. The authors reason that in the fetal brain, intronic RNAs are subjected to brain specific regulatory pathways that regulate alternative splicing programs to control neuronal development (Ameur et al., 2011).

A detailed analysis of *Auts2* mRNA and protein expression in the developing mouse brain was published in 2010 (Table 1) (Bedogni, Hodge, Nelson, et al., 2010). The authors found that *Auts2* is expressed in the developing cerebral cortex and cerebellum and is located in the nuclei of neurons and some neuronal progenitors (Table 1). *Auts2* expression was identified in numerous neuronal cell types, including glutamatergic neurons (cortex, olfactory bulb, hippocampus), GABAergic neurons (Purkinje cells), and tyrosine hydroxylase (TH) positive

dopaminergic neurons (substantia nirgra and ventral tegmental area). The authors showed colocalization of *Auts2* with only a subset of eomesodermin (Tbr2) and paired box 6 (Pax6) positive cells in the ventricular and subventricular zones, suggesting that *Auts2* might be expressed in the transition between radial glial and intermediate progenitors (Bedogni, Hodge, Nelson, et al., 2010). The authors suggest that *Auts2* and T-box, brain, 1 (*Tbr1*) are coexpressed mostly in glutamatergic neuron populations in the forebrain and other transcription factors likely influence expression of *Auts2* in other regions. The authors also note that *Auts2* could be expressed in a transient phase of neuronal maturation or differentiation in the cortex (Bedogni, Hodge, Nelson, et al., 2010). In summary, AUTS2 has been shown to be a nuclear protein that is primarily expressed in the brain in various cell types as well as regions implicated in ASD such as the neocortex.

1.5 *AUTS2* and autism spectrum disorder, intellectual disability, and developmental delay

AUTS2 has been repeatedly implicated as an ASD candidate gene in recent years. Following the initial finding of an *AUTS2* translocation in twins with autism (Sultana et al., 2002), over 50 unrelated individuals with ASD, ID, or DD were identified with distinct structural variants disrupting the *AUTS2* region in numerous different reports (Figure 2) (Bakkaloglu et al., 2008; Ben-David et al., 2011; Beunders et al., 2013; Cuscó, Medrano, & Gener, 2009; Girirajan et al., 2011, 2013; Glessner et al., 2009; Huang, Zou, Maher, Newton, & Milunsky, 2010; Jolley et al., 2013; Kalscheuer et al., 2007; Nagamani et al., 2013; Pinto et al., 2010; Talkowski et al., 2012; Tropeano et al., 2013). Some of the structural variants are exclusively noncoding, suggesting that improper regulation and subsequent expression of *AUTS2* could be involved in the progression of the individual's disorder (Oksenberg, Stevison, Wall, & Ahituv, 2013). Along

with ASD, ID, and DD, many of these individuals also have other phenotypes including epilepsy, brain malformations, or dysmorphic features. One group described an "AUTS2 syndrome" in individuals with varying severity of growth and feeding problems, neurodevelopmental features, neurological disorders, dysmorphic features, skeletal abnormalities, and congenital malformations (Beunders et al., 2013). The spectrum of phenotypes observed in individuals with *AUTS2* mutations is consistent with the wide range of ASD phenotypes. This suggests that *AUTS2* is not associated with a specific sub-type of ASD. It has also been noted that dysmorphic features were more pronounced in individuals with 3' *AUTS2* deletions, where most of the coding region resides (Beunders et al., 2013). However, copy number variations (CNVs) at the *AUTS2* locus have also been observed in unaffected individuals, indicating that structural rearrangements are tolerated in some cases (Bakkaloglu et al., 2008; Redon et al., 2006). This suggests that disruptions in *AUTS2* may lead to neurodevelopmental disorders by being one of multiple genomic 'hits'. The large number of independent publications implicating *AUTS2* with ASD, ID, or DD provide strong evidence for its involvement in these disorders. It is worth noting however, that no publication has shown single base pair variants in the *AUTS2* locus affiliated with ASD, despite numerous ASD-related exome sequencing studies (Chahrour et al., 2012; O'Roak et al., 2011, 2012; Sanders et al., 2012).

The observation that *AUTS2* variants are mostly CNVs may be due to the susceptibility of this region to chromosomal breakpoints. A 2011 report showed that the offspring of older male mice have an increased risk of *de novo* CNVs in certain locations, including the *Auts2* locus (Flatscher-Bader et al., 2011). Another report found that hydroxyurea, a ribonucleotide reductase inhibitor, as well as aphidicolin, a DNA polymerase inhibitor, induce a high frequency of *de novo* CNVs in cultured human cells and found a clustering of CNVs in *AUTS2* (M. Arlt &

Ozdemir, 2011). Aphidocolin also induced CNV formation in the *Auts2* locus in nonhomologous end joining deficient mouse embryonic stem cells (M. F. Arlt, Rajendran, Birkeland, Wilson, & Glover, 2012). Because the *AUTS2* locus is a hotspot for CNVs, and individuals with ASD generally carry more CNVs than their unaffected siblings (Sebat et al., 2007), examining if these high numbers of ASD associated CNVs around *AUTS2* are consequential, and not merely a result of their susceptibility to CNVs, warrants investigation. There is also the possibility that these CNVs are affecting regulatory regions of other genes, including the nearby WBS critical region.

In 2013, a genome wide analysis of DNA methylation was published on ASD discordant and concordant MZ twins. A region in the *AUTS2* promoter (chr7: 68701907; hg18) was the 42nd most differentially methylated CpG site in the genome, suggesting that not only sequence variation, but also epigenetic changes to the *AUTS2* locus could be involved in the development of ASD related traits (Wong et al., 2013). Significant DNA methylation differences were often observed near other genes that have been previously implicated in ASD, including *methyl-CpG binding domain protein 4* (*MBD4*) and *microtubule-associated protein 2* (*MAP2*). The authors cautioned, however, that it is difficult to draw conclusions about the causality of the differentially methylated sites due to small sample size, lack of corresponding RNA expression data, the use of whole blood rather than brain tissue, and potential epigenetic effects due to medicine (Wong et al., 2013).

Combined, the evidence for a causative role of *AUTS2* in DD and ID is quite convincing. However, for ASD the evidence presented so far suggests that disruptions in *AUTS2* can play a causative role, but to demonstrate causality, more research needs to done on cohorts of well-defined ASD patients and on the functional consequence of these disruptions.

1.6 *AUTS2* and other neurological conditions

In addition to ASD, ID, and DD, *AUTS2* has been implicated in other neurological disorders. Some of these disorders, like epilepsy, have been shown to be linked to ASD. However, other *AUTS2*-associated phenotypes are ASD independent. *AUTS2* expression was found to have significant association with nicotine dependence, cannabis dependence and antisocial personality disorder, although this study had a small numbers of cases and would need to be repeated with larger cohorts (Philibert et al., 2007). The study also suggested, although it did not quite reach significance, that *AUTS2* expression is implicated in alcohol dependence (Philibert et al., 2007). In 2011, a genome-wide association meta-analysis found an *AUTS2* non-coding SNP, rs6943555, to be significantly associated with alcohol consumption (Schumann et al., 2011). The authors also reported increased *AUTS2* expression in carriers of the minor A allele of rs6943555 compared with the T allele in 96 human prefrontal cortex samples. Additionally, they identified significant differences in expression of *Auts2* in whole-brain extracts of mice with differences in voluntary alcohol consumption. The authors also showed that down-regulation of *tay*, which has sequence similarity to *AUTS2*, caused reduction in alcohol sensitivity in *Drosophila* (Schumann et al., 2011). Also implicating *AUTS2* in drug dependence was a 2011 study showing that *AUTS2* has a 3.01 fold change (downregulation) between 19 male heroin-dependent individuals and 20 controls in lymphoblastoid cell lines (Liao, Cheng, & Lai, 2011). A follow-up study compared *AUTS2* transcript levels of lymphoblastoid cell lines between 124 heroin-dependent and 116 control males using quantitative PCR and found that average transcript levels of *AUTS2* in the heroin-dependent group was significantly lower than controls. They also found that AA homozygotes for rs6943555 were significantly over-

represented in the heroin-dependent subjects (Chen, Liao, Lai, & Chen, 2013). Taken together, these reports show strong evidence for *AUTS2*'s involvement in addiction and dependence.

Additionally, the *AUTS2* locus has been shown to be implicated or altered in individuals with schizoaffective disorder (Hamshere et al., 2009), bipolar disorder (Hattori et al., 2009; Lee, Woo, & Greenwood, 2012), epilepsy (Mefford et al., 2010), ADHD (Elia et al., 2010), differential processing speed (Luciano, Hansell, & Lahti, 2011), suicidal tendencies under the influence of alcohol (Chojnicka et al., 2013), and dyslexia (Girirajan et al., 2011), either through CNV or genome-wide association studies. A 2012 article sequenced balanced chromosomal abnormalities in patients with neurodevelopmental disorders and found the *AUTS2* locus to be perturbed in individuals with microcephaly, macrocephaly, ataxia, visual impairment, language disability, seizure disorder, dysmorphic features, behavioral problems, motor delay, and Rubinstein-Taybi syndrome (Talkowski et al., 2012). It could be that the observation that most cases of *AUTS2* structural variants are associated with ASD is attributed to more individuals with ASD being tested in this locus than patients with other neurological disorders leading to an underestimate in the link between *AUTS2* and other neurological phenotypes. Taken together, these observations suggest that *AUTS2* dysfunction is not restricted to ASD, DD or ID, but rather involved in a wide-range of neurological disorders. Additionally, there are a few studies implicating *AUTS2* in non-neurological disorders and traits.

1.7 *AUTS2* and non-neurological disorders and traits

A few reports have implicated *AUTS2* in non-neurological disorders and traits. In 2004, a report examined 18 cases of childhood hyperdiploid acute lymphoblastic leukemia (ALL) cases in order to identify the relationship between extra copies of chromosomes and increased gene expression. The authors identified multiple regions with increased expression that correlated

poorly or not at all with the presence of extra copies of chromosomes, including 7q11.2. *AUTS2* showed consistently higher expression levels in the cDNA samples of patients than in normal mononuclear cells, possibly implicating the gene in ALL (Gruszka-Westwood et al., 2004). In 2008, it was reported that paired box 5 (PAX5) can be rearranged with a variety of partners, including *AUTS2* (one case) in pediatric ALL (Kawamata et al., 2008). A couple of years later, a second case of *PAX5-AUTS2* fusion was identified in pediatric ALL (Coyaud et al., 2010). In 2012, the third case of *PAX5-AUTS2* fusion was identified in a patient with pediatric ALL, providing additional evidence that *PAX5-AUTS2* is a recurring gene fusion in ALL (Denk et al., 2012). Two of the three *PAX5-AUTS2* cases had CNS diseases either at the time of diagnosis or relapse (Denk et al., 2012). Individual reports, some of which identify single patients, have also implicated the *AUTS2* locus in the aging of human skin (Lener et al., 2006), lung adenocarcinoma (Weir, Woo, Getz, & Perner, 2007), lethal prostate cancer (Penney et al., 2010), the number of corpora lutea in pigs (Sato, Hayashi, & Kobayashi, 2011), early-onset androgenetic alopecia (R. Li et al., 2012), and metastatic nonseminomatous testicular cancer (Stadler et al., 2012). Despite a number of reports suggesting *AUTS2*'s role in non-neurological disorders and traits, disruption of *AUTS2* is most often reported to be associated with neurological phenotypes.

1.8 *AUTS2* and human evolution

In 2006, a comparative genomics approach was used to search the human genome for regions that have significantly changed in humans in the last 5 million years since the divergence from chimpanzees, but are highly conserved in other species (Pollard, Salama, King, et al., 2006; Pollard, Salama, Lambert, et al., 2006). They identified 202 such regions that they termed human

accelerated regions (HARs). These HARs are great candidates for sequences responsible for the evolution of human-specific traits. An intronic region in *AUTS2* ranked as the 31st most accelerated region in their study. Similarly, in 2006 a different group combed the genome for conserved non-coding sequences that were accelerated in the human lineage (Prabhakar, Noonan, Pääbo, & Rubin, 2006). The authors identified 902 human accelerated conserved non-coding sequences (HACNSs). HACNS 174 and 369 both lay within introns of *AUTS2*. With the publication of the draft sequence of the Neanderthal genome in 2011, it was found that the first half of *AUTS2* displayed the strongest statistical signal in a genomic screen differentiating modern humans from Neanderthals (Green et al., 2010). This region contains 293 consecutive SNPs where only ancestral alleles were observed in the Neanderthals, only two of which are coding variants (a G to C nonsynonymous substitution at chr7:68,702,743 (hg18) only in the Han Chinese and a C to T synonymous change at chr7:68,702,866 (hg18) within the Yoruba and Melanesian populations). Other regions that were found to have the most significant human-Neanderthal changes also include genes that are involved in cognition and social interaction, including dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 1A (*DYRK1A*), neuregulin *3* (*NRG3*) and Ca++-dependent secretion activator 2 (*CADPS2*) (Green et al., 2010). The authors conclude that multiple genes involved in cognitive development were positively selected during the evolution of modern humans (Green et al., 2010). Taken together, these studies suggest that significant changes in *AUTS2* occurred specifically in modern humans and it is conceivable, based on the neurological role that this gene plays, that these changes could lead to cognitive traits specific to humans.

1.9 *AUTS2* gene pathways

A 2010 study used radiation hybrid genotyping data to test for interaction of 99% of all possible gene pairs across the mammalian genome (Lin, Wang, Ahn, Park, & Smith, 2010). *AUTS2* was the known gene with the greatest number of edges, or connectivity (Lin et al., 2010). Despite that finding, little is known about the genetic pathways that *AUTS2* is involved in. However, a few articles have provided evidence linking AUTS2 to other proteins and pathways.

One potential pathway was revealed by examining genes that can oscillate expression during somitogenesis. Two papers found that the expression of *AUTS2* oscillates in phase with other notch pathway genes, suggesting that it is a component of the notch signaling pathway (Dequéant et al., 2006; William et al., 2007). Notch signaling has been shown to be involved in neuronal migration through its interaction with *Reelin*, a gene implicated in ASD and a target of *Tbr1* (Hashimoto-Torii et al., 2008; Wang et al., 2004).

Although not reaching significance, a group found that *Auts2* has a 1.33 fold change in cerebellar gene expression in methyl CpG binding protein 2 (*Mecp2*) null mice. Loss of *MECP2* function can cause a number of neurodevelopmental disorders including Rett syndrome and autism (Ben-Shachar, Chahrour, Thaller, Shaw, & Zoghbi, 2009). The authors also compared their data with data generated from other gene expression studies. They found that *Auts2* is consistently altered in both their datasets, as well as in post-mortem Rett syndrome patient brains and mutated in fibroblast and lymphocytes (Ben-Shachar et al., 2009).

Starting at mouse E12, *Auts2* mRNA is expressed in the cortical preplate, where it colocalizes with *Tbr1*, a transcription factor that exerts positive and negative control of regional and laminar identity in postmitotic neurons (Bedogni, Hodge, Elsen, et al., 2010; Bedogni, Hodge, Nelson, et al., 2010). Using Tbr1 antibodies for chromatin immunoprecipitation (ChIP)

of E14.5 cortex, it was shown that the *Auts2* promoter is a direct transcriptional target of Tbr1 in the developing neocortex and is involved in frontal identity (Bedogni, Hodge, Elsen, et al., 2010).

SATB homeobox 2 (*Satb2*) is one of four genes (including *Tbr1*) that regulates projection identity within the layers of the mammalian cortex. In 2012, a report showed that in mice, Tbr1 expression is dually regulated by *Satb2* and B-cell CLL/lymphoma 11B (*Ctip2*) in cortical layers 2-5. The authors also demonstrated that *Satb2* regulates *Auts2.* They showed that like Tbr1, Auts2 is expressed in the deep and upper layers of the cortex. They investigated whether the loss of Tbr1 expression in the upper layer neurons in *Satb2* mutants coincides with changes in Auts2 expression. They observed that there was a significant loss of Auts2 expression in the upper layers of *Satb2* mutants similar to the loss of Tbr1 in *Satb2* mutants. The authors did not observe any changes in Auts2 expression in layers 5 or 6. Their results suggest that *Satb2* regulates the expression of Tbr1, which in turn regulates Auts2 expression in callosal projection neurons (Srinivasan et al., 2012).

GTF2I repeat domain containing 1 (*GTF2IRD1*) is one of 26 genes deleted in WBS and encodes a putative transcription factor expressed throughout the brain during development. *Gtf2ird1* knockout mice display reduced innate fear and increased sociability, phenotypes consistent with WBS (Young et al., 2008). Microarray screens were used to find transcriptional targets of *Gtf2ird1* on brain tissue from *Gtf2ird1* knockout mice at two time points, E15.5 and birth (P0), compared to wild type littermates. *Auts2* was one of only two genes identified in both (E15.5 and P0) microarray experiments to be altered compared to controls. In P0 mouse brains of knockout mice, *Auts2* was increased by 1.3 fold, and in contrast in E15.5 embryos, it was

decreased by 1.5 fold (O'Leary & Osborne, 2011). It is unclear if *AUTS2* is a target of *Gtf2ird1*, or if this observation is caused by the proximity of the two genes.

Zinc finger, matrin-type 3 (also known as *wig-1*), a transcription factor regulated by p53, plays an important role in RNA protection and stabilization and as part of the p53 pathway, is a casual factor in neurodegenerative diseases. wig-1 down regulation by antisense oligonucleotide treatment led to a significant reduction in *Auts2* mRNA levels in the brains of BACHD mice, a mouse model for Huntington's disease. The authors also reported a trend in reduction of *Auts2* mRNA levels in the livers of BALB/c mice but no reduction in *Auts2* levels in FVB (background strain of BACHD) mouse brains (Sedaghat, Mazur, Sabripour, Hung, & Monia, 2012). These results suggest a role for *wig-1* in the regulation of *Auts2* expression and further links *Auts2* with pathways involved in the CNS.

Polycomb Repressive Complex 1 *(PRC1)* is a polycomb-group (PcG) gene which acts as a developmental regulator through transcriptional repression. It is critical for many biological processes in mammals, including differentiation. There are six major groups of PRC1 complexes, each containing a distinct PCGF subunit (PCGF1-6), a RING1 A/B ubiquitin ligase, and unique associated polypeptides. Using tandem affinity purification of PCGF3 and PCGF5, AUTS2 was recovered, implicating a role for AUTS2 in transcriptional repression during development (Gao et al., 2012).

In 2013, the regulatory pathway for *SEMA5A*, an autism candidate gene, was mapped *in silico* using expression quantitative trait locus (eQTL) mapping. The authors found that the *SEMA5A* regulatory network significantly overlaps rare CNVs around ASD-associated genes, including *AUTS2*. Given the extensive *trans*-regulatory network associated with *SEMA5A*, the authors also investigated the possibility that there are a number of upstream master regulators

that control this network. Performing eQTL mapping for expression levels of the eQTL-associated genes within the network (eQTLs of the eQTLs of *SEMA5A*), the authors identified 12 regions associated with the expression of 10 or more primary *SEMA5A* eQTL genes, including *AUTS2*. This study suggests that *AUTS2* is involved, and may be a master regulator in ASD-related pathways (Cheng, Quinn, & Weiss, 2013).

1.10 Concluding remarks

As we identify the genes involved in ASD, DD and ID, our ability to genetically diagnose these disorders improves, and future screens should assess *AUTS2* for potential causative CNVs. However, before we are able to use *AUTS2* as a diagnostic tool, we must determine what makes a CNV in or around *AUTS2* causative or benign and for what disorders (e.g. ID, DD, ASD, ASD with ID/DD, etc). This includes a deeper investigation of the function and regulatory network of this gene, which was the focus of my thesis and described in Chapter 2. A major step in developing future ASD and ASD-related phenotype treatments relies on a solid understanding of the pathways involved and how they interact, which is discussed in detail in Chapter 3. Obtaining a better understanding of the pathways associated with *AUTS2* will allow us to better comprehend the biological systems that can be perturbed when the function of this gene is disrupted, as well as how nucleotide changes in it led to human-specific traits. Multiple reports have implicated *AUTS2* in addiction and other neurological phenotypes, but the mechanism and certainty of these involvements remain unclear, highlighting the importance of deeper investigations into the function of this gene and its role in development and disease. Future work using an *Auts2* mouse knockout should reveal greater detail of the function of this gene. In summary, we can presume that this gene is involved in neurodevelopment, and may play a role

in ASD and ASD-related phenotypes. There is also significant data suggesting that *AUTS2* has human-specific variants that could possibly contribute to human cognition.

Chapter 1 tables

Table 1: *Auts2* expression in the developing mouse brain (summary based on (Bedogni, Hodge, Nelson, et al., 2010))

Time point	*Auts2* expression
E11	mRNA barely detectable.
E12-13	Colocalization with Tbr1 in the cortical preplate. Tbr1 is a transcription factor specific for postmitotc projection neurons.
E12-14	High expression in the developing cortex, thalamus and cerebellum. There is continued expression in these regions throughout development, but levels fluctuate and are found in gradients. Different markers show *Auts2* expression in multiple neuronal subtypes in the developing cortex.
E14	Expression in the hippocampal primordium. Transient expression in the locus ceruleus and vestibular nuclei.
E16	Expression in the cerebral cortex is now a gradient of high rostral to low caudal expression.
E19	Highest expression in inferior and superior colliculi and the pretectum.
P0	*Auts2* expression becomes progressively more superficial in the frontal cortex. Coexpression with Tbr1 becomes rare as Tbr1 becomes more selective to layer 6.
E16-P21	*Auts2* is expressed mostly in the frontal cortex, hippocampus, and the cerebellum. In addition, high expression levels were detected in the developing dorsal thalamus, olfactory bulb, inferior colliculus and the substantia nigra.
P21	Expression in developing thalamic areas, including the anterior thalamic nucluei and in ventrolateral/ventromedial nuclei. *Auts2* is restricted to superficial layers in frontal cortex. *Auts2* is expressed throughout the subgranular zone and the granule cell layer of the hippocampus.

Chapter 1 figures

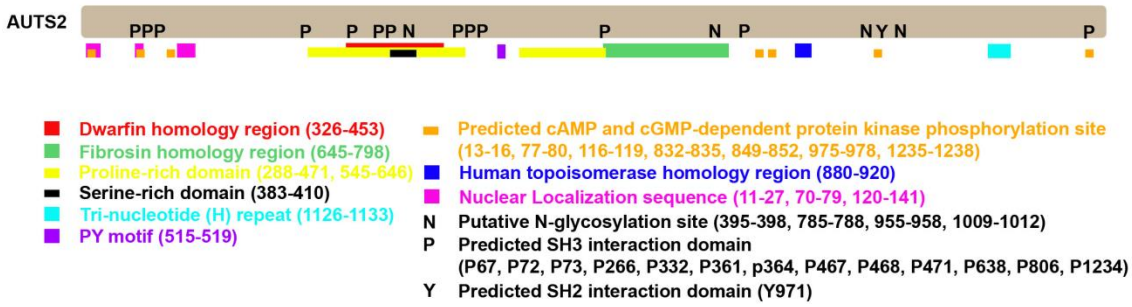


Figure 1: Schematic of the AUTS2 protein

AUTS2 (1,259 amino acids) is shown as a grey bar. The locations of predicted domains, motifs, regions of homology, and other characterized sequences are shown below and within the protein. Numbers in parenthesis represent the amino acid location. The figure is based on predicted features in (Bedogni, Hodge, Nelson, et al., 2010; Sultana et al., 2002).

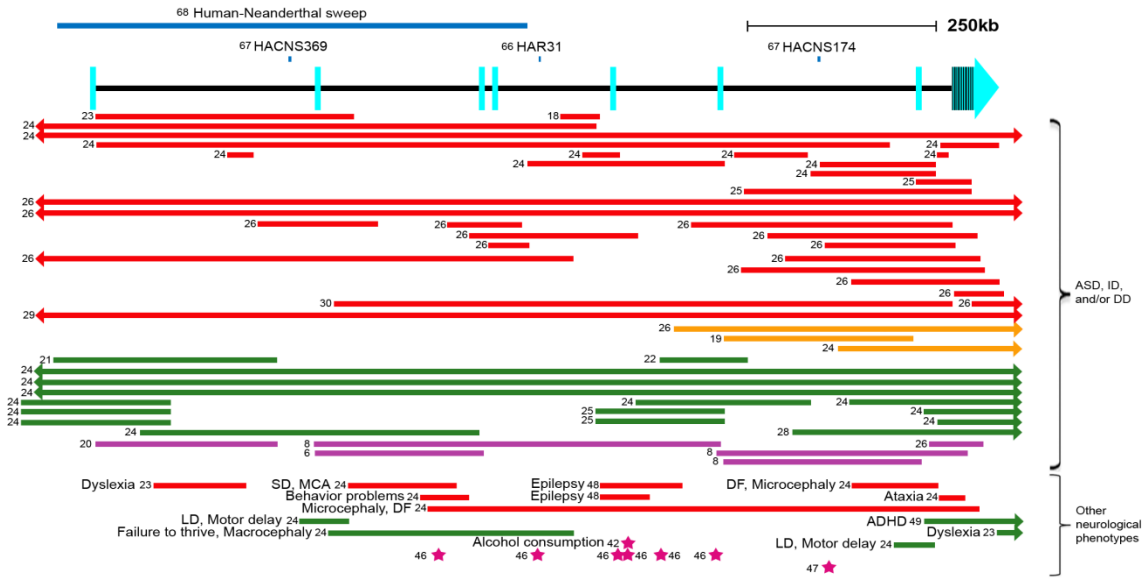


Figure 2: Schematic of the *AUTS2* genomic region

Numbers to the left of the lines correspond to reference numbers published in (Oksenberg & Ahituv, 2013). Human accelerated sequences are shown as blue lines above the gene (Green et al., 2010; Pollard, Salama, King, et al., 2006; Prabhakar et al., 2006). Structural variants (Bakkaloglu et al., 2008; Ben-David et al., 2011; Beunders et al., 2013; Cuscó et al., 2009; Elia et al., 2010; Girirajan et al., 2011; Glessner et al., 2009; Huang et al., 2010; Jolley et al., 2013; Kalscheuer et al., 2007; Mefford et al., 2010; Nagamani et al., 2013; Pinto et al., 2010; Sultana et al., 2002; Talkowski et al., 2012; Tropeano et al., 2013) are represented as colored lines (red: deletion, orange: inversion, green: duplication, purple: translocation). SNPs are shown as magenta stars. rs6943555 is associated with alcohol consumption (Schumann et al., 2011). SNPS in (Hattori et al., 2009; Lee et al., 2012) are associated with bipolar disorder. SNPs in (Hattori et al., 2009) are reported to be in strong linkage disequilibrium with each other. Arrows in bars signify that the structural variant extends past the gene in that direction. Exons are depicted as light blue rectangles, as defined by the RefSeq genes track in the UCSC Genome Browser. Human Accelerated Conserved Non-coding Sequence (HACNS), Human Accelerated Region

(HAR), developmental delay (DD), intellectual disability (ID), dysmorphic features (DF), seizure disorder (SD), multiple congenital anomalies (MCA), language disability (LD). Figure adapted from (Oksenberg et al., 2013).

Chapter 2 – The function and regulation of *AUTS2*

2.1 Abstract

Nucleotide changes in the *AUTS2* locus, some of which affect only noncoding regions, are associated with autism and other neurological disorders including attention deficit hyperactivity disorder, epilepsy, dyslexia, motor delay, language delay, visual impairment, microcephaly, and alcohol consumption. In addition, *AUTS2* contains the most significantly accelerated genomic region differentiating humans from Neanderthals, which is primarily composed of noncoding variants. However, the function and regulation of this gene remained largely unknown and were the main focus of my thesis. To characterize *auts2* function, I knocked it down in zebrafish, which lead to a smaller head size, neuronal reduction and decreased mobility. To characterize *AUTS2* regulatory elements, I tested sequences for enhancer activity in zebrafish and mice. I identified twenty-three functional zebrafish enhancers, ten of which were active in the brain. Mouse enhancer assays that I carried out characterized three mouse brain enhancers that overlap an ASD-associated deletion and four enhancers that reside in regions implicated in human evolution, two of which are active in the brain. Combined, my results show that *AUTS2* is important for neurodevelopment and expose candidate enhancer sequences in which nucleotide variation could lead to neurological disease and human-specific traits.

2.2 Introduction

Autism spectrum disorders (ASDs) are common (1/88 in the United States) (Baio, 2012) childhood neurodevelopmental disorders known as pervasive developmental disorders (reviewed

in (Pardo & Eberhart, 2007)). ASDs are highly heritable, signifying a substantial genetic etiology (Risch et al., 1999; Talkowski, Minikel, & Gusella, 2014). A balanced translocation involving the *autism susceptibility candidate 2* (*AUTS2*; GenBank NM_001127231.1) gene in a pair of monozygotic twins with ASD was the first to link this gene to autism (Sultana et al., 2002) (see Figure 2 from Chapter 1). Following this finding, thirty-six additional unrelated individuals with ASD, intellectual disability, or developmental delay were found to have distinct heterozygous structural variants disrupting the *AUTS2* region (Bakkaloglu et al., 2008; Ben-David et al., 2011; Girirajan et al., 2011; Glessner et al., 2009; Huang et al., 2010; Kalscheuer et al., 2007; Nagamani et al., 2013; Pinto et al., 2010; Talkowski et al., 2012), four exclusively in noncoding regions (Pinto et al., 2010; Talkowski et al., 2012). Additional structural variants in *AUTS2*, some of which are only intronic, were also shown to be associated with attention deficit hyperactivity disorder (ADHD) (Elia et al., 2010), epilepsy (Mefford et al., 2010; Talkowski et al., 2012), dyslexia (Girirajan et al., 2011), motor delay, language delay, visual impairment, microcephaly and others (Talkowski et al., 2012). In addition, a genome-wide association meta-analysis study identified SNP rs6943555 within the fourth intron of *AUTS2* to be the most statistically significant SNP associated with alcohol consumption (Schumann et al., 2011) (see Figure 2 from Chapter 1). These various *AUTS2*-associated phenotypes suggest this gene has an important neurological function. It is worth noting though that some individuals with disrupted *AUTS2* and mental retardation or autism have additional, potentially non-neuronal phenotypes, such as hypotonia, short stature, urogenital abnormalities, and skeletal abnormalities (Kalscheuer et al., 2007; Sultana et al., 2002)

In addition to *AUTS2's* role in neurological disease, it was also shown to be important for human-specific evolution. The first half of *AUTS2* displayed the strongest statistical signal in a

genomic screen differentiating modern humans from Neanderthals (Green et al., 2010). This is attributed to a stretch of 293 consecutive SNPs, only two of which are coding variants: (a G to C nonsynonymous substitution at chr7:68,702,743 (hg18) only in the Han Chinese and a C to T synonymous change in chr7:68,702,866 (hg18) within the Yoruba and Melanesian populations). Other regions identified to have the most significant human-Neanderthal sweeps also include genes that are involved in cognition and social interaction, including *DYRK1A, NRG3* and *CADPS2* (Green et al., 2010), reinforcing our interest in *AUTS2*'s role in cognition and human-Neanderthal differences. In addition, three different evolutionary conserved noncoding intronic regions in *AUTS2* (HAR31, HACNS174 and HACNS369) have been found to be significantly accelerated when compared to primates in two different studies (Pollard, Salama, King, et al., 2006; Prabhakar et al., 2006) (see Figure 2 from Chapter 1). Combined, these data suggest that altered regulation of *AUTS2* could be associated with human specific traits.

The functional role of *AUTS2* is not well known, although some studies have identified a putative role in transcriptional regulation during neuronal development. The predicted AUTS2 protein contains a PY motif, a putative WW-domain-binding region (Sultana et al., 2002) present in various transcription factors, implying that *AUTS2* may be involved in transcriptional regulation (Kalscheuer et al., 2007) (see Figure 1 from Chapter 1).

I used zebrafish morpholinos to functionally characterize *auts2*. I show that knocking down this gene leads to an overall stunted developmental phenotype that includes a smaller head, body and reduced movement. Further characterization of morphant fish revealed a reduction in developing midbrain neurons and also in sensory and motor neurons. To characterize *AUTS2* enhancers, I used both zebrafish and mouse transgenic enhancer assays. I identified three functional enhancers within an ASD-associated deletion and six brain enhancer in regions

associated with human specific evolution. Combined, I found that *AUTS2* is important for neuronal development and characterized several functional enhancers within this locus, where nucleotide changes could be associated with neurodevelopmental disease and human specific evolution.

2.3 Results

2.3a *auts2* zebrafish expression

Zebrafish can be an effective tool to study ASD (Tropepe & Sive, 2003). Using whole mount *in situ* hybridization, I determined that *auts2* is expressed in zebrafish at 24 hours post fertilization (hpf) in the forebrain, midbrain and hindbrain (Figure 1). Additionally, *auts2* is expressed in the trunk (including the spinal cord), with stronger expression towards the caudal peduncle. At 48 hpf, *auts2* is expressed in the brain and pectoral fin and from 72-120 hpf its expression is restricted primarily to the brain. *auts2* is also weakly expressed in the eye from 24-120 hpf. Overall, I observed that the zebrafish expression largely correlates with the previously characterized mouse expression (Bedogni, Hodge, Nelson, et al., 2010; Visel, Thaller, & Eichele, 2004).

2.3b Phenotypic characterization of *auts2* morphants

I next used morpholinos (MOs) to knockdown *auts2* in zebrafish during development. Fish injected with an *auts2* translational blocking MO displayed a stunted developmental phenotype with smaller heads, eyes, body and pectoral fins (Figure 1C and Figure 2A). A second *auts2* MO that disrupts the splice junction between intron two and exon three exhibited similar but less severe phenotypes (Figure 1D). These phenotypes appeared in 80-90% of injected fish

and were rescued by co-injecting the full length human *AUTS2* mRNA along with the translational blocking MO (68% of injected fish showed a partial to full rescue) (Figure 1E). Injection of a 5 base pair (bp) mismatch *auts2* translational MO control did not show any phenotype (Figure 1B and Figure 2A), further validating the specificity of our MOs to effectively knockdown *auts2* in zebrafish.

To further characterize the neurological function of *auts2*, I injected the translational MO into the HuC-GFP transgenic zebrafish line (Park et al., 2000), where developing neurons express green fluorescent protein (GFP). Compared to the 5bp mismatch control, translational MO injected fish showed a dramatic decrease in GFP at 48 and 72 hpf in the dorsal region of the midbrain, including the optic tectum, the midbrain-hindbrain boundary (which includes the cerebellum), the hindbrain and the retina (Figure 2B). This phenotype was also observed by staining neurons with Nissl at 48 hpf (Figure 3A). TUNEL staining of 48 hpf embryos revealed that morphant fish exhibit increased apoptosis in the midbrain in the same location where fewer neurons where observed (Figure 2C and Figure 3B). Anti-proliferating Cell Nuclear Antigen (PCNA) staining showed increased amounts of cell proliferation in morphant fish in the forebrain, midbrain and hindbrain (Figure 2D and Figure 3C-E). While seemingly contradictory, increased amounts of both TUNEL and PCNA positive cells has been previously shown, as cell death and proliferation could be coupled (Alenzi, 2004; Evan & Littlewood, 1998). It is conceivable that the increased PCNA positive cells are the result of morphant cells failing to differentiate into mature neurons, as seen in the HuC-GFP line. These results suggest that *auts2* may be involved in the production and maintenance of neurons in the zebrafish brain.

Both the translational and splicing morphant fish also showed a decreased movement response when gently prodded with a pipette tip compared to controls that began at 48 hpf. This

phenotype was observed until 120 hours when the zebrafish were euthanized. In order to determine whether motor neuron defects could explain this phenotype, I injected the translational MOs into the Tg(mnx1:GFP) zebrafish line, which expresses GFP in developing motor neurons (Flanagan-Steet, Fox, Meyer, & Sanes, 2005). At 48 hpf, morphant fish displayed fewer GFP labeled motor neuron cell bodies in the spinal cord. Additionally, motor neuron projections were weaker and perpendicular to the spinal cord, in contrast to the angled projections of the control injected fish (Figure 2E). This phenotype was also confirmed using the znp-1 antibody to mark motor neuron axons (Gordon, Gribble, Syrett, & Granato, 2012) in control and morphant fish. Morphant fish consistently showed more branching of axons compared to controls (Figure 4). To assess sensory neuron defects, Rohon-Beard neurons were stained with anti-HNK-1 in control and translational MO injected fish at 48 hpf. Morphant fish displayed on average 60% fewer sensory neurons in the spinal cord (Figure 2F). These results suggest that loss of *auts2* in zebrafish could lead to motor and sensory neuron defects, which may play a role in their reduced movement and decreased response to touch.

2.3c *AUTS2* enhancer characterization

Due to the observations that noncoding regions in the *AUTS2* locus are associated with neurological phenotypes and human-specific evolution, I set out to identify enhancers in this locus. To focus my search, limited my candidates to be between the first exon and fifth intron, due to this region encompassing the human-Neanderthal sweep (exon 1-4; chr7:68,662,946-69,274,862 (hg18)) (Green et al., 2010) and several noncoding nucleotide changes that have been associated with neurological phenotypes (Girirajan et al., 2011; Pinto et al., 2010; Schumann et al., 2011; Talkowski et al., 2012). *AUTS2* enhancer candidate (AEC) sequences were selected based on evolutionary conservation, embryonic mouse forebrain and midbrain ChIP-seq datasets

(Visel et al., 2009) and nucleotide variants that define the human-Neanderthal sweep (Green et al., 2010) (see methods). I also tested the human accelerated region (HAR) in intron four, HAR31 (Pollard, Salama, King, et al., 2006), and the human accelerated conserved non-coding sequences (HACNS) in introns one and six, HACNS 369 and HACNS 174 respectively (Prabhakar et al., 2006). Using these criteria, 40 AECs were selected for zebrafish enhancer assays (Table 1). These human sequences were cloned into the E1b-GFP-Tol2 enhancer assay vector and injected into zebrafish (Q. Li et al., 2010). Of the 40 candidates, 23 were found to be functional enhancers, 22 of which showed enhancer activity in locations that overlap *auts2* expression in zebrafish and 10 that were active in the brain (Table 1 and Figure 5).

To further characterize the regulatory elements within a 33,519bp deletion associated with ASD in *AUTS2* intron four (Pinto et al., 2010), the three positive zebrafish enhancers in this region (AEC27, AEC29, AEC32) were analyzed in mice using a similar transgenic assay (Visel, Minovitsky, Dubchak, & Pennacchio, 2007). AEC27 showed enhancer expression in the somitic muscle in zebrafish, while examination of its enhancer activity at E11.5 (hs658;(Visel et al., 2007)) found it to be active in the midbrain and neural tube (Figure 6). At E12.5, AEC29 had enhancer activity in the olfactory epithelium similar to zebrafish and also displayed enhancer expression in the eye (Figure 6). AEC32 recapitulated the zebrafish enhancer expression in the midbrain and hindbrain with additional enhancer expression in the forebrain at E12.5. Histological sections of AEC32 showed enhancer activity in the mouse cerebellum (Figure 6), a region thought to play a role in ASD (Pardo & Eberhart, 2007). The removal of these three brain enhancers and potentially other functional sequences in this region could contribute to the neurological phenotypes in patients with deletions in this intron.

I next set out to characterize enhancers in regions implicated in human-specific evolution. Four of the sixteen positive zebrafish enhancers identified in this region (Table 1 and Figure 5) were analyzed for enhancer activity in mice. These four sequences were positive mouse enhancers active in the brain, the otic vesicle, or eye (Figure 7). Interestingly, two of these enhancers (AEC10 and 21) show enhancer expression in the developing tectum, a region in the brain that is thought to control auditory and visual responses.

2.4 Discussion

Using MOs to knockdown *auts2*, I observed an overall phenotype of stunted development, making it difficult to characterize discrete phenotypes. The phenotypes observed in our knockdown, along with those from another group (Beunders et al., 2013), are summarized in Table 2. Using HuC, a neuronal marker, both groups observed a decrease in neuronal cells in the brain (Figure 2 and Figure 8), including in the cerebellum. The cerebellum is involved in cognitive and emotional function and has been repeatedly implicated in ASD (Pardo & Eberhart, 2007).

I showed morphant fish display increased apoptosis and cell proliferation in the brain, possibly as a result of morphant cells failing to differentiate into mature neurons, which matches the HuC results (Oksenberg et al., 2013). Although increased cell proliferation was observed in my study (Oksenberg et al., 2013), the other study described decreased cell proliferation (Beunders et al., 2013). The differences in this phenotype could be due to differences in the stains used (PCNA, which marks cells in early G1 and S phase versus phosphohistone-H3, a marker of cells in G2 and M phase). Both reports, however, found that *auts2* knock-down cells show more replicating DNA, but fewer cells dividing into daughter cells. The craniofacial

phenotype of the morphant fish was also characterized in the other study, finding that they have micrognathia (undersized jaw) and retrognathia (receded jaw) (Figure 8C) (Beunders et al., 2013). Given that migrating neural crest cells play an important role in craniofacial development (Gilbert, 2000), it is possible that this phenotype is a result of defects in neuronal cell development.

Additionally, less movement was reported in morphant fish in my study, which could be caused by fewer motor neuron cell bodies in the spinal cord along with improperly angled and weaker projections, and/or fewer sensory neurons, both of which were observed in morphant fish (Oksenberg et al., 2013). In addition, the cerebellum plays a major role in motor control, and it is possible that the defects detected in cerebellar neurons could partially explain the reduced movement phenotype observed in morphant fish. It is worth noting that two individuals with *AUTS2* structural variants had motor delay phenotypes (Figure 2, Chapter 1) (Talkowski et al., 2012).

While I observed an overall stunted development (Oksenberg et al., 2013), the other group reported a phenotype restricted to the brain and jaw (Figure 8C) (Beunders et al., 2013). A potential cause for the difference in this phenotype, along with the differences in cell proliferation phenotypes, could be due to the use of different morpholinos for these assays: an *auts2* translational morpholino (Oksenberg et al., 2013) versus splicing morpholinos (Beunders et al., 2013). Both groups were able to rescue the morphant phenotype by injecting the full length human *AUTS2 m*RNA along with the morpholino (Beunders et al., 2013; Oksenberg et al., 2013). Beunders et al. also rescued the phenotype by injecting the shorter, C-terminal isoform of *AUTS2*, suggesting that the final nine exons of *AUTS2* contains the critical region of the gene, at least for the dysmorphic phenotype observed in knock-down fish. This is in line with the group's

observation that dysmorphic features were more pronounced in individuals with 3' *AUTS2* deletions (Beunders et al., 2013). My *auts2* MOs were designed to disrupt *auts2* activity on chromosome 10 (build Zv9). It is worth noting, that there is also a putative, less characterized version of *auts2* with an incomplete coding sequence located on zebrafish chromosome 15 (ENSDART00000012712). Knocking down this gene along with the *auts2* gene that was assayed in our study may lead to more severe phenotypes. In order to better understand *AUTS2* function, a conditional knockout mouse should be developed.

My enhancer search focused primarily on the first five introns due to the numerous reports of cognitive-related structural variations in that region (Ben-David et al., 2011; Girirajan et al., 2011; Glessner et al., 2009; Huang et al., 2010; Kalscheuer et al., 2007; Mefford et al., 2010; Nagamani et al., 2013; Pinto et al., 2010; Schumann et al., 2011; Sultana et al., 2002; Talkowski et al., 2012), along with the region's putative role in evolution. There could be numerous functional enhancers outside this region that we have not tested in this study. For example, there is an intragenic SNP (rs6961611) associated with processing speed (Luciano, Hansell, Lahti, Davies, & Medland, 2012) 1.6 mega bases downstream of *AUTS2* which could be associated with a regulatory element for this gene. While the expression of my enhancers largely recapitulated *Auts2* expression, it is possible that the enhancers we identified could regulate a neighboring gene. Future experiments such as chromatin interaction analyses (Fullwood et al., 2010; Lieberman-Aiden et al., 2009) could be able to distinguish what promoters our enhancers are interacting with.

Previous work has shown that human enhancer sequences can function as active enhancers in zebrafish, even without homologous sequences in zebrafish (Fisher, Grice, Vinton, Bessling, & McCallion, 2006; Mcgaughey et al., 2008; Navratilova et al., 2009). My results

confirm these findings for some of our enhancers. For example, AEC10, 13 and 29, which do not have homologous sequences in zebrafish, have similar enhancer expression patterns in zebrafish and mouse (Table 1). However, AEC21 and 27, which are conserved down to zebrafish, and AEC 24, which is conserved down to chicken, don't have matching expression patterns in zebrafish and mice.

I found three positive human enhancers in both zebrafish and mouse that reside within a 33,519bp deletion detected in an individual with ASD, one of which, AEC32, is expressed in the cerebellum. This deletion was inherited from the individual's mother who was not diagnosed with ASD (Pinto et al., 2010). ASDs are likely caused by multiple genomic aberrations in combination with environmental factors. While it is possible that in this individual, this deletion leads to ASD due to the loss of these enhancers and potentially other functional sequences, it is also possible that the loss of these enhancers is one of multiple "hits" (Poot, Smagt, Brilstra, Bourgeron, & van der Smagt, 2011) or that the deletion is not causative. With the constantly growing number of individuals with ASDs or other neurological phenotypes that have *AUTS2* mutations, some of which are purely noncoding, it is likely that improper regulation of this gene is involved in the progression of these disorders.

I also characterized enhancers in locations associated with other neurological phenotypes. In an 84kb deletion in intron one of an individual with dyslexia, I identified four positive human enhancers in zebrafish (AEC3-6) (Table 1 and Figure 5), one of which is expressed in the midbrain. In addition, one of the candidates that was negative for zebrafish enhancer activity (AEC35) was a sequence that included the alcohol consumption associated SNP (rs6943555) (Schumann et al., 2011). It is possible that zebrafish is not a good model system for this region/phenotype or that the actual functional region/variant is further away from this tag SNP.

By characterizing the regulatory landscape of this region we have obtained a better understanding of the functional units within this gene, which now pose as candidates for mutation analysis in individuals with various neurological phenotypes.

AUTS2 has been singled out as a gene that is rapidly evolving in humans in three different studies (Green et al., 2010; Pollard, Salama, King, et al., 2006; Prabhakar et al., 2006). Using zebrafish enhancer assays, I identified sixteen different enhancers that lie within regions that were implicated in human evolution, six of which show expression in the brain. I tested four of the enhancers in mice and two of them had midbrain enhancer activity. My enhancer results shed light on the tight regulation of this gene; however, other experiments should determine the targets of the AUTS2 protein, which is described in Chapter 3.

2.5 Materials and Methods

2.5a Whole-mount *in situ* hybridization

Zebrafish embryos were collected from ABs or caspers (White et al., 2008) between 24 to 120 hpf and fixed in 4% paraformaldehyde buffered with 1X PBS (PFA). The zebrafish *auts2* (Open Biosystems EDR1052-4681254) cDNA clone was used to generate digoxygenin labeled probes. Whole-mount *in situ* hybridizations were performed according to standard protocols (Thisse et al., 2004).

2.5b Morpholino assays

Two morpholino (MO) antisense oligonucleotides targeting *auts2* were designed by Gene-Tools. One MO was designed to target the translational start site of *auts2* (GTGGAGAGTGTGTCAACACTAAAAT). The second was designed to target the splice junction between intron 2 and exon 3 of Ensembl Transcript ENSDART00000137928

(TCGACTACTGCTGTGAACAAAGAGA). A third 5 bp mismatch control for the translational MO (GTGGACACTGTGTGAAGACAAAAAT) was also designed. The MOs were diluted to 1mM in deionized water and injected using standard techniques (Nasevicius & Ekker, 2000) into one cell-stage embryos. To rescue the morphant phenotypes, I transcribed full length human *AUTS2* RNA (Open Biosystems MHS1010-9204165) using the T7 message machine (Ambion) and co-injected it along with the translational MO at a concentration of 168ng/ul. The HuC line was generously donated by Dr. Su Guo (UCSF). The Tg(mnx1:GFP) (AB) line (formerly known as hb9) was obtained from the Zebrafish International Resource Center (ZIRC; http://zebrafish.org/zirc/home/guide.php). Fish were injected with MOs as described above and annotated using the Leica M165 FC microscope. At least 50 translational MO injected fish and controls were compared in all zebrafish lines used.

2.5c Immunohistochemistry on zebrafish sections

AB zebrafish embryos injected with the *auts2* translational MO or the 5bp control were fixed at 48 hpf in 4% PFA overnight at 4°C, then washed for 15 minutes at room temperature in PBS. Zebrafish were frozen into blocks using Tissue-Tek O.C.T. (Sakura Finetek) then sectioned (10-20 microns) using a Leica CM1850 cryostat and stained with Nissl (FD NeuroTechnologies). Morphant and control sections represent comparable planes. Staining with PCNA (DAKO, Monoclonal Mouse PCNA clone PC10) was done according to the manufacturer's protocol. Cell nuclei were visualized using DAPI (Invitrogen). Staining sections with TUNEL (Roche, *In Situ* Cell Death Detection Kit, TMR red) was done according to the manufacturer's protocol. Zebrafish sections were analyzed using the Leica M165 FC or the Nikon Eclipse E800 microscope. At least 25 fish were analyzed in each condition. Control and morphant pictures were taken with identical exposures and are representative of each condition.

For TUNEL staining on sections, criteria for amount of cell death was based on the number of individual TUNEL positive cells identified in the midbrain and eye, indicative of cell death in those regions. For PCNA staining (cell cycle marker) on sections, criteria for amount of proliferation in the forebrain, midbrain and hindbrain was qualitatively evaluated due to the larger number of PCNA positive cells in morphants compared to controls.

2.5d Zebrafish whole-mount immunohistochemistry

Casper zebrafish embryos injected with the *auts2* translational MO or the 5bp control were fixed at 48 hpf overnight at 4°C in 4% PFA. For TUNEL staining, embryos were transferred to methanol for 30 minutes followed by rehydration in methanol/PBST (PBS with 0.1% tween). They were then placed in Proteinase K (10µg/ml) for 5 minutes and postfixed in 4% PFA for 20 minutes. Embryos were later placed in prechilled ethanol:acetic acid (2:1) at -20°C for 10 minutes and then washed in PBST for 20 minutes followed by TUNEL staining using the *In Situ* Cell Death Detection Kit, TMR red (Roche) according to the manufacturer's protocol. Sensory neurons were analyzed using anti-HNK-1 (Sigma) followed by the goat anti-mouse IgM HRP secondary antibody (abcam, ab5930) using previously described methods (Holder & Hill, 1991). HNK-1 positive cells where manually counted in 6 different control and morphant fish. Fish were analyzed using the Leica M165 FC or the Nikon Eclipse E800 microscope. At least 25 fish were analyzed in each condition. Control and morphant pictures were taken with identical exposures and are representative of each condition. For TUNEL whole mount staining, criteria for amount of cell death was based on the number of viewable individual TUNEL positive cells in the forebrain, midbrain and hindbrain. For HNK-1 staining, criteria for amount of sensory neurons was based on the number of individual HNK-1 positive cells counted in equal lengths of the trunk. Motor neuron axons were analyzed using anti-znp-1

(Developmental Studies Hybridoma Bank) followed by anti-mouse IgG HRP (GE Healthcare) using previously described methods (Westerfield, 2007).

2.5e Transgenic enhancer assays

AUTS2 enhancer candidate (AEC) sequences were selected based on evolutionary conservation (sequences showing ≥70% identity for at least 100bp between human and chicken), E1A binding protein p300 (EP300) forebrain or hindbrain ChIP-Seq datasets (Visel et al., 2009), and nucleotide variants that define the human-Neanderthal sweep (Green et al., 2010) (Table S1). PCR was carried out on human genomic DNA (Qiagen) using primers designed to amplify the AEC sequences (Table 1). Primers were designed such that they will have additional flanking sequences to the conserved, ChIP-Seq or human-Neanderthal accelerated sequences based on previous experiments that have shown this to be a reliable method for obtaining positive enhancer activity (Pennacchio et al., 2006). PCR products were cloned into the E1b-GFP-Tol2 enhancer assay vector containing an E1b minimal promoter followed by GFP (Q. Li et al., 2010). They were then injected following standard procedures (Nusslein-Volhard, 2002; Westerfield, 2007) into at least 100 embryos per construct along with Tol2 mRNA (Kawakami, 2005), to facilitate genomic integration. GFP expression was observed and annotated up to 48 hpf. An enhancer was considered positive if at least 15% of all fish surviving to 48 hpf showed a consistent expression pattern after subtracting out percentages of tissue expression in fish injected with the empty enhancer vector. Notably, the empty vector showed particularly high background for heart and somitic muscle and as described all enhancer results were obtained after deducting its expression pattern. Thus, in order to call positive somitic muscle enhancer activity, over 26% (24hpf) or 40% (48hpf) of alive fish needed to show positive enhancer activity. To call a positive heart enhancer, 32% (24hpf) or 50% (48hpf) of alive fish needed show

positive heart activity. For each construct, at least 50 fish were analyzed for GFP expression at 48 hpf. For the mouse enhancer assays, the same human genomic fragment used in zebrafish was transferred into a vector containing the *Hsp68* minimal promoter followed by a *LacZ* reporter gene (Kothary R., Clapoff S., Brown A., Campbell R., 1988; Pennacchio et al., 2006) and sequence verified to ensure the insert matched the human reference sequence. Sequences having rare variants were changed to the reference human genomic sequence by site-directed mutagenesis (Mutagenex or Quickchange II, Stratagene) and sequence verified for having the reference sequence. Transgenic mice were generated by Cyagen Biosciences using standard procedures (Nagy A., Gertsenstein M. , Vintersten K., 2003). Embryos were harvested at E12.5 and stained for LacZ expression using standard procedures (Pennacchio et al., 2006). Mouse embryos selected for sectioning were placed in an overnight cryoprotection stage using 30% sucrose in PBS. Mice were frozen into blocks using Tissue-Tek O.C.T. (Sakura Finetek) then sectioned (20 microns) using a Leica CM1850 cryostat and stained with Nuclear Fast Red Solution (Sigma-Aldrich) for one minute. There is no human subjects work involved in this article. All animal work was approved by the UCSF Institutional Animal Care and Use Committee (protocol number AN084690).

2.6 Acknowledgments

I would like to thank Lauren A. Weiss, Ophir D. Klein and members of the Ahituv lab for helpful comments on this manuscript. I would like to thank Yien-Ming Kuo and Michael Berberoglu for their support with histological sections and staining. I would also like to thank Shoa L. Clarke and Gill Bejerano (Stanford) for computational assistance and Len A. Pennacchio

and Axel Visel (LBL) for mouse transgenic enhancer embryos. I would also like to thank Erik Sistermans for providing us with figure 8C.

Chapter 2 tables

Table 1: *AUTS2* enhancer candidates (AECs) selected for enhancer assays

AEC #	Chromosomal position (hg18)	Zebrafish enhancer?	Tissue **	Mouse Enhancer? ◊	Tissue ◊◊
1	chr7:68,724,902-68,728,333	No	Negative	Not Tested	
2	chr7:68,812,867-68,814,245	No	Negative	Not Tested	
3	chr7:68,821,855-68,823,820	Yes	Dots around caudal end of tail in somite area (18%, 24hpf)	Not Tested	
4	chr7:68,825,835-68,827,626	Yes	Notochord (31%, 24hpf; 30%, 48hpf) Otic vesicle (16%, 24hpf; 28%, 48hpf)	Not Tested	
5	chr7:68,841,424-68,843,440	Yes	Midbrain (39%, 24hpf; 27%, 48hpf) Somitic Muscle plus dots at end of tail (60%, 24hpf) Otic vesicle (15%, 48hpf) Somitic Muscle (16%, 48hpf)	Not Tested	
6	chr7:68,875,482-68,877,048	Yes	Pericardium (29%, 24hpf; 34%, 48hpf) Eye (28%, 48hpf)	Not Tested	
7	chr7:68,908,734-68,909,828	No	Negative	Not Tested	

8	chr7:68,922,864-68,924,172	No	Negative	Not Tested	
9	chr7:68,928,408-68,929,934	No	Negative	Not Tested	
10	chr7:68,963,804-68,965,620	Yes	Midbrain (60%, 24hpf; 58%, 48hpf), Dots all over mostly above notochord (55%, 24hpf; 39%, 48hpf) Spinal cord (18%, 48hpf)	Yes	Midbrain (12/13, E12.5) Eye (6/13, E12.5) Proximal limb (6/13, E12.5)
11	chr7:68,966,890-68,968,624	No	Negative	Not Tested	
12	chr7:68,996,372-68,998,072	Yes	Epidermis (59%, 24hpf; 50%, 48hpf) Forebrain (39% 24hpf; 34%, 48hpf) Heart (15%, 24hpf) Otic Vesicle (24%, 24hpf; 34%, 48hpf)	Not Tested	
13	chr7:69,036,983-69,038,688	Yes	Otic Vesicle (19%, 48hpf)	Yes (hs 1660)	Ear (5/8, E11.5)
14	chr7:69,108,779-69,109,947	Yes	Somitic Muscle (18%, 24hpf)	Not Tested	
15	chr7:69,166,208-69,168,337	Yes	Eye (19%, 48hpf)	Not Tested	
16	chr7:69,183,577-69,186,403	Yes	Ventral Caudal region (25%, 24hpf; 26%, 48hpf)	Not Tested	

17	chr7:69,195,062-69,196,938	No	Negative	Not Tested	
18	chr7:69,200,457-69,202,216	Yes	Spots on top or end of tail (18%, 24hpf)	Not Tested	
19	chr7:69,206,936-69,208,600	Yes	Heart (42%, 24hpf; 33%, 48hpf) Hindbrain (23%, 24hpf) Somitic Muscle (38%, 48hpf)	Not Tested	
20	chr7:69,217,233-69,219,228	Yes	Midbrain (22%, 24hpf)	Not Tested	
21	chr7:69,234,838-69,236,265	Yes	Spinal Cord (15%, 48hpf)	Yes (hs 1425)	Midbrain (4/6, E11.5)
22	chr7:69,236,069-69,238,171	No	Negative	Not Tested	
23	chr7:69,274,050-69,276,432	Yes	Somitic Muscle (50%, 24hpf; 47%, 48hpf)	Not Tested	
24	chr7:69,284,310-69,286,965	Yes	Hindbrain (17%, 24hpf) Spinal cord (66%, 24hpf; 31%, 48hpf)	Yes	Eye (4/5, E11.5)
25	chr7:69,312,912-69,316,619	Yes	Blood islands-yolk (18%, 24hpf)	Not Tested	
26	chr7:69,316,920-69,317,888	Yes	Forebrain (45%, 24hpf; 40%, 48hpf) Otic Vesicle (23%, 48hpf)	Not Tested	
27	chr7:69,340,167-69,341,551	Yes	Somitic Muscle (30%, 48hpf)	Yes (hs 658)	Midbrain (11/15, E11.5) Neural tube (9/15,

					E11.5)
28	chr7:69,343,769-69,346,450	No	Negative	Not Tested	
29	chr7:69,358,340-69,359,796	Yes	Forebrain (olfactory organ) (15%, 48hpf) Heart (23%, 48hpf)	Yes	Olfactory epithelium (3/6, E12.5) Eye (6/6, E12.5)
30	chr7:69,359,964-69,362,937	No	Negative	Not Tested	
31	chr7:69,365,749-69,367,561	No	Negative	Not Tested	
32	chr7:69,369,740-69,372,052	Yes	Midbrain (19%, 48hpf) Hindbrain (18%, 48hpf)	Yes	Midbrain (4/4, E12.5) Forebrain (4/4, E12.5) Hindbrain (4/4, E12.5) Eye (4/4, E12.5)
33	chr7:69,374,313-69,375,754	No	Negative	Not Tested	
34	chr7:69,439,813-69,441,520	No	Negative	Not Tested	
35	chr7:69,443,563-69,444,772	No	Negative	Not Tested	
36	chr7:69,456,365-69,457,073	Yes	Hindbrain (18%, 24hpf)	Not Tested	
37	chr7:69,458,163-69,460,029	No	Negative	Not Tested	

38	chr7:69,479,431-69,480,094	Yes	Epidermis, only on head (24%, 24hpf)	Not Tested	
39	chr7:69,534,927-69,536,470	No	Negative	Not Tested	
40	chr7:69,654,802-69,656,119	No	Negative	Not Tested	

** Zebrafish tissue data: the tissues specific expression of sequences positive for enhancer activity are mentioned along with the percent positive of total alive fish after subtracting out the empty vector followed by the hours post fertilization (hpf) of expression.

◊ Mouse enhancer data for (hs 1425), (hs 658) and (hs 1660) were acquired from the VISTA enhancer browser: http://enhancer.lbl.gov/

◊◊ Mouse tissue data: the tissue specific expression of the sequences positive for enhancer activity out of the total LacZ stained mice along with the embryonic day (E) that they were assayed at.

A more complete version of this table available using reference: (Oksenberg et al., 2013)

Table 2: *auts2* morpholino knock-down phenotypes.

Assay following morpholino injection	**Developmental phenotype**	**Reference**
Whole mount	Overall stunted development, including smaller head and eyes. Less movement when prodded. Microcephaly with no overall developmental delay.	(Oksenberg et al., 2013) (Beunders et al., 2013)
Alcian blue staining	Micrognathia (undersized jaw) and retrognathia (receded jaw).	(Beunders et al., 2013)
HuC-GFP zebrafish line	Fewer developing neurons in the dorsal region of the midbrain, including the optic tectum, the midbrain-hindbrain boundary (including the cerebellum), the hindbrain and the retina (Oksenberg et al., 2013).	(Oksenberg et al., 2013)
HuC/D staining	Reduction in HuC/D-positive postmitotic neurons as well as a loss of bilateral symmetry.	(Beunders et al., 2013)
TUNEL staining	Increased apoptosis in the midbrain.	(Oksenberg et al., 2013)
PCNA staining	Increased cell proliferation in the forebrain, midbrain and hindbrain.	(Oksenberg et al., 2013)
Phosphohistone H3	Decreased cell proliferation in the brain.	(Beunders et al., 2013)
Tg(mnx1:GFP) zebrafsih line	Fewer motor neuron cell bodies in the spinal cord and weaker, improperly angled projections.	(Oksenberg et al., 2013)
HNK-1 staining	Fewer sensory neurons in the spinal cord.	(Oksenberg et al., 2013)

Chapter 2 figures

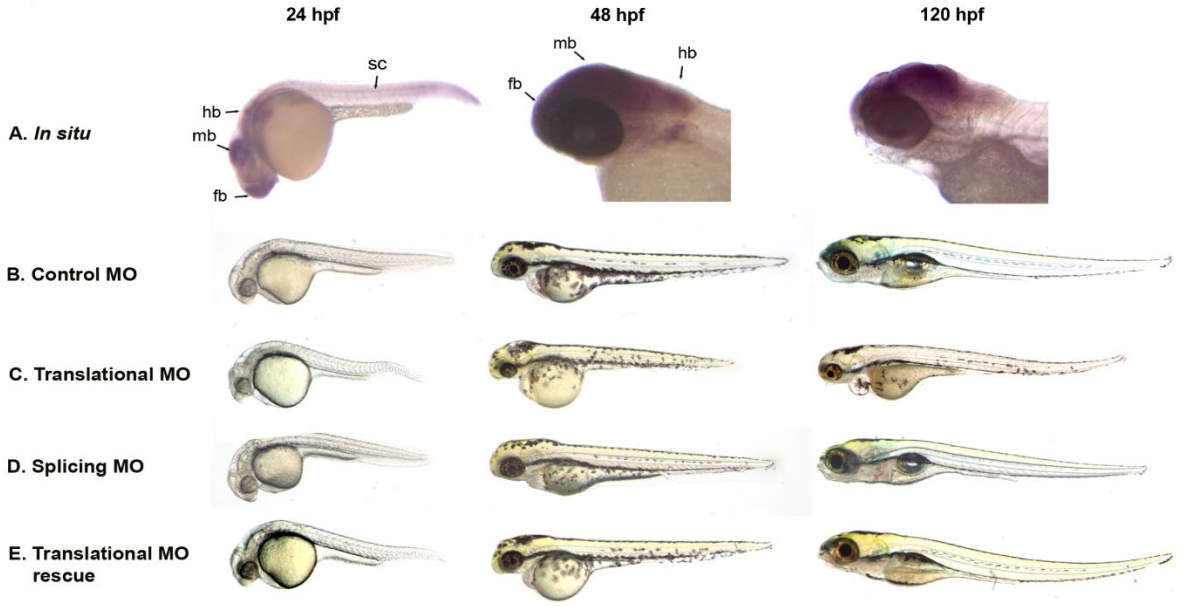


Figure 1: *auts2* expression and morphant phenotype

(A) Whole-mount *in situ* hybridization of *auts2* shows that it is expressed in the forebrain (fb) (including olfactory organs), midbrain (mb), hindbrain (hb), spinal cord (sc), the caudal peduncle and eye at 24hpf. At 48hpf, *auts2* is expressed in the brain, pectoral fin and eye. At 120 hpf expression is restricted to the brain, primarily the midbrain, and weakly in the eye. (B) Fish injected with the 5 bp translational MO mismatch control have indistinguishable morphology as wild type fish at 24, 48 and 120 hpf. (C) Injection of the *auts2* translational MO results in fish with a stunted development phenotype that includes smaller heads, eyes, bodies and fins. (E) Injection of the *auts2* splice-blocking MO shows a similar but less severe phenotype than the *auts2* translational MO. (E) The *auts2* translational MO phenotype is partially rescued by co-injecting the full length human *AUTS2*. Note the longer body and larger brain compared to the translational and splicing morphant fish. MO injected fish in C, D, and E are scaled to the 5bp injected control fish in B.

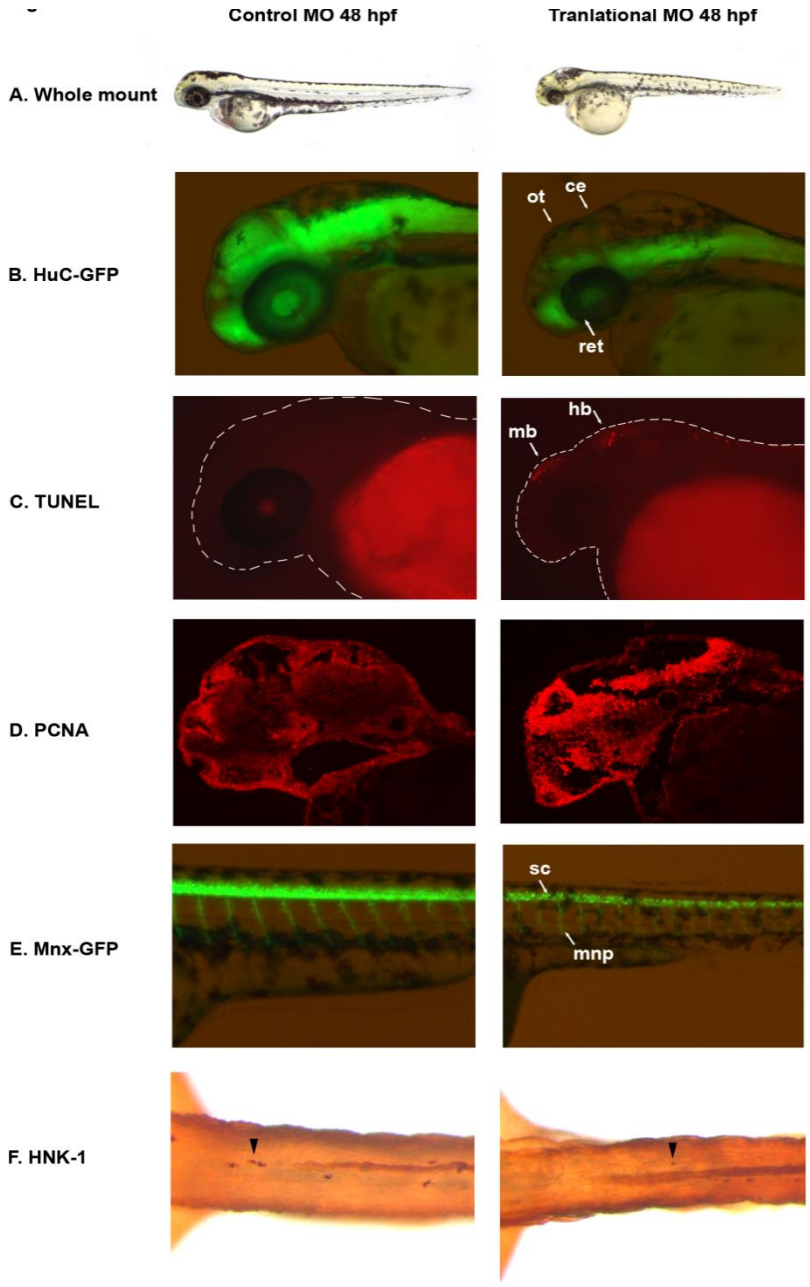


Figure 2: *auts2* 48 hpf morphant phenotype.

(A) Fish injected with the 5 base-pair translational MO mismatch control have similar morphology as wild type fish. Injection of the *auts2* translational MO results in fish with a stunted development phenotype that includes a smaller head, eyes, body and fins. (B) HuC-GFP

fish injected with the 5bp control MO display normal levels of developing neurons in the brain. HuC-GFP translational MO injected fish display considerably less developing neurons in the optic tectum (ot), retina (ret), and cerebellum (ce). (C) 5bp mismatch control injected fish have little to non-observable apoptosis in the brain as observed by TUNEL staining, while translational MO injected fish display high levels of apoptosis, primarily in the midbrain (mb) and hindbrain (hb). (D) PCNA cell proliferation assay in the 5bp MO control injected fish shows lower levels of cell proliferation in the brain compared to the translational MO injected fish. (E) Tg(mnx1:GFP) fish injected with the 5bp MO control display normal levels of motor neurons versus the *auts2* translational MO injected fish which have fewer motor neurons in the spinal cord (sc). In addition, motor neuron projections (mnp) are weaker and more perpendicular to the spinal cord. (F) Translational MO injected fish display fewer Rohon-Beard cells (arrowheads) in the spinal cord than morphants. All morphant fish are scaled to their 5bp control counterparts.

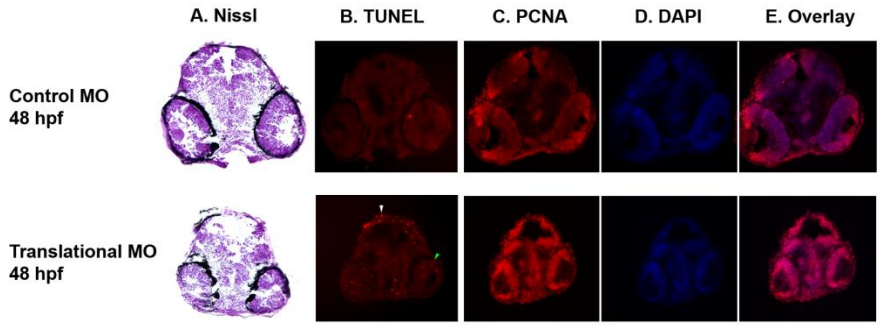


Figure 3: Histological phenotype of *auts2* morphants

(A) Nissl staining shows a reduction in neuron territory, primarily in the midbrain, of fish injected with the translational MO compared to 5bp mismatch controls at 48 hpf. (B) TUNEL stained sections show fewer apoptotic cells in the optic tectum (white arrowhead) and the retina (green arrowhead) in 48 hpf *auts2* morphants versus the 5bp translational MO mismatch control. (C-E) Coronal sections stained with PCNA, DAPI and overlays show an increase in cell proliferation in the translational morphant fish compared to the 5bp mismatch control in the mesencephalon, diencephalon and retina.

Figure 4: znp-1 antibody on control and morphant fish

The motor neurons axons of the morphant fish are different than the controls, signified by a drastic increase in the amount of branching (red arrow).

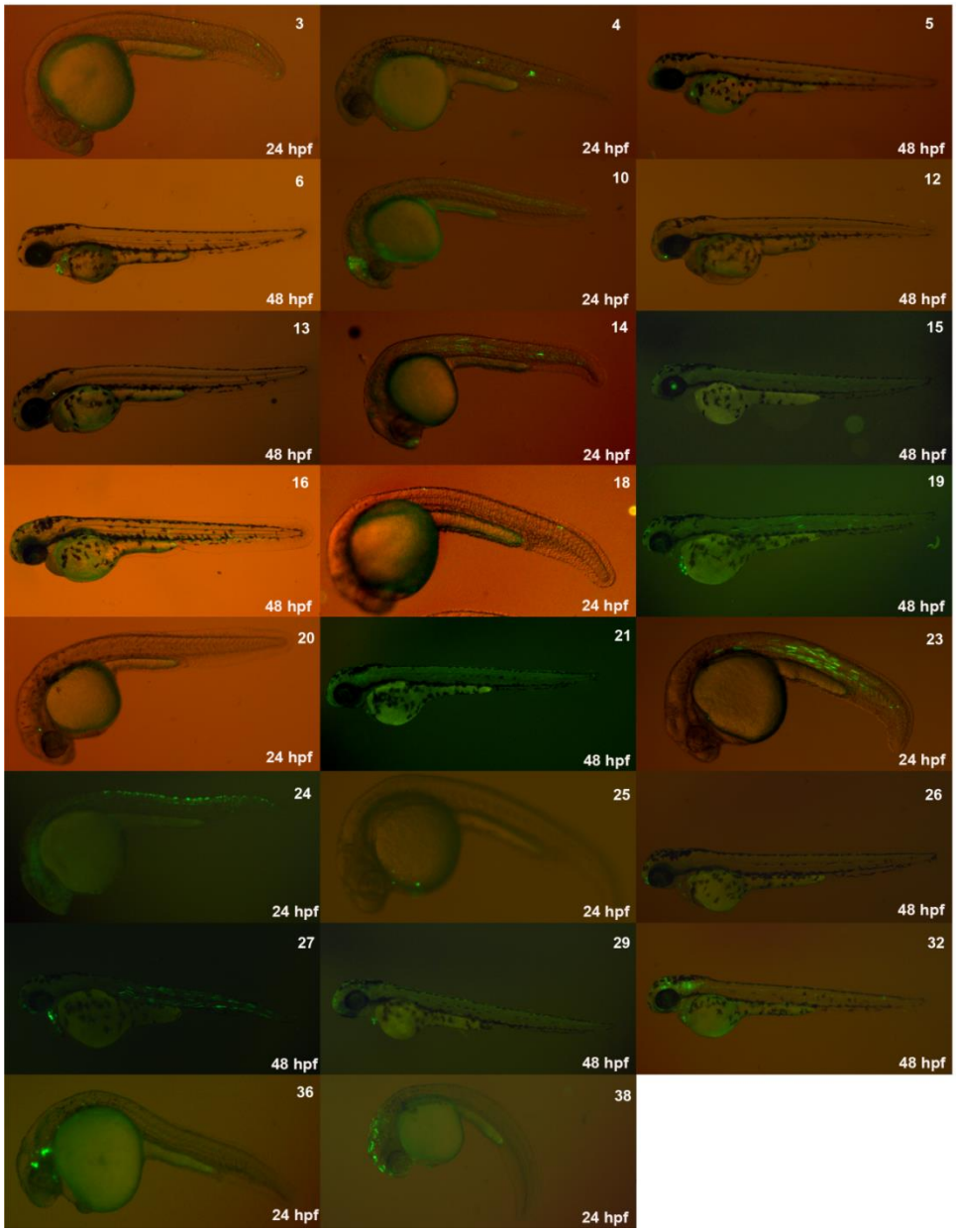


Figure 5: *AUTS2* enhancer candidates (AECs) positive for enhancer activity in zebrafish

A representative fish of each positive AEC enhancer is shown. The number in the top right of every image is the AEC number and the hours post fertilization (hpf) when the picture was taken is indicated in the bottom right. Their tissue-specific expression pattern is denoted in Table S1 and http://zen.ucsf.edu.

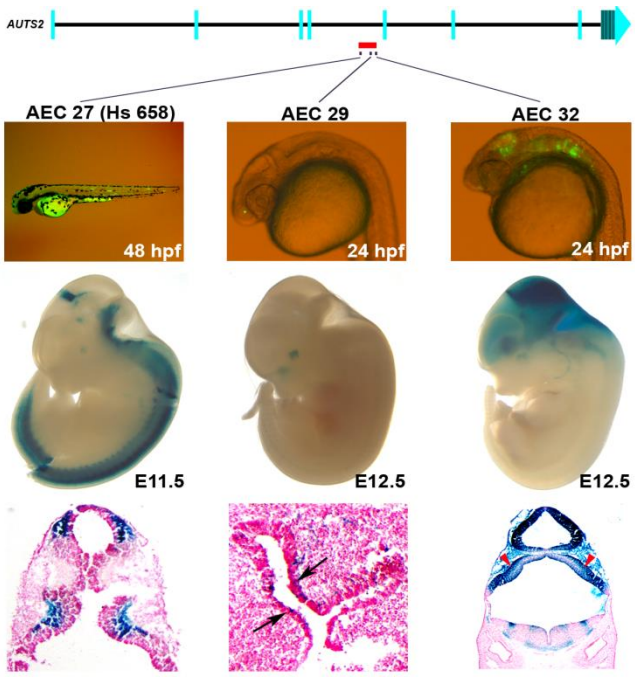


Figure 6: Enhancers within an ASD-associated *AUTS2* intronic deletion.

Three positive enhancers (AEC27, 29, 32) show positive enhancer activity in zebrafish (24 or 48 hpf) and in mice (E11.5 or 12.5). AEC27 shows enhancer expression in the somitic muscle in zebrafish, while in mouse at E11.5 (hs658; (Visel et al., 2007)) it is active in the midbrain, medulla, and neural tube at E11.5. The histological section below shows its enhancer activity in the pretectum and the pons. At E12.5, AEC29 shows enhancer activity in the olfactory epithelium (arrows in histological section) similar to zebrafish and in addition also displays enhancer expression in the eye. AEC32 recapitulates the zebrafish enhancer expression displaying strong enhancer activity in the midbrain (tectum) and hindbrain and in addition also displays enhancer expression in the forebrain at E12.5. Histological sections of AEC32 show enhancer activity in the mouse cerebellum (red arrowheads).

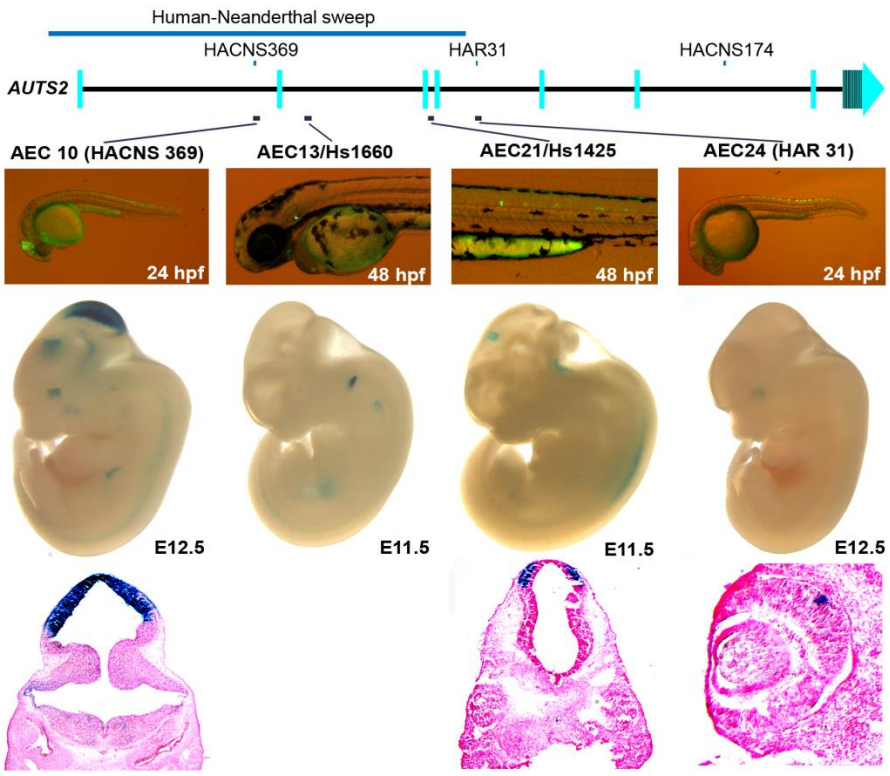


Figure 7: Four positive zebrafish and mouse enhancers in regions implicated in human evolution

At E12.5, AEC10 shows zebrafish and mouse enhancer expression in the midbrain and eye. The histological section below highlights its expression in the tectum. AEC13, is expressed in the otic vesicle both in zebrafish and E11.5 mouse embryos (hs1660 ;(Visel et al., 2007)). AEC21 is expressed in the spinal cord in zebrafish, while in the mouse it showed midbrain expression at E11.5 (hs1425; (Visel et al., 2007)). Histological sections below show its expression in the pretectum of the midbrain. AEC24 was expressed in the spinal cord and hindbrain in zebrafish and in the eye in mouse at E12.5.

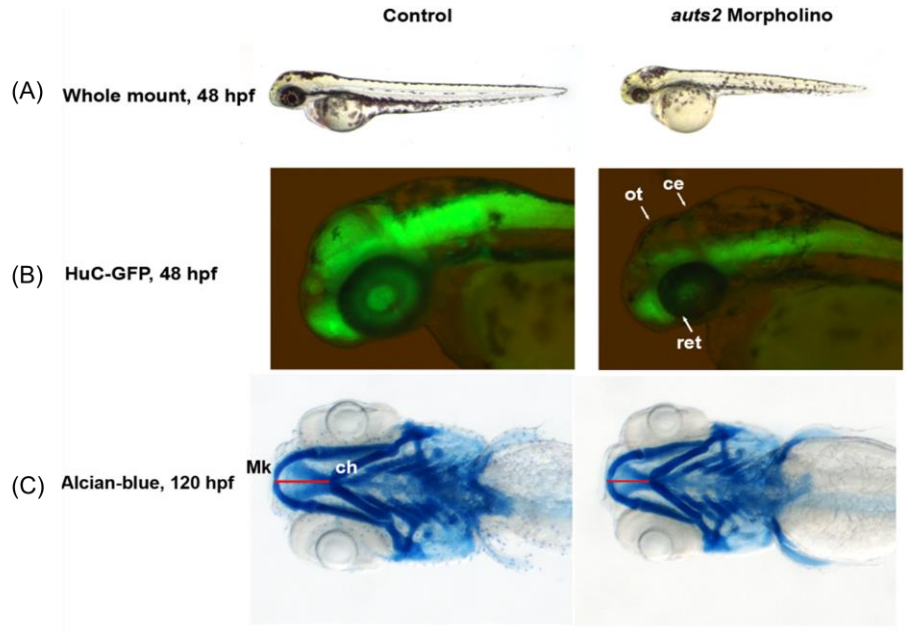


Figure 8: *auts2* zebrafish knockdown phenotype

(**A**) At 48 hours post fertilization (hpf), fish injected with a 5 base-pair *auts2* morpholino (MO) mismatch control have similar morphology as wild type fish, whereas fish injected with a corresponding translational MO display a stunted developmental phenotype that includes a smaller head, eyes, body and fins. (**B**) At 48 hpf, HuC-GFP fish injected with a 5 base-pair *auts2* mismatch control MO display normal levels of developing neurons in the brain, while translational MO injected fish display less developing neurons in the cerebellum (ce), optic tectum (ot) and retina (ret). (**C**) At 120 hpf, fish injected with an *auts2* splicing MO and stained with alcian-blue show a significant reduction in the distance between the Meckel (Mk) and ceratohyal cartilages (ch) (shown as a red line) compared to controls, indicating a reduced lower-jaw size. (**A**, **B**) adapted from (Oksenberg et al., 2013), (**C**) adapted from (Beunders et al., 2013).

Chapter 3 – Genome-wide distribution of Auts2 binding localizes with active neurodevelopmental genes

3.1 Abstract

The autism susceptibility candidate 2 gene (*AUTS2*) has been associated with multiple neurological diseases including autism spectrum disorders (ASD). Previous studies showed that *AUTS2* has an important neurodevelopmental function and is a suspected master regulator of genes implicated in ASD-related pathways. However, the regulatory role and targets of Auts2 are not well known. Here, by using ChIP-seq and RNA-seq on mouse embryonic day 16.5 forebrains, I elucidated the gene regulatory networks of Auts2. I find that the majority of promoters bound by Auts2 belong to genes highly expressed in the developing forebrain, suggesting that Auts2 is involved in transcriptional activation. Auts2 non-promoter bound regions significantly overlap developing brain-associated enhancer marks and are located near genes involved in neurodevelopment. Auts2 marked sequences are enriched for binding site motifs of neurodevelopmental transcription factors, including Pitx3 and TCF3. Additionally, I characterized two functional brain enhancers marked by Auts2 near *NRXN1* and *ATP2B2*, both ASD implicated genes. My results implicate Auts2 as an active regulator of important neurodevelopmental genes and pathways and identify novel genomic regions which could be associated with ASD and other neurodevelopmental diseases.

3.2 Introduction

AUTS2 has been increasingly implicated as an ASD candidate gene with over 50 unrelated individuals with ASD or ASD-related phenotypes, such as intellectual disability or

developmental delay, having structural variants disrupting the *AUTS2* region (Bakkaloglu et al., 2008; Ben-David et al., 2011; Beunders et al., 2013; Cuscó et al., 2009; Girirajan et al., 2011, 2013; Glessner et al., 2009; Huang et al., 2010; Jolley et al., 2013; Kalscheuer et al., 2007; Nagamani et al., 2013; Pinto et al., 2010; Talkowski et al., 2012; Tropeano et al., 2013). *AUTS2* structural variants have also been associated with other neurological phenotypes including epilepsy (Mefford et al., 2010), schizoaffective disorder (Hamshere et al., 2009), bipolar disorder (Hattori et al., 2009; Lee et al., 2012), ADHD (Elia et al., 2010), differential processing speed (Luciano et al., 2011), suicidal tendencies under the influence of alcohol (Chojnicka et al., 2013), and dyslexia (Girirajan et al., 2011). Additionally, a non-coding single-nucleotide polymorphism within *AUTS2* (rs6943555) is significantly associated with alcohol consumption (Schumann et al., 2011). *AUTS2* has also been implicated in human specific evolution (Green et al., 2010; Pollard, Salama, Lambert, et al., 2006; Prabhakar et al., 2006).

The precise function of *AUTS2* is not well known; however, zebrafish knock-downs as I've described in chapter 2 have shown *auts2* to be critical in neurodevelopment. *auts2* morphant fish display microcephaly with a decrease in neuronal cells in the brain (Beunders et al., 2013; Oksenberg et al., 2013), which may be caused by failure of cells to differentiate into mature neurons (Oksenberg et al., 2013). *auts2* knock-down also leads to craniofacial abnormalities in zebrafish (Beunders et al., 2013) and reduced movement, possibly caused by fewer motor and/or sensory neurons (Oksenberg et al., 2013). Sequence analysis of AUTS2 identified a predicted PY motif (PPPY) at amino acids 515-519 (Sultana et al., 2002), a potential WW-domain-binding region involved in protein-protein interactions. This motif is thought to be involved in the activation of transcription factors, suggesting that AUTS2 may be involved in transcriptional regulation (Kalscheuer et al., 2007).

As described in the introduction in chapter 1, several proteins are suggested to affect the expression of AUTS2. T-box, brain, 1 (Tbr1), a transcription factor that has been implicated in ASD (O'Roak et al., 2012), regulates regional and laminar identity in postmitotic neurons (Bedogni, Hodge, Nelson, et al., 2010) and is critical for proper neocortex development (Bedogni, Hodge, Elsen, et al., 2010), interacts with the *Auts2* promoter in the developing neocortex (Bedogni, Hodge, Elsen, et al., 2010). *Tbr1* deficient mice have decreased levels of reelin (Bedogni, Hodge, Elsen, et al., 2010), a protein necessary for proper neuronal migration in developing brains that can be expressed at decreased levels in individuals with ASD (Fatemi et al., 2005). *Auts2* expression is reduced in SATB homeobox 2 (*Satb2*) null mice, with *Satb2* being a known regulator of Tbr1 expression in callosal projection neurons (Srinivasan et al., 2012). *Auts2* has been reported to have a 1.33 fold change in cerebellar gene expression in methyl CpG binding protein 2 (*Mecp2*) null mice, a gene implicated in neurodevelopmental disorders including Rett syndrome and autism (Ben-Shachar et al., 2009). *AUTS2* is a potential target of GTF2I repeat domain-containing 1 (*GTF2IRD1*), one of 26 genes deleted in neurodevelopmental disorder Williams-Beuren syndrome (O'Leary & Osborne, 2011; Young et al., 2008). Zinc finger matrin-type 3 (*Zmat3*, also known as *wig-1*) is a transcription factor regulated by p53 and plays an important role in RNA protection and stabilization. wig-1 down regulation leads to a significant reduction in *Auts2* mRNA levels in the brains of BACHD mice, a Huntington's disease mouse model (Sedaghat et al., 2012).

AUTS2 is also thought to interact with other genes and pathways including the notch and ERK signaling pathways, *PRC1,* and *SEMA5A* (Molnar & de Celis, 2013; Oksenberg & Ahituv, 2013). Notch signaling has been shown to be involved in neuronal migration through its interaction with *Reelin* (Hashimoto-Torii et al., 2008; Wang et al., 2004) and *AUTS2* expression

was found to oscillate in phase with other notch pathway genes (Dequéant et al., 2006; William et al., 2007). Polycomb-group repressive complex 1 (*PRC1*), a polycomb-group gene often involved in transcriptional repression physically interacts with AUTS2, implicating a role for AUTS2 in developmental transcriptional regulation (Gao et al., 2012). The regulatory pathway for autism candidate gene semaphorin 5A (*SEMA5A*) contains rare copy number variations around ASD-associated genes, including *AUTS2.* Twelve regulators of *SEMA5A*-regulated genes were identified, including *AUTS2*, suggesting that *AUTS2* is a master regulator in ASD-related pathways (Cheng et al., 2013). The *drosophila melanogaster* Tay bridge gene, which has homology with AUTS2 in the C-terminal region, is a component of the EGFR/Erk signaling pathway that antagonizes EGFR signaling and interacts with Erk. This suggests that AUTS2's normal function in humans is involved in Erk signaling, playing a role in the differentiation and survival of cells during development (Molnar & de Celis, 2013). Combined, these interactions implicate *AUTS2* in neurodevelopment pathways, including processes important for cell differentiation and ASD (Oksenberg & Ahituv, 2013). However, the actual downstream regulatory targets of *AUTS2* still remain largely unknown.

Given *AUTS2*'s suspected role in neurodevelopment and ASD, and that *AUTS2* may be a master regulator in ASD-related pathways (Cheng et al., 2013), I explored the genomic targets of *Auts2* in E16.5 mouse forebrains using chromatin immunoprecipitation followed by deep sequencing (ChIP-seq), RNA-seq and zebrafish enhancer assays. I found multiple lines of evidence that Auts2 is associated with promoters and distal enhancers of genes that are active during neurodevelopment. Additionally, motif analysis of Auts2 marked sites found enrichment for known motifs involved in neurodevelopment including paired-like homeodomain 3 (Pitx3), transcription factor 3 (TCF3), and forkhead box O3 (FOXO3). Finally, I identified two novel

brain enhancers marked by Auts2, which are located near neurexin 1 (*NRXN1*) and ATPase Ca++ transporting plasma membrane 2 (*ATP2B2*), two genes implicated in ASD.

3.3 Results

3.3a Overview

I performed RNA-seq and ChIP-seq using an Auts2 antibody on E16.5 mouse forebrains. E16.5 was chosen because of the reported strong *Auts2* expression in the forebrain (Bedogni, Hodge, Nelson, et al., 2010) and the established neurogenesis for many relevant brain structures at this time point (Finlay & Darlington, 1995). Through RNA-seq, I identified 8,897 transcripts expressed at this time point (quantified as FPKM>0.3). My Auts2 ChIP-seq found 1,930 marked sites, the majority of which (1,146=59%) do not overlap gene promoters (2,500bp upstream and 500bp downstream of a TSS). Nonetheless, the 784 Auts2-marked promoters I detected are significantly more than expected by chance ($p<0.001$; permutation test), and most promoter peaks (602=31% of all peaks) directly overlap the TSS (Figure 1A).

3.3b Promoters of actively transcribed neurodevelopmental genes are marked by Auts2

I initially focused on the 784 Auts2 marked sites that reside within promoter regions. These promoter peaks correspond to 776 genes, as a few genes have multiple Auts2 marked sites within their promoter region. My RNA-seq analysis showed these genes display significantly higher expression levels at E16.5 than transcripts that were not marked by Auts2 ($p<2.2e^{-16}$; Wilcoxon test) (Figure 1B). Consistent with their association with highly expressed genes, 88% of Auts2 marks at promoters (689/784) are also marked by the active promoter histone modification H3K4me3 in previously published ChIP-seq data generated from E14.5 whole brain

(Shen et al., 2012) ($p<0.001$; permutation test) (Figure 1C). These results indicate that the presence of Auts2 at promoters correlates with transcriptional activation.

Using Ingenuity Pathway Analysis (IPA), I comprehensively analyzed pathways and networks of the 776 genes whose promoters overlap Auts2 marked sites, and therefore may be directly regulated by the Auts2 protein. I found that these genes are significantly enriched for diseases and biological functions related to neurodevelopment, including epileptic seizures, disorders of the basal ganglia, migration of neural precursor cells, axonogenesis, differentiation of neurons, morphology of the cerebral cortex, brain mass, and more (Figure 2A, Table 1). Interestingly, these genes are also enriched for processes involved in gene expression and cell cycle, including expression of RNA, transcription, proliferation of cells, splicing of RNA, expression of DNA, and cell death (Figure 2A).

3.3c Non-promoter Auts2 marked regions function as enhancers

I next analyzed the 1,146 Auts2 ChIP-seq marked sites that did not overlap promoters. Supporting the hypothesis that these distal peaks are functionally important genomic elements, 74% are evolutionarily conserved (phastcons 30-way mammal conserved elements) (Siepel et al., 2005), significantly more than expected by chance ($p<0.001$; permutation test). Further suggesting that Auts2 is primarily an activator, 26% (294/1,146) of Auts2 distal peaks overlap previously reported mouse E14.5 whole brain H3K27ac ChIP-seq peaks (Shen et al., 2012), an active enhancer mark, which is significantly more than expected by chance ($p<0.001$; permutation test). In contrast, only 2% (20/1,146) overlap E14.5 mouse forebrain H3K27me3 peaks (Shen et al., 2012), a repressive mark ($p=0.081$; permutation test) (Figure 1C). Five Auts2 marked sites were identified within the *Auts2* gene itself, all of which overlap mouse E14.5 forebrain H3K4me1 marks (Shen et al., 2012), and two that overlap H3K27ac marks (Shen et al.,

2012), suggesting an auto-regulatory active role for Auts2. Combined, these results imply that many of the non-promoter Auts2 marked sites likely function as activate regulatory elements.

I investigated the regulatory functions of the 1,146 non-promoter Auts2 marked sites using the gene ontology tool GREAT. These Auts2 marked non-promoter sites reside near genes involved in mouse brain development (Figure 2B) including corpus callosum size and neuron number, both of which have been implicated in ASD (Courchesne et al., 2011; Egaas, Courchesne, & Saitoh, 1995; Lainhart & Lange, 2011). Taken together, these data support a neurodevelopmental and gene expression role for genes associated with Auts2-marked regulatory regions.

Given my observed correlation between Auts2 marked sites and enhancer marks, I next tested whether these sequences function as enhancers *in vivo*. Ten Auts2 marked enhancer candidates (AMECs) (Table 2) were tested for enhancer activity using a zebrafish transgenic enhancer assay. Candidates were selected based on proximity to ASD-related genes, conservation, and overlap with enhancer-associated histone modifications (Shen et al., 2012). Four of the ten candidates were positive enhancers at 24 or 48 hours post fertilization (hpf). AMEC1 lies in an intron of *NRXN1* and showed positive enhancer activity in the heart and forebrain (olfactory epithelium) at 48 hpf (Figure 3A, Figure 4A). AMEC2, which lies ~56kb upstream of contactin 4 (*CNTN4*), displayed enhancer activity in the somitic muscles at 48 hpf (Figure 4B). AMEC5, which lies ~571kb upstream of the RNA binding protein fox-1 homolog (*RBFOX1*), had enhancer activity in the notochord at 48 hpf (Figure 4D). AMEC8, which lies in an intron of *ATP2B2*, showed enhancer activity in the midbrain and hindbrain (potentially in the trigeminal sensory neuron) and spinal cord at 24 hpf, and in somitic muscles at 48 hpf (Figure

3B, Figure 4D). These results show that Auts2 occupies functionally active enhancers, some of which drive expression in the developing brain.

3.3d Additional genome wide pathway analyses confirm neurodevelopmental function

I examined the function of nearby genes for all 1,930 Auts2 marked sites and the subset of 784 promoter marked sites using GREAT (Table 1) and observed an enrichment for gene expression in the mouse cerebral cortex (promoter sites) and in the lower jaw (all marked sites) (Table 1), which fits with previous *auts2* morpholino knock-down phenotypes of smaller heads (Beunders et al., 2013; Oksenberg et al., 2013) and undersized and reduced jaws (Beunders et al., 2013). GREAT disease ontology enrichment also identified motor neuron disease and hereditary degenerative disease of central nervous system (promoter marked sites) (Table 1), which also matches the *auts2* knock-down phenotype of fewer motor neuron cell bodies in the spinal cord along with improperly angled weaker projections (Oksenberg et al., 2013).

I also performed IPA analysis on all Auts2 marked sites and the subset of 1,146 non-promoter Auts2 marked sites using the gene whose TSS was closest to the ChIP-seq peak (Table 1). For both datasets, I found significant enrichment of many diseases and functions including multiple neurodevelopmental related categories (e.g., migration of neurons, differentiation of neurons, development of the cerebral cortex, disorders of the basal ganglia, Parkinson's disease, schizophrenia, etc.) (Table 1). Additionally, I observed enrichment for genes involved in the axonal guidance signaling canonical pathway ($p=4.6e^{-9}$; right-tailed Fisher Exact Test; for all Auts2 marked sites) (Figure 5, Table 1) and ERK/MAPK signaling canonical pathway ($p=6.79e^{-7}$; right-tailed Fisher Exact Test; for all Auts2 marked sites) (Figure 6, Table 1).

3.3e Motif analysis identifies transcription factors involved in neuronal development

To identify known motifs present in Auts2 marked sites, I compared the sequences of those sites with position weight matrices of several hundreds of experimentally determined transcription factor binding sites using MEME-ChIP (Machanick & Bailey, 2011). I analyzed gene ontology terms linked to identified motifs with the gene ontology tool GOMO (Buske, Bodén, Bauer, & Bailey, 2010). In Auts2 marked promoter regions, we identified motifs involved in many functions including olfactory receptor activity, translation, transcription, neuron fate commitment, and structural constituent of ribosomes (Table 3). Among the neuro-associated enriched motifs is a binding motif for Pitx3, a transcriptional regulator involved in the differentiation and maintenance of dopaminergic neurons during development (Chung et al., 2005). A paralog of PITX3 with a similar binding motif is PITX1, which has been implicated in autism (Philippi et al., 2007). In Auts2 marked non-promoter regions, we identified motifs involved in many functions including calcium ion binding, translation, structural constituent of ribosomes, and olfactory receptor activity (Table 3). Among the known motifs enriched in non-promoter marked regions is the T-cell acute lymphocyte leukemia/TCF3 heterodimer (TAL1::TCF3). Tcf3 is a transcriptional regulator expressed in the developing cerebral cortex involved in the regulation cell growth, differentiation, and commitment of multiple cell lineages including neurons (Gray et al., 2004; Massari & Murre, 2000; Slattery, Ryan, & McMorrow, 2008). TAL1 and TCF3 are also associated with Acute Lymphoblastic Leukemia (ALL) (Barber et al., 2007; Kelliher, Seldin, & Leder, 1996), which has been associated with *AUTS2* (Oksenberg & Ahituv, 2013). Also enriched is FOXO3, a transcriptional activator involved in neuronal cell death (Xie et al., 2012) (Table 3). Using MEME and ChIPmunk, two *de novo* motif algorithms (Bailey & Gribskov, 1994; Kulakovskiy, Boeva, Favorov, & Makeev, 2010), I was

only able to identify motifs comprised of simple repeats, for instance [CA]ACA[CA]ACA[CA]ACA.

3.4 Discussion

My mouse E16.5 forebrain RNA-seq and ChIP-seq analysis lends further support to a neurodevelopmental role for Auts2. I showed that Auts2 is frequently and significantly localized near promoters of active genes, suggesting that Auts2 is involved in the activation or maintenance of gene expression. IPA and GREAT analyses identified significant associations with neurodevelopmental genes and pathways near Auts2 ChIP-seq peaks. Both GREAT and IPA found highly significant enrichment for genes involved in gene expression and ribosome related proteins. Interestingly, dynamic regulation of individual ribosomal proteins control gene expression and mammalian development within vertebrate embryos (Kondrashov et al., 2011). Future experiments are needed to determine if and how *AUTS2* and ribosomes work together to affect neurodevelopment. IPA analysis also identified axonal guidance signaling as a significantly enriched canonical pathway, which includes the *SEMA5A* gene and supports *AUTS2's* role in *SEMA5A* related pathways including neurodevelopment and ASD progression (Cheng et al., 2013).

My analyses identified several interesting genes that Auts2 may target. Glycerophosphodiester phosphodiesterase domain containing 1 (*Gdpd1*) contains an Auts2 mark at its promoter and is highly expressed in our RNA-seq dataset (FPKM=2.58). *GDPD1* encodes a glycerophosphodiester phosphodiesterase and could be a candidate region linking AUTS2 to alcoholism given that other classes of phosphodiesterases are involved in alcohol seeking and consumption behaviors (Wen et al., 2012). GABA B receptor 1 (*Gabbr1*), which also contains an

Auts2 marked site at its promoter and is highly expressed in our dataset (FPKM=11.18), is a key component of GABAergic signaling important for synaptic regulation and is implicated in autism, epilepsy, and alcohol and cocaine addiction (Enoch et al., 2012; Fatemi, Folsom, Reutiman, & Thuras, 2009; Xi et al., 2011). Ubiquitin carboxyl-terminal esterase L1 (*Uchl1*) is very highly expressed in our dataset (FPKM= 43.36) and displays Auts2 localization at its promoter. *Uchl1* is expressed specifically in neurons and is a strong Parkinson's disease susceptibility gene (Maraganore et al., 2004), providing insight on a potential new role for *AUTS2* in the progression of Parkinson's disease. Another very highly expressed gene (FPKM=28.08) that has a promoter marked with Auts2 is neruocan (*Ncan*), a gene thought to be involved in axon guidance (Watanabe et al., 1995) that is implicated in bipolar disorder and schizophrenia (Cichon et al., 2011; Mühleisen et al., 2012). These genes represent possible connections between *AUTS2* and a wide range of neurological disorders and provide a list of candidate interactions for further functional studies of the role of *AUTS2* in neurodevelopment and disease.

In addition to neurodevelopment, *AUTS2* may play a role in cancer. Previous studies have linked *AUTS2* with ALL (Oksenberg & Ahituv, 2013), matching my finding of enrichment of the TCF3 motif in Auts2 marked sites, since TCF3 is associated with ALL (Barber et al., 2007). Additionally, our GREAT results identified the MSigDB perturbation term "genes whose DNA methylation differ between primary ALL cells and peripheral blood samples" as significant ($p=7.82e^{-6}$; Binomial Test; non-promoter marks).

De novo motif analysis of Auts2 marked regions did not find any non-repetitive novel motifs, and the distribution of known motifs was not centrally enriched within the Auts2 marked regions, supporting the theory that Auts2 acts as a co-factor rather than by binding directly to

DNA. AUTS2 lacks identified DNA-binding motifs but contains several predicted protein-protein interaction domains including an SH2, a PY, and 13 SH3 domains (Bedogni, Hodge, Nelson, et al., 2010; Oksenberg & Ahituv, 2013; Sultana et al., 2002). More research needs to be done to determine Auts2 binding partners and how they converge to bind DNA.

Auts2 binding analysis revealed that non-promoter Auts2 marked sites overlap the enhancer mark H3K27ac significantly more than expected by chance, but not with the repressive H3K27me3 mark, suggesting that Auts2 binds to active enhancer regions throughout the genome. Taken together, this suggests that Auts2 also has an activating role in distant gene regulatory elements. I functionally characterized ten Auts2 marked sequences near genes implicated in ASD using a transgenic zebrafish enhancer assay. Four AMECs showed positive zebrafish enhancer expression, two of which were positive in the brain. AMEC1, which showed positive zebrafish expression in the olfactory epithelium, is in the intron of *NRXN1*, a gene involved in synapse formation and signaling that has been implicated in ASD and other neurological disorders (Ching et al., 2010; Missler & Südhof, 1998). *Nrxn1* is also expressed in the olfactory epithelium at E14.5 (Diez-Roux et al., 2011), suggesting this enhancer could regulate Nrxn1. Several other lines of evidence, including our motif analysis, known expression patterns of Auts2 in mouse and zebrafish (Bedogni, Hodge, Nelson, et al., 2010; Oksenberg et al., 2013), and my IPA and GREAT analyses, further support an olfactory role for Auts2. AMEC8 lies in an intron of *ATP2B2*, which is implicated in ASD, likely due to altered Ca^{2+} signaling when *ATP2B2* is defective (Carayol et al., 2011; Yang et al., 2013), This sequence showed positive zebrafish enhancer activity in trigeminal sensory neurons matching *Atp2b2*'s expression in the mouse trigeminal ganglion (Diez-Roux et al., 2011).

Using ChIP-seq and RNA-seq on mouse E16.5 forebrains I identified potential Auts2 regulated regions and found that they mark active regulatory elements. I located 1,930 targets of Auts2, about 40% of which lie in promoter regions. Transcripts with Auts2-bound promoters had significantly higher expression levels than ones without Auts2 binding. Both the genes and regulatory sequences that are bound by Auts2 provide distinctive candidate regions to investigate nucleotide variation associated with neurodevelopmental disorders. *AUTS2* is emerging as a critical regulator of active neurodevelopmental genes, and future studies such as mouse knockouts could confirm and increase our understanding of the neurodevelopmental function of this gene.

3.5 Materials and Methods

3.5a ChIP-seq:

Mouse embryos were harvested from timed pregnant CD-1 females (Charles River) at E16.5. The forebrains were dissected in cold PBS and batches of 8 forebrains each were collected in a tube and washed twice with PBS. Forebrains were cut to <1mm size and crosslinked with 1% formaldehyde for 10 minutes. Chromatin from forebrain tissue was isolated and sheared using a Bioruptor (Diagenode) and immunoprecipitation was performed using 5 micrograms of anti-AUTS2 antibody (HPA000390, Sigma Aldrich) (previously verified to be specific for the Auts2 protein) (Bedogni, Hodge, Nelson, et al., 2010). Chromatin from the same sample was processed for the input control. Illumina libraries were constructed from ChIP and input DNA by the UC Davis Genomics core and sequenced on a HiSeq2500 (Illumina). Auts2 ChIP-seq data from this study is available in GEO (http://www.ncbi.nlm.nih.gov/geo/) [accession

number TBD, currently SRA experiment SRX433885]. ChIP-seq reads were demultiplexed and aligned to the mouse genome (mm9) using Bowtie(Langmead, Trapnell, Pop, & Salzberg, 2009) allowing one mismatch per read and retaining only reads with a single reportable alignment (example command: bowtie -p 12 -m 1 -v 1 -S). Resulting SAM files were converted to BAM format for peak calling with MACS2 (version 2.0.10) (Y. Zhang et al., 2008). For peak calling, MACS2 extended the reads to 300bp and kept only peaks where FDR≤0.01 after normalizing relative to input DNA (example command: macs2 callpeak -f BAM -g mm --keep-dup 1 --nomodel --extsize 300 -q 0.01).

3.5b RNA-seq:

Total RNA was extracted from two replicates of E16.5 mouse forebrain tissue and purified using the RNeasy Maxi Kit (Qiagen) and sent to Otogenetics for ribosomal RNA depletion, cDNA production using random primers, library preparation, and pair end sequencing (~100bp) using Illumina HiSeq2000. Resulting reads were demultiplexed and aligned to the mouse genome (mm9) using TopHat v2.0.7(Kim et al., 2013; Langmead et al., 2009; Trapnell, Pachter, & Salzberg, 2009). The two replicates where merged and read counts mapping to each transcript were obtained using HTseq (Anders et al., 2013). Expression of each transcript was quantified as fragments per kilobase of transcript per million fragments mapped (FPKM) by dividing the total number of reads mapping to each transcript by transcript length, and the total number of reads aligned to the genome divided by one million. The Wilcoxon test from the statistical toolkit R was used to calculate differences in gene expression between genes whose promoter contains an Auts2 marked site and all other genes. RNA-seq data from this study is available in GEO (http://www.ncbi.nlm.nih.gov/geo/) [accession number TBD, currently SRA experiment SRX433887].

3.5c Analyzing the genomic context near Auts2 peaks:

Distance from Auts2 marked sites to nearest transcription start site (TSS) was assessed using genomic coordinates downloaded from the UCSC Genome Browser's mm9 RefSeq Genes track. Histone modification ChIP-seq data from mouse E14.5 whole brain (Shen et al., 2012) was downloaded from UCSC Genome Browser (H3K4me3, H3K27ac, and H3K27me3). All datasets were downloaded in October 2013. BedTools'(Quinlan & Hall, 2010) *intersectBed* command was used to identify overlap between this study's Auts2 marked sites and histone modification peaks from E14.5 mouse whole brain. A single base pair overlap was sufficient to consider two regions overlapping. Significance was calculated using a permutation p-test (1000 permutations). To determine how many Auts2 marked sites reside within each gene in the mouse genome, genes were defined by transcription start and end sites in mm9 RefSeq. BedTools' *windowBed* command (with 1bp window) was used to count the number of Auts2 marked sites within each gene.

3.5d Gene ontology, pathway, and motif analysis:

Pathway analysis was performed using Ingenuity Pathway Analysis (IPA) (Ingenuity® Systems, www.ingenuity.com) by inputting a list of genes near Auts2 marked regions (one gene per Auts2 marked site, based on closest distance to TSS). Gene ontology analysis was performed using Genomic Regions Enrichment of Annotations Tool (GREAT) (McLean et al., 2010). To associate genomic regions with genes in GREAT, I used the "basal plus extension" setting, except when examining promoter marked sites, in which case "single nearest gene" was chosen. Background regions were defined using GREAT's "whole genome" setting except when examining promoter marked sites, where we defined a promoter-specific background containing all promoter regions (2,500bp upstream and 500bp downstream of a TSS) plus a window the size

of the largest Auts2 ChIP-seq peak (1,457bp). MEME and ChIPmunk (Bailey & Gribskov, 1994; Kulakovskiy et al., 2010) were used to search for *de novo* motifs within Auts2 marked sites. MEME-ChIP and GOMO (Buske et al., 2010; Machanick & Bailey, 2011) were used to identify experimentally characterized motifs and their gene ontologies.

3.5e Transgenic enhancer assays:

Enhancer candidate sequences were selected from Auts2 marked sites based on proximity to SFARI genes that had a gene score of 1-3 (corresponding to high confidence genes, strong candidates and genes with suggestive evidence),(Abrahams et al., 2013) evolutionary conservation (sequences showing ≥70% identity for at least 100 bp), and/or overlap with enhancer-associated histone marks (H3K27ac, H3K4me1), but not promoters (H3K4me3) from whole brain E14.5 ChIP-seq datasets (Shen et al., 2012). PCR was done on human genomic DNA (Qiagen) and products were cloned into the E1b-GFP-Tol2 enhancer assay vector containing an E1b minimal promoter followed by GFP (Q. Li et al., 2010) and verified by sequencing. Constructs were injected following standard procedures (Nusslein-Volhard, 2002; Westerfield, 2007) into at least 100 embryos along with Tol2 mRNA (Kawakami, 2005), to facilitate genomic integration. GFP expression was observed and annotated up to 48 hours post-fertilization (hpf). An enhancer was considered positive if at least 15% of all fish surviving to 48 hpf showed a consistent expression pattern after subtracting out percentages of tissue expression in fish injected with the empty enhancer vector. For each construct, at least 50 fish were analyzed for GFP expression at 48 hpf. All animal work was approved by the UCSF Institutional Animal Care and Use Committee (protocol number AN100466).

3.6 Acknowledgements

I would like to thank Megan Laurance for her assistance with IPA and Mariel Mckenzie Finucane with her assistance with statistics.

Chapter 3 tables

Table 1: Complete IPA and GREAT data: All data from both IPA and GREAT (including all Auts2 marked regions, promoter Auts2 marked regions, and non-promoter Auts2 marked regions).

Due to the size of this table, it will only be available through download of the excel file once the paper is released. Table name: Table S1. Manuscript name: Genome-wide distribution of Auts2 binding localizes with active neurodevelopmental genes

Table 2: Auts2 marked enhancer candidates selected for enhancer assays

AMEC	mm9	hg19 (inj seq)	Nearby gene	Distance to ASD gene (TSShg19)	Whole brain E14.5 marks (mm9)**	Enhancer results
1	chr17:91,051,949-91,054,230	chr2:50,791,381-50,793,765	Nrxn1	Intron		Forebrain (17.6%, 48hpf) Heart (21.9%, 48hpf)
2	chr6:105,553,911-105,556,644	chr3:2,082,346-2,084,261	Cntn4	56kb		Somitic muscle (19%, 48hpf)
3	chr4:65,271,249-65,273,628	chr9:119,455,679-119,459,219	Astn2	Intron	H3K4me1+ (H3K4me3 -)	Negative
4	chr5:132,299,103-132,302,306	chr7:69,848,219-69,849,944	Auts2	Intron	H3K4me1+ (H3K4me3 -) H3K27ac +	Negative
5	chr16:5,431,204-5,435,251	chr16:5,494,064-5,498,310	Rbfox1	571kb		Notochord (21.9% 48hpf)
6	chr5:21,767,725-21,768,928	chr7:103,530,668-103,533,035	Reln	Intron		Negative
7	chr5:22,181,802-22,192,961	chr7:103,946,391-103,948,122	Reln	316kb		Negative

8	chr6:113,804,242-113,807,961	chr3:10,503,994-10,507,102	Atp2b2	Intron	H3K4me1+ (H3K4me3 -) H3K27ac +	Hindbrain (15.7%, 24hpf) Midbrain (24.1%, 24hpf) Spinal cord (18.1%, 24hpf) Somitic muscle (19%, 48hpf)
9	chr6:14,769,541-14,771,373	chr7:113,626,516-113,628,368	Foxp2	300kb		Negative
#	chr6:45,438,657-45,445,755	chr7:146,302,256-146,305,323	Cntnap2	Intron		Negative

** Shen, Y. *et al.* (2012) A map of the cis-regulatory sequences in the mouse genome. *Nature* 488, 116–20

Table 3: MEME-ChIP and GOMO results: Known motifs enriched in our datasets and their related gene ontology.

Due to the size of this table, it will only be available through download of the excel file once the paper is released. Table name: Table S3. Manuscript name: Genome-wide distribution of Auts2 binding localizes with active neurodevelopmental genes

Chapter 3 figures

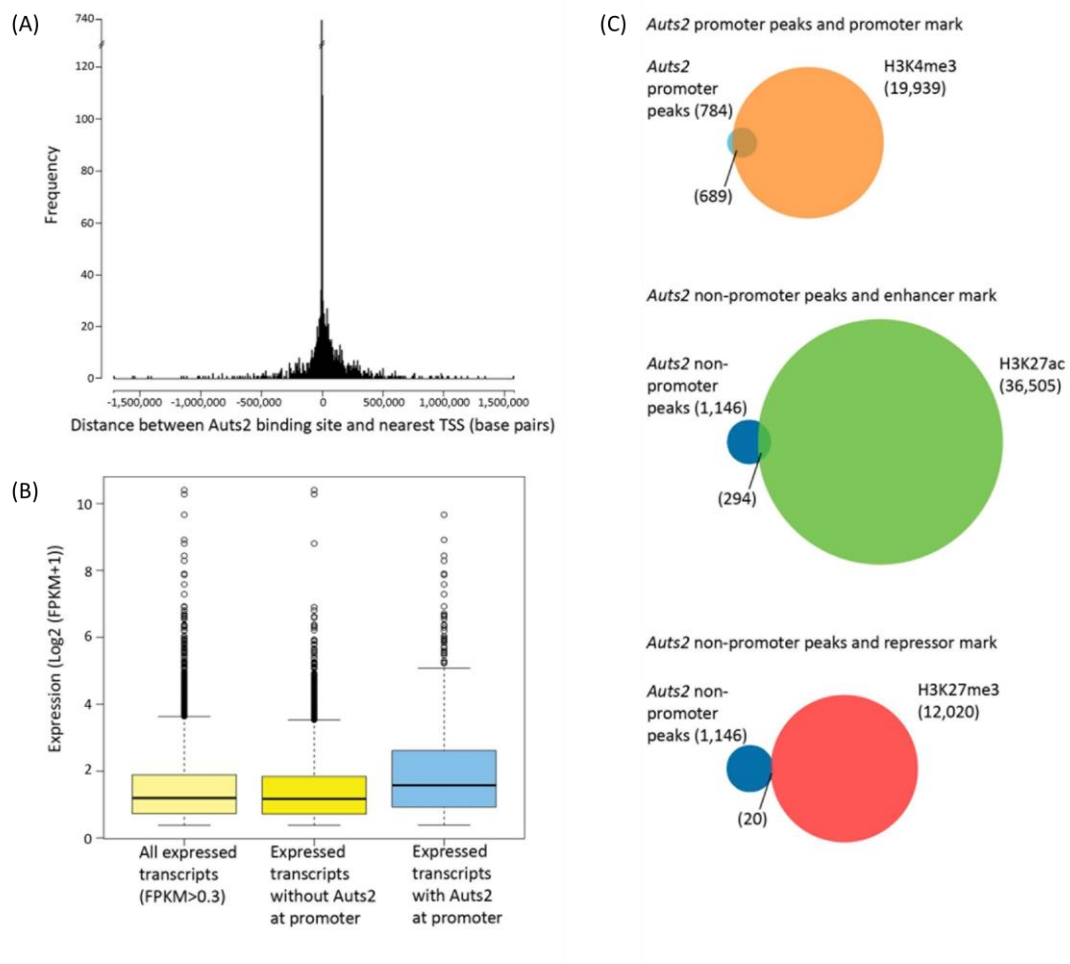


Figure 1: Analysis of Auts2 ChIP-seq peaks

(A) Distance distribution of the 1,930 Auts2 marked sites to the nearest TSS shows preferential binding near TSSs. Histogram displays bins of 5kb. (B) FPKM transcript expression scores (FPKM>0.3) for genes whose promoters localized with Auts2 display significantly higher expression than those that do not ($p<2.2e^{-16}$; Wilcoxon test). (C) Overlaps of Auts2 marked sites with histone modifications show significant localization of Auts2 at promoters (H3K4me3) and active enhancers (H3K27ac) ($p<0.001$; permutation test) but not repressed regions (H3K27me3)

(p value =0.081; permutation test). Histone data was acquired from previously reported ChIP-seq for mouse E14.5 whole brain (Shen et al., 2012).

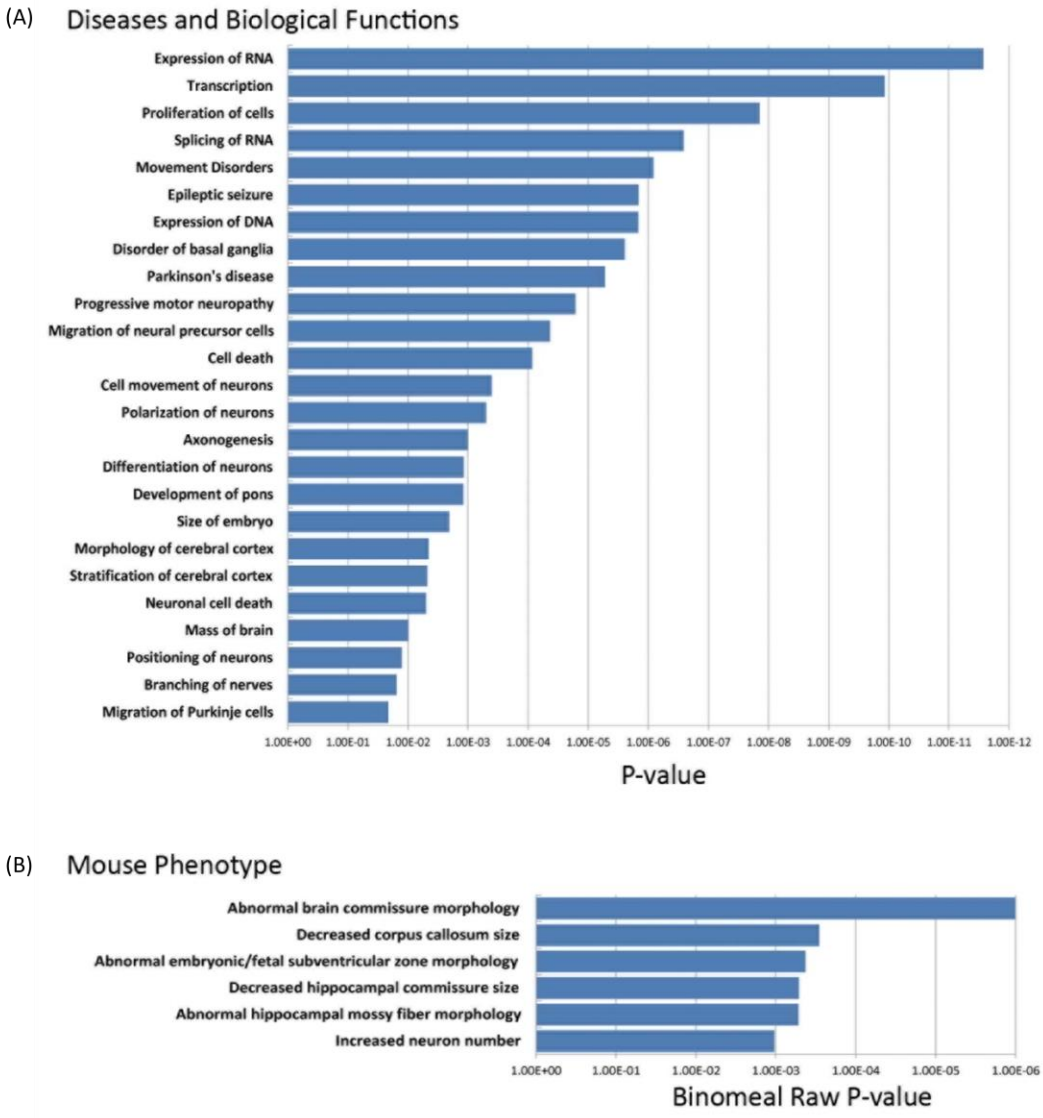


Figure 2: Pathway analysis and gene ontology of Auts2 marked sites

(A) IPA pathway analysis of genes whose promoters contain an Auts2 marked site; the figure shows selected significant neurological, gene expression, and cell cycle related disease and biological functions. (B) GREAT (McLean et al., 2010) gene ontology analysis of non-promoter Auts2 marked sites; the figure shows all significant neurological related mouse phenotypes.

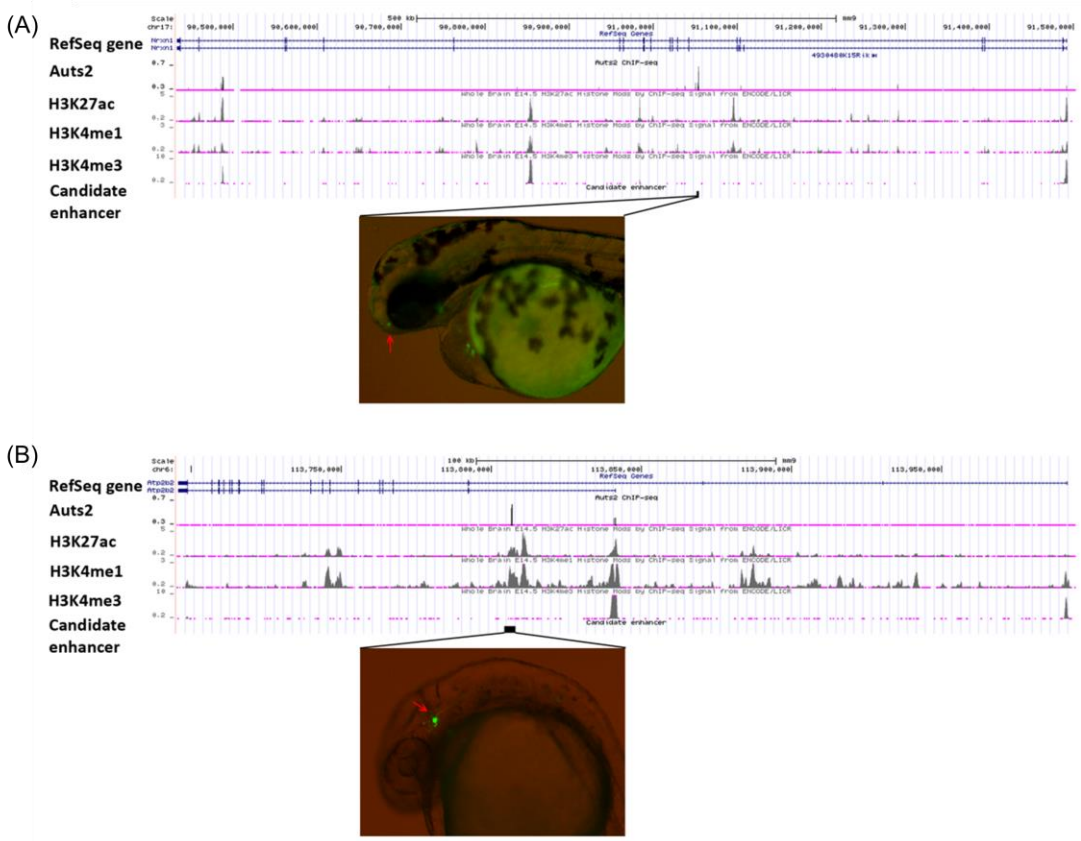


Figure 3: Auts2 marked enhancers

(A) A UCSC Genome Browser snapshot of the *Nrxn1* locus in mm9, including tracks for the RefSeq gene, Auts2 ChIP-seq, whole brain E14.5 H3K27ac/H3k4me1/H3kme3 ChIP-seq,(Shen et al., 2012) and the Auts2 marked enhancer candidate (AMEC) 1. A representative picture of AMEC1 shows positive enhancer activity in the zebrafish heart and forebrain (olfactory epithelium (red arrow)) at 48 hours post fertilization (hpf) is shown below. (B) UCSC browser snapshot of the *Atp2b2* region including RefSeq and ChIP-seq tracks. Below, a representative 24 hpf embryo shows enhancer activity of AMEC8 in the midbrain (red arrow) and hindbrain (trigeminal sensory neurons).

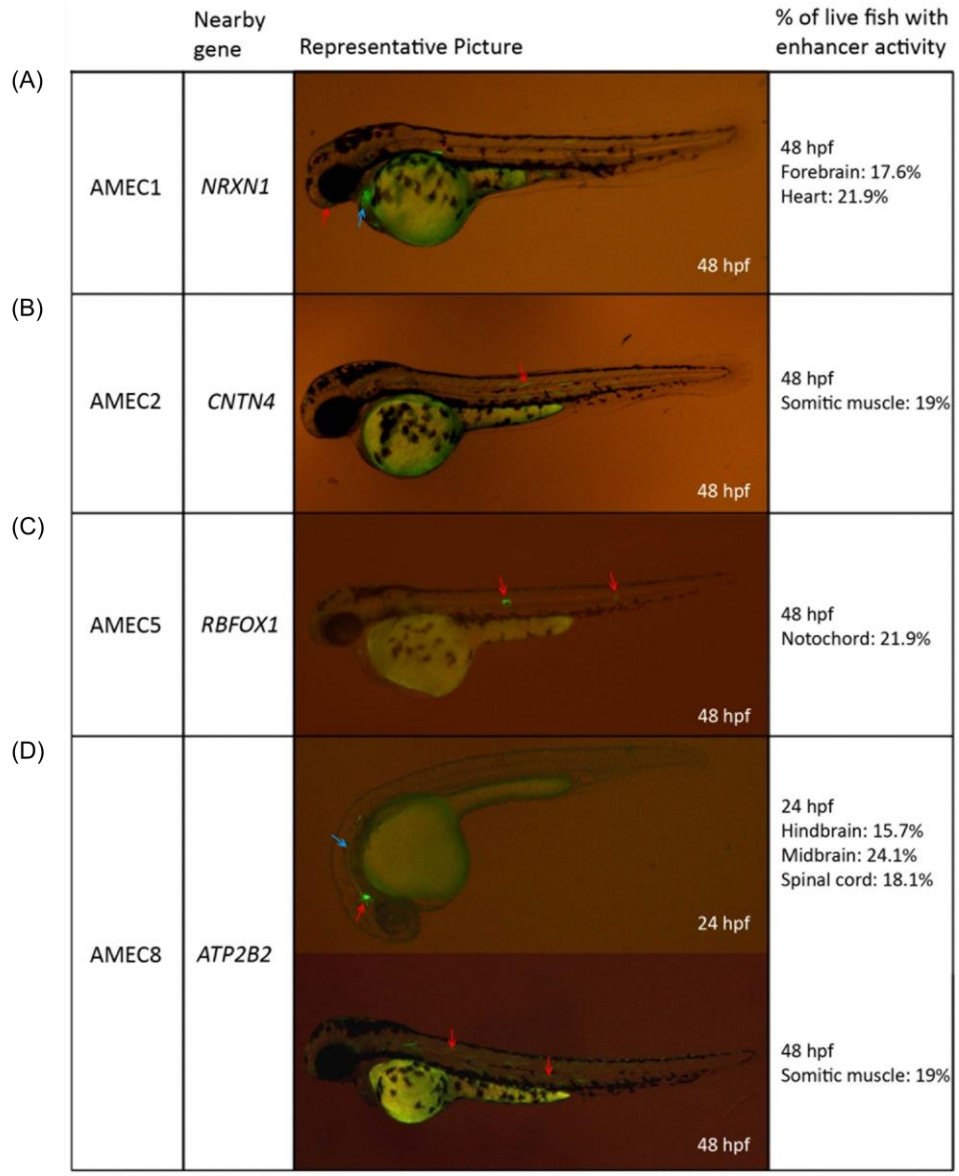


Figure 4: Positive zebrafish enhancers

(A) AMEC1 lies in the intron of *NRXN1*, and is positive for enhancer activity in the forebrain (red arrow) and heart (blue arrow) at 48 hours post fertilization (hpf). (B) AMEC2 is located ~56kb upstream of *CNTN4*, and is positive in the somitic muscle (red arrow) at 48 hpf. (C) AMEC5 resides ~571kb upstream of to *RBFOX1*, and is positive in the notochord (red arrows) at 48 hpf. (D) AMEC8, located in an intron of *ATP2B2*, was found to be positive for enhancer

activity in the hindbrain (blue arrow) and the midbrain (red arrow) at 24 hpf and in somitic muscles (red arrows) at 48 hpf. AMEC: Auts2 marked enhancer candidate.

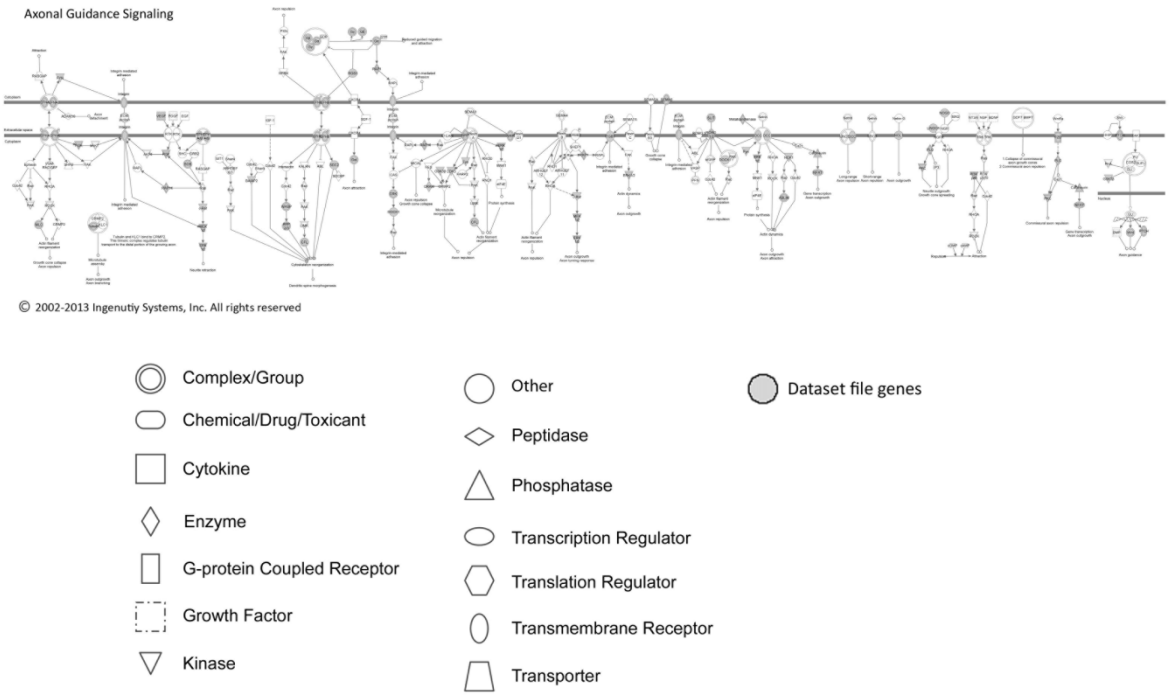
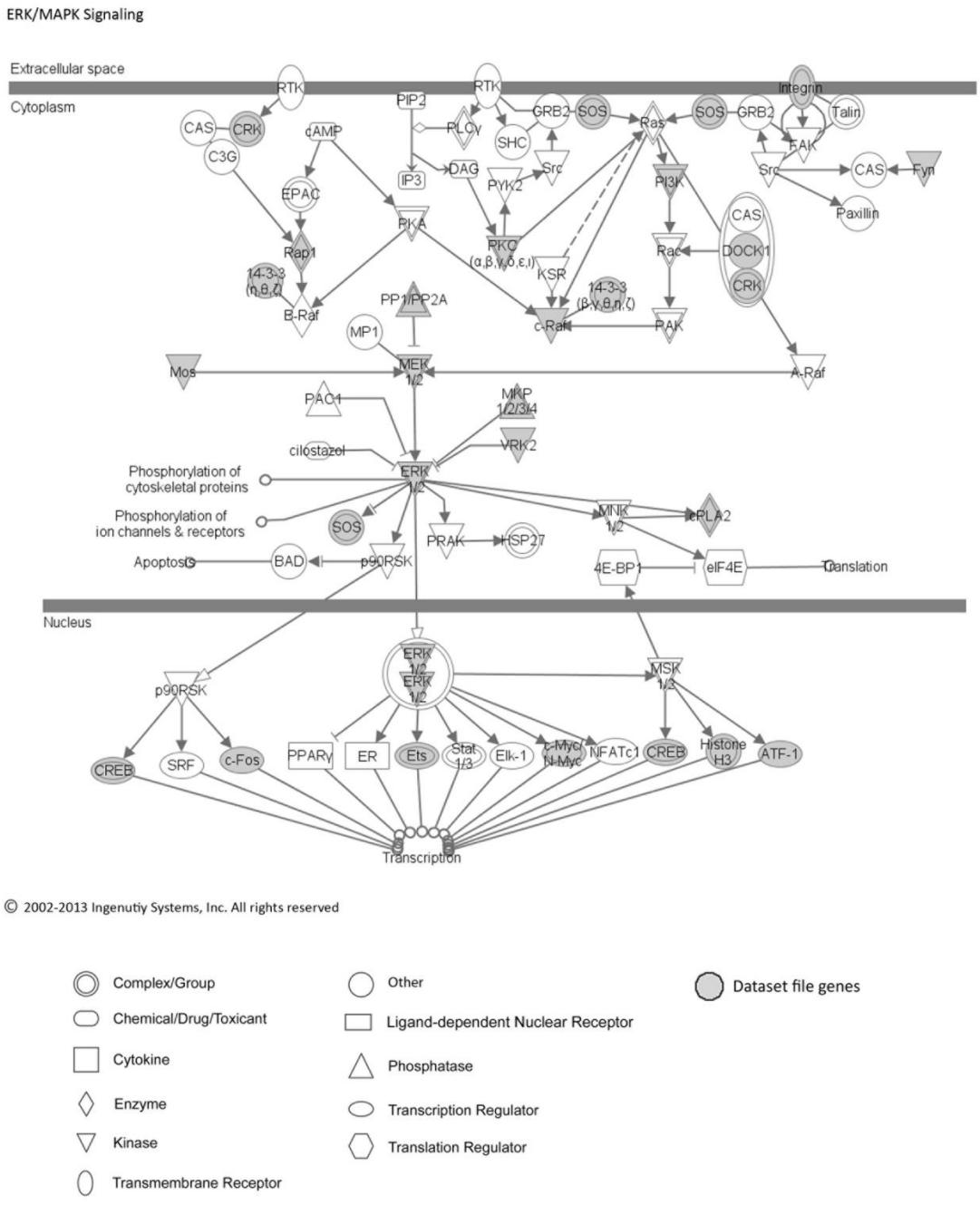


Figure 5 IPA canonical pathway: Axonal guidance signaling.

Highlighted in grey are genes whose TSS are nearest to an Auts2 marked site.



Complex/Group
Chemical/Drug/Toxicant
Cytokine
Enzyme
Kinase
Transmembrane Receptor
Other
Ligand-dependent Nuclear Receptor
Phosphatase
Transcription Regulator
Translation Regulator
Dataset file genes

Figure 6: IPA canonical pathway: ERK/MAPK signaling

Highlighted in grey are genes whose TSS are nearest to an Auts2 marked site.

Chapter 4 – Conclusions and future directions

Over the past several years, *AUTS2* has emerged as critical neurodevelopmental gene that has been implicated as a causative gene in ASD and ASD related phenotypes in dozens of individuals. In the near future, mutations around *AUTS2* may act as a diagnostic, but first we must determine what makes a CNV in or around *AUTS2* causative or benign and for what disorders (e.g. ID, DD, ASD, ASD with ID/DD, etc).

Knockdown of *auts2* leads to stunted development and a reduction in motor and sensory neurons in the spinal cord and developing neurons in regions that include the midbrain and cerebellum. The overall stunted development of these zebrafish makes it difficult to determine the precise function of this gene.

My search for enhancers of *AUTS2* focused primarily on the first five introns due to the numerous reports of cognitive-related structural variations in that region (Ben-David et al., 2011; Girirajan et al., 2011; Glessner et al., 2009; Huang et al., 2010; Kalscheuer et al., 2007; Mefford et al., 2010; Nagamani et al., 2013; Pinto et al., 2010; Schumann et al., 2011; Sultana et al., 2002; Talkowski et al., 2012), along with the region's putative role in evolution. While the expression of our enhancers largely recapitulated *Auts2* expression, it is possible that the enhancers we identified could regulate a neighboring gene. Future experiments such as chromatin interaction analyses (Fullwood et al., 2010; Lieberman-Aiden et al., 2009) could be able to distinguish what promoters our enhancers are interacting with. I found three positive human enhancers in both zebrafish and mouse that reside within a 33,519bp deletion detected in an individual with ASD, one of which, AEC32, is expressed in the cerebellum. With the constantly growing number of individuals with ASDs or other neurological phenotypes that have *AUTS2* mutations, some of which are purely noncoding, it is likely that improper regulation of

this gene is involved in the progression of these disorders. By characterizing the regulatory landscape of this region I have obtained a better understanding of the functional units within this gene.

AUTS2 has been singled out as a gene that is rapidly evolving in humans in three different studies (Green et al., 2010; Pollard, Salama, King, et al., 2006; Prabhakar et al., 2006). I identified sixteen different enhancers that lie within regions that were implicated in human evolution, six of which show expression in the brain. I tested four of the enhancers in mice and two of them had midbrain enhancer activity. My enhancer results, combined with the observation that human-specific neurological disorders are associated with mutations in this gene, suggest that *AUTS2* has an important role in the evolution of human cognitive traits. Future experiments should dissect what effects these nucleotide changes have on transcription factor binding and regulatory function.

My mouse E16.5 forebrain RNA-seq and ChIP-seq analysis lends further support to a neurodevelopmental role for Auts2. Auts2 is frequently and significantly localized near promoters of active genes, suggesting that Auts2 is involved in the activation or maintenance of gene expression. IPA and GREAT analyses identified significant associations with neurodevelopmental genes and pathways near Auts2 ChIP-seq peaks. It would be interesting to investigate Auts2 binding at different time points and tissues to reveal a better understanding of the temporal and spatial regulatory effects of this protein.

De novo motif analysis of Auts2 marked regions did not find any non-repetitive novel motifs, and the distribution of known motifs was not centrally enriched within the Auts2 marked regions, supporting the theory that Auts2 acts as a co-factor rather than by binding directly to DNA. AUTS2 lacks identified DNA-binding motifs but contains several predicted protein-

protein interaction domains including an SH2, a PY, and 13 SH3 domains (Bedogni, Hodge, Nelson, et al., 2010; Oksenberg & Ahituv, 2013; Sultana et al., 2002). More research needs to be done to determine Auts2 binding partners and how they converge to bind DNA. Future experiments should take a closer look at potential Auts2 binding partners to determine the mechanism by which the protein affects neurodevelopment. Experiments showing protein-protein and protein-DNA interactions such as EMSAs would be critical for accomplishing this.

Auts2 binding analysis revealed that non-promoter Auts2 marked sites overlap the enhancer mark H3K27ac significantly more than expected by chance, but not with the repressive H3K27me3 mark, suggesting that Auts2 binds to active enhancer regions throughout the genome. This suggests that Auts2 has an activating role in distant gene regulatory elements. I functionally characterized ten Auts2 marked sequences near genes implicated in ASD using a transgenic zebrafish enhancer assay. Four AMECs showed positive zebrafish enhancer expression, two of which were positive in the brain. AMEC1, which showed positive zebrafish expression in the olfactory epithelium, is in the intron of *NRXN1*, a gene involved in synapse formation and signaling that has been implicated in ASD and other neurological disorders (Ching et al., 2010; Missler & Südhof, 1998). AMEC8 lies in an intron of *ATP2B2,* which is implicated in ASD, likely due to altered Ca^{2+} signaling when *ATP2B2* is defective (Carayol et al., 2011; Yang et al., 2013). This sequence showed positive zebrafish enhancer activity in trigeminal sensory neurons matching *Atp2b2*'s expression in the mouse trigeminal ganglion (Diez-Roux et al., 2011).

In summary, *AUTS2* is emerging as both a critical neurodevelopmental gene and a regulator of active neurodevelopmental genes. The next crucial experiment to perform is an *Auts2* mouse knockout. This will help elucidate the function of this gene in a mammalian model.

The combination of mouse knock out along with an RNA-seq to determine what genes are affected by improper function of *Auts2* will yield interesting potential mechanisms of gene function, especially if this data is paired with my ChIP-seq and RNA-seq data described above.

References

Abrahams, B. S., Arking, D. E., Campbell, D. B., Mefford, H. C., Morrow, E. M., Weiss, L. a, … Packer, A. (2013). SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). *Molecular Autism*, *4*(1), 36. doi:10.1186/2040-2392-4-36

Abrahams, B. S., & Geschwind, D. H. (2008). Advances in autism genetics: on the threshold of a new neurobiology. *Nature Reviews. Genetics*, *9*(5), 341–55. doi:10.1038/nrg2346

Alenzi, F. Q. B. (2004). Links between apoptosis, proliferation and the cell cycle. *Br J Biomed Sci*, *61*(2), 99–102. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15250676

Ameur, A., Zaghlool, A., Halvardson, J., Wetterbom, A., Gyllensten, U., Cavelier, L., & Feuk, L. (2011). Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. *Nature Structural & Molecular Biology*, *18*(12), 1435–40. doi:10.1038/nsmb.2143

Anders, S., McCarthy, D. J., Chen, Y., Okoniewski, M., Smyth, G. K., Huber, W., & Robinson, M. D. (2013). Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nature Protocols*, *8*(9), 1765–86. doi:10.1038/nprot.2013.099

Arlt, M. F., Rajendran, S., Birkeland, S. R., Wilson, T. E., & Glover, T. W. (2012). De novo CNV formation in mouse embryonic stem cells occurs in the absence of Xrcc4-dependent nonhomologous end joining. *PLoS Genetics*, *8*(9), e1002981. doi:10.1371/journal.pgen.1002981

Arlt, M., & Ozdemir, A. (2011). Hydroxyurea induces de novo copy number variants in human cells. *Proc Natl Acad Sci U S A*, *2011*. doi:10.1073/pnas.1109272108/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1109272108

Bailey, T. L., & Gribskov, M. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. In *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology* (pp. 28–36). Menlo Park, California.

Baio, J. (2012). Prevalence of Autism Spectrum Disorders — Autism and Developmental Disabilities Monitoring Network , 14 Sites , United States , 2008. *Centers for Disease Control and Prevention MMWR Surveillance Summaries*, *61*(3).

Bakkaloglu, B., Roak, B. J. O., Louvi, A., Gupta, A. R., Abelson, J. F., Morgan, T. M., … State, M. W. (2008). Molecular Cytogenetic Analysis and Resequencing of Contactin Associated Protein-Like 2 in Autism Spectrum Disorders. *Amer J Hum Genet*, (January), 165–173. doi:10.1016/j.ajhg.2007.09.017.

Barber, K. E., Harrison, C. J., Broadfield, Z. J., Stewart, A. R. M., Wright, S. L., Martineau, M., … Moorman, A. V. (2007). Molecular Cytogenetic Characterization of TCF3 Acute Lymphoblastic Leukemia. *Genes, Chromosomes and Cancer*, *46*(5), 478–486. doi:10.1002/gcc

Bedogni, F., Hodge, R. D., Elsen, G. E., Nelson, B. R., Daza, R. a M., Beyer, R. P., … Hevner, R. F. (2010). Tbr1 regulates regional and laminar identity of postmitotic neurons in developing neocortex. *Proc Natl Acad Sci U S A*, *107*(29), 13129–34. doi:10.1073/pnas.1002285107

Bedogni, F., Hodge, R. D., Nelson, B. R., Frederick, E. a, Shiba, N., Daza, R. a, & Hevner, R. F. (2010). Autism susceptibility candidate 2 (Auts2) encodes a nuclear protein expressed in developing brain regions implicated in autism neuropathology. *Gene Expr Patterns*, *10*(1), 9–15. doi:10.1016/j.gep.2009.11.005

Ben-David, E., Granot-Hershkovitz, E., Monderer-Rothkoff, G., Lerer, E., Levi, S., Yaari, M., … Shifman, S. (2011). Identification of a functional rare variant in autism using genome-wide screen for monoallelic expression. *Human Molecular Genetics*, *20*(18), 3632–41. doi:10.1093/hmg/ddr283

Ben-Shachar, S., Chahrour, M., Thaller, C., Shaw, C. a, & Zoghbi, H. Y. (2009). Mouse models of MeCP2 disorders share gene expression changes in the cerebellum and hypothalamus. *Human Molecular Genetics*, *18*(13), 2431–42. doi:10.1093/hmg/ddp181

Beunders, G., Voorhoeve, E., Golzio, C., Pardo, L. M., Rosenfeld, J. a, Talkowski, M. E., … Sistermans, E. a. (2013). Exonic Deletions in AUTS2 Cause a Syndromic Form of Intellectual Disability and Suggest a Critical Role for the C Terminus. *American Journal of Human Genetics*, *92*(2), 210–20. doi:10.1016/j.ajhg.2012.12.011

Buske, F. a, Bodén, M., Bauer, D. C., & Bailey, T. L. (2010). Assigning roles to DNA regulatory motifs using comparative genomics. *Bioinformatics (Oxford, England)*, *26*(7), 860–6. doi:10.1093/bioinformatics/btq049

Carayol, J., Sacco, R., Tores, F., Rousseau, F., Lewin, P., Hager, J., & Persico, A. M. (2011). Converging evidence for an association of ATP2B2 allelic variants with autism in male subjects. *Biological Psychiatry*, *70*(9), 880–7. doi:10.1016/j.biopsych.2011.05.020

Chahrour, M. H., Yu, T. W., Lim, E. T., Ataman, B., Coulter, M. E., Hill, R. S., … Walsh, C. a. (2012). Whole-exome sequencing and homozygosity analysis implicate depolarization-regulated neuronal genes in autism. *PLoS Genetics*, *8*(4), e1002635. doi:10.1371/journal.pgen.1002635

Chen, Y.-H., Liao, D.-L., Lai, C.-H., & Chen, C.-H. (2013). Genetic analysis of AUTS2 as a susceptibility gene of heroin dependence. *Drug and Alcohol Dependence*, *128*(3), 238–42. doi:10.1016/j.drugalcdep.2012.08.029

Cheng, Y., Quinn, J. F., & Weiss, L. A. (2013). An eQTL mapping approach reveals that rare variants in the SEMA5A regulatory network impact autism risk. *Hum Mol Genet*, 1–13. doi:10.1093/hmg/ddt150

Ching, M. S. L., Shen, Y., Tan, W.-H., Jeste, S. S., Morrow, E. M., Chen, X., … Wu, B.-L. (2010). Deletions of NRXN1 (neurexin-1) predispose to a wide spectrum of developmental disorders. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics : The Official Publication of the International Society of Psychiatric Genetics*, *153B*(4), 937–47. doi:10.1002/ajmg.b.31063

Chojnicka, I., Gajos, K., Strawa, K., Broda, G., Fudalej, S., Fudalej, M., … Płoski, R. (2013). Possible Association between Suicide Committed under Influence of Ethanol and a Variant in the AUTS2 Gene. *PloS One*, *8*(2), e57199. doi:10.1371/journal.pone.0057199

Chung, S., Hedlund, E., Hwang, M., Kim, D. W., Shin, B.-S., Hwang, D.-Y., … Kim, K.-S. (2005). The homeodomain transcription factor Pitx3 facilitates differentiation of mouse embryonic stem cells into AHD2-expressing dopaminergic neurons. *Molecular and Cellular Neurosciences*, *28*(2), 241–52. doi:10.1016/j.mcn.2004.09.008

Cichon, S., Mühleisen, T. W., Degenhardt, F. a, Mattheisen, M., Miró, X., Strohmaier, J., … Nöthen, M. M. (2011). Genome-wide association study identifies genetic variation in neurocan as a susceptibility factor for bipolar disorder. *American Journal of Human Genetics*, *88*(3), 372–81. doi:10.1016/j.ajhg.2011.01.017

Courchesne, E., Mouton, P. R., Calhoun, M. E., Semendeferi, K., Ahrens-Barbeau, C., Hallet, M. J., … Pierce, K. (2011). Neuron number and size in prefrontal cortex of children with autism. *JAMA : The Journal of the American Medical Association*, *306*(18), 2001–10. doi:10.1001/jama.2011.1638

Coyaud, E., Struski, S., Dastugue, N., Brousset, P., Broccardo, C., & Bradtke, J. (2010). PAX5-AUTS2 fusion resulting from t(7;9)(q11.2;p13.2) can now be classified as recurrent in B cell acute lymphoblastic leukemia. *Leukemia Research*, *34*(12), e323–5. doi:10.1016/j.leukres.2010.07.035

Cuscó, I., Medrano, A., & Gener, B. (2009). Autism-specific copy number variants further implicate the phosphatidylinositol signaling pathway and the glutamatergic synapse in the etiology of the disorder. *Human Molecular Genetics*, *18*(10), 1795–1804.

Denk, D., Nebral, K., Bradtke, J., Pass, G., Möricke, A., Attarbaschi, A., & Strehl, S. (2012). PAX5-AUTS2: a recurrent fusion gene in childhood B-cell precursor acute lymphoblastic leukemia. *Leukemia Research*, *36*(8), e178–81. doi:10.1016/j.leukres.2012.04.015

Dequéant, M.-L., Glynn, E., Gaudenz, K., Wahl, M., Chen, J., Mushegian, A., & Pourquié, O. (2006). A complex oscillating network of signaling genes underlies the mouse segmentation clock. *Science*, *314*(5805), 1595–8. doi:10.1126/science.1133141

Diez-Roux, G., Banfi, S., Sultan, M., Geffers, L., Anand, S., Rozado, D., … Ballabio, A. (2011). A high-resolution anatomical atlas of the transcriptome in the mouse embryo. *PLoS Biology*, *9*(1), e1000582. doi:10.1371/journal.pbio.1000582

Egaas, B., Courchesne, E., & Saitoh, O. (1995). Reduced size of corpus callosum in autism. *Archives of Neurology*, *52*. Retrieved from http://medcontent.metapress.com/index/A65RM03P4874243N.pdf

Elia, J., Gai, X., Xie, H. M., Perin, J. C., Geiger, E., Glessner, J. T., … White, P. S. (2010). Rare structural variants found in attention-deficit hyperactivity disorder are preferentially associated with neurodevelopmental genes. *Mol Psychiatry*, *15*(6), 637–46. doi:10.1038/mp.2009.57

Enoch, M.-A., Zhou, Z., Kimura, M., Mash, D. C., Yuan, Q., & Goldman, D. (2012). GABAergic gene expression in postmortem hippocampus from alcoholics and cocaine addicts; corresponding findings in alcohol-naïve P and NP rats. *PloS One*, *7*(1), e29369. doi:10.1371/journal.pone.0029369

Evan, G., & Littlewood, T. (1998). A Matter of Life and Cell Death. *Science*, *281*(August), 1317–1322.

Fatemi, S. H., Folsom, T. D., Reutiman, T. J., & Thuras, P. D. (2009). Expression of GABA(B) receptors is altered in brains of subjects with autism. *Cerebellum (London, England)*, *8*(1), 64–9. doi:10.1007/s12311-008-0075-3

Fatemi, S. H., Snow, A. V, Stary, J. M., Araghi-Niknam, M., Reutiman, T. J., Lee, S., … Pearce, D. a. (2005). Reelin signaling is impaired in autism. *Biol Psychiatry*, *57*(7), 777–87. doi:10.1016/j.biopsych.2004.12.018

Finlay, B., & Darlington, R. (1995). Linked regularities in the development and evolution of mammalian brains. *Science*, *268*(5217), 1578–1584. Retrieved from http://www.sciencemag.org/content/268/5217/1578.short

Fisher, S., Grice, E. a, Vinton, R. M., Bessling, S. L., & McCallion, A. S. (2006). Conservation of RET regulatory function from human to zebrafish without sequence similarity. *Science*, *312*(5771), 276–9. doi:10.1126/science.1124070

Flanagan-Steet, H., Fox, M. a, Meyer, D., & Sanes, J. R. (2005). Neuromuscular synapses can form in vivo by incorporation of initially aneural postsynaptic specializations. *Development*, *132*(20), 4471–81. doi:10.1242/dev.02044

Flatscher-Bader, T., Foldi, C. J., Chong, S., Whitelaw, E., Moser, R. J., Burne, T. H. J., … McGrath, J. J. (2011). Increased de novo copy number variants in the offspring of older males. *Translational Psychiatry*, *1*(8), e34. doi:10.1038/tp.2011.30

Fleischhacker, W. W., & Brooks, D. J. (2006). *Neurodevelopmental Disorders*. Berlin: Springer.

Fullwood, M. J., Liu, M. H., Pan, Y. F., Liu, J., Han, X., Bin, Y., … Liu, E. T. (2010). NIH Public Access. *Nature*, *462*(7269), 58–64. doi:10.1038/nature08497.An

Gao, Z., Zhang, J., Bonasio, R., Strino, F., Sawai, A., Parisi, F., … Reinberg, D. (2012). PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. *Molecular Cell*, *45*(3), 344–56. doi:10.1016/j.molcel.2012.01.002

Geschwind, D. H. (2009). Advances in autism. *Annual Review of Medicine*, *60*, 367–80. doi:10.1146/annurev.med.60.053107.121225

Gilbert, S. F. (2000). *Developmental Biology* (6th ed.). Sunderland: Sinauer Associates. Retrieved from http://www.ncbi.nlm.nih.gov/books/NBK10065/

Girirajan, S., Brkanac, Z., Coe, B. P., Baker, C., Vives, L., Vu, T. H., … Eichler, E. E. (2011). Relative Burden of Large CNVs on a Range of Neurodevelopmental Phenotypes. *PLoS Genet*, *7*(11), e1002334. doi:10.1371/journal.pgen.1002334

Girirajan, S., Johnson, R. L., Tassone, F., Balciuniene, J., Katiyar, N., Fox, K., … Selleck, S. B. (2013). Global increases in both common and rare copy number load associated with autism. *Human Molecular Genetics*. doi:10.1093/hmg/ddt136

Glessner, J. T., Wang, K., Cai, G., Korvatska, O., Kim, C. E., Wood, S., … Hakonarson, H. (2009). Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature*, *459*(7246), 569–73. doi:10.1038/nature07953

Gordon, L. R., Gribble, K. D., Syrett, C. M., & Granato, M. (2012). Initiation of synapse formation by Wnt-induced MuSK endocytosis. *Development*, *139*(5), 1023–33. doi:10.1242/dev.071555

Gray, P. a, Fu, H., Luo, P., Zhao, Q., Yu, J., Ferrari, A., … Ma, Q. (2004). Mouse brain organization revealed through direct genome-scale TF expression analysis. *Science (New York, N.Y.)*, *306*(5705), 2255–7. doi:10.1126/science.1104935

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., … Pääbo, S. (2010). A draft sequence of the Neandertal genome. *Science*, *328*(5979), 710–22. doi:10.1126/science.1188021

Gruszka-Westwood, A. M., Horsley, S. W., Martinez-Ramirez, A., Harrison, C. J., Kempski, H., Moorman, A. V., … Kearney, L. (2004). Comparative expressed sequence hybridization studies of high-hyperdiploid childhood acute lymphoblastic leukemia. *Genes, Chromosomes and Cancer*, *41*(3), 191–202. doi:10.1002/gcc.20085

Hamshere, M. L., Green, E. K., Jones, I. R., Jones, L., Moskvina, V., Kirov, G., … Craddock, N. (2009). Genetic utility of broadly defined bipolar schizoaffective disorder as a diagnostic concept. *The British Journal of Psychiatry : The Journal of Mental Science*, *195*(1), 23–9. doi:10.1192/bjp.bp.108.061424

Hashimoto-Torii, K., Torii, M., Sarkisian, M. R., Bartley, C. M., Shen, J., Radtke, F., … Rakic, P. (2008). Interaction between Reelin and Notch signaling regulates neuronal migration in the cerebral cortex. *Neuron*, *60*(2), 273–84. doi:10.1016/j.neuron.2008.09.026

Hattori, E., Toyota, T., Ishitsuka, Y., Iwayama, Y., Yamada, K., Ujike, H., … Yoshikawa, T. (2009). Preliminary genome-wide association study of bipolar disorder in the Japanese population. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics : The Official Publication of the International Society of Psychiatric Genetics*, *150B*(8), 1110–7. doi:10.1002/ajmg.b.30941

Holder, N., & Hill, J. (1991). Retinoic acid modifies development of the midbrain-hindbrain border and affects cranial ganglion formation in zebrafish embryos. *Development (Cambridge, England)*, *113*(4), 1159–70. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1811934

Huang, X.-L., Zou, Y. S., Maher, T. a, Newton, S., & Milunsky, J. M. (2010). A de novo balanced translocation breakpoint truncating the autism susceptibility candidate 2 (AUTS2) gene in a patient

with autism. *American Journal of Medical Genetics. Part A*, *152A*(8), 2112–4. doi:10.1002/ajmg.a.33497

Jolley, A., Corbett, M., McGregor, L., Waters, W., Brown, S., Nicholl, J., & Yu, S. (2013). De Novo Intragenic Deletion of the Autism Susceptibility Candidate 2 (AUTS2) Gene in a Patient With Developmental Delay: A Case Report and Literature Review. *American Journal of Medical Genetics. Part A*, *2*, 1–5. doi:10.1002/ajmg.a.35922

Kalscheuer, V. M., FitzPatrick, D., Tommerup, N., Bugge, M., Niebuhr, E., Neumann, L. M., … Ullmann, R. (2007). Mutations in autism susceptibility candidate 2 (AUTS2) in patients with mental retardation. *Hum Genet*, *121*(3-4), 501–9. doi:10.1007/s00439-006-0284-0

Kawakami, K. (2005). Transposon tools and methods in zebrafish. *Dev Dyn*, *234*(2), 244–54. doi:10.1002/dvdy.20516

Kawamata, N., Ogawa, S., Zimmermann, M., Niebuhr, B., Stocking, C., Sanada, M., … Koeffler, H. P. (2008). Cloning of genes involved in chromosomal translocations by high-resolution single nucleotide polymorphism genomic microarray. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(33), 11921–6. doi:10.1073/pnas.0711039105

Kelliher, M. a, Seldin, D. C., & Leder, P. (1996). Tal-1 induces T cell acute lymphoblastic leukemia accelerated by casein kinase IIalpha. *The EMBO Journal*, *15*(19), 5160–6. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=452259&tool=pmcentrez&rendertype=abstract

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., & Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*, *14*(4), R36. doi:10.1186/gb-2013-14-4-r36

Kondrashov, N., Pusic, A., Stumpf, C. R., Shimizu, K., Hsieh, A. C., Xue, S., … Barna, M. (2011). Ribosome-mediated specificity in Hox mRNA translation and vertebrate tissue patterning. *Cell*, *145*(3), 383–97. doi:10.1016/j.cell.2011.03.028

Kothary R., Clapoff S., Brown A., Campbell R., P. A. & R. J. (1988). A transgene containing lacZ inserted into the dystonia locus is expressed in neural tube. *Nature*, *335*, 435–437.

Kulakovskiy, I. V, Boeva, V. a, Favorov, a V, & Makeev, V. J. (2010). Deep and wide digging for binding motifs in ChIP-Seq data. *Bioinformatics (Oxford, England)*, *26*(20), 2622–3. doi:10.1093/bioinformatics/btq488

Lainhart, J., & Lange, N. (2011). Increased neuron number and head size in autism. *JAMA: The Journal of the American Medical …*, *306*(18), 2031–2032. doi:10.1093/cercor/bhr018.10.

Lamond, A. I., & Spector, D. L. (2003). Nuclear speckles: a model for nuclear organelles. *Nature Reviews. Molecular Cell Biology*, *4*(8), 605–12. doi:10.1038/nrm1172

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, *10*(3), R25. doi:10.1186/gb-2009-10-3-r25

Lee, H., Woo, H., & Greenwood, T. (2012). A genome-wide association study of seasonal pattern mania identifies NF1A as a possible susceptibility gene for bipolar disorder. *Journal of Affective Disorders*, *145*, 200–207. Retrieved from http://www.sciencedirect.com/science/article/pii/S0165032712005630

Lener, T., Moll, P. R., Rinnerthaler, M., Bauer, J., Aberger, F., & Richter, K. (2006). Expression profiling of aging in the human skin. *Experimental Gerontology*, *41*(4), 387–97. doi:10.1016/j.exger.2006.01.012

Lepagnol-Bestel, a-M., Maussion, G., Boda, B., Cardona, a, Iwayama, Y., Delezoide, a-L., … Simonneau, M. (2008). SLC25A12 expression is associated with neurite outgrowth and is upregulated in the prefrontal cortex of autistic subjects. *Molecular Psychiatry*, *13*(4), 385–97. doi:10.1038/sj.mp.4002120

Li, Q., Ritter, D., Yang, N., Dong, Z., Li, H., Chuang, J. H., & Guo, S. (2010). A systematic approach to identify functional motifs within vertebrate developmental enhancers. *Dev Biol*, *337*(2), 484–95. doi:10.1016/j.ydbio.2009.10.019

Li, R., Brockschmidt, F. F., Kiefer, A. K., Stefansson, H., Nyholt, D. R., Song, K., … Richards, J. B. (2012). Six novel susceptibility Loci for early-onset androgenetic alopecia and their unexpected association with common diseases. *PLoS Genetics*, *8*(5), e1002746. doi:10.1371/journal.pgen.1002746

Liao, D., Cheng, M., & Lai, C. (2011). Comparative gene expression profiling analysis of lymphoblastoid cells reveals neuron-□specific enolase gene (ENO2) as a susceptibility gene of heroin dependence. *Addiction Biology*.

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., … Dekker, J. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, *326*(5950), 289–93. doi:10.1126/science.1181369

Lin, A., Wang, R. T., Ahn, S., Park, C. C., & Smith, D. J. (2010). A genome-wide map of human genetic interactions inferred from radiation hybrid genotypes. *Genome Research*, *20*(8), 1122–32. doi:10.1101/gr.104216.109

Luciano, M., Hansell, N. K., Lahti, J., Davies, G., & Medland, S. E. (2012). UKPMC Funders Group Whole genome association scan for genetic polymorphisms influencing information processing speed. *Biol Psychol*, *86*(3), 193–202. doi:10.1016/j.biopsycho.2010.11.008.Whole

Luciano, M., Hansell, N., & Lahti, J. (2011). Whole genome association scan for genetic polymorphisms influencing information processing speed. *Biological Psychology*, *86*(3), 193–202. doi:10.1016/j.biopsycho.2010.11.008.Whole

Machanick, P., & Bailey, T. L. (2011). MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics (Oxford, England)*, *27*(12), 1696–7. doi:10.1093/bioinformatics/btr189

Maraganore, D. M., Lesnick, T. G., Elbaz, A., Chartier-Harlin, M.-C., Gasser, T., Krüger, R., … Toda, T. (2004). UCHL1 is a Parkinson's disease susceptibility gene. *Annals of Neurology*, *55*(4), 512–21. doi:10.1002/ana.20017

Martens, M. a, Wilson, S. J., & Reutens, D. C. (2008). Research Review: Williams syndrome: a critical review of the cognitive, behavioral, and neuroanatomical phenotype. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *49*(6), 576–608. doi:10.1111/j.1469-7610.2008.01887.x

Massari, M., & Murre, C. (2000). Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. *Molecular and Cellular Biology*, *20*(2), 429–440. Retrieved from http://mcb.asm.org/content/20/2/429.short

Mcgaughey, D. M., Vinton, R. M., Huynh, J., Al-saif, A., Beer, M. A., & Mccallion, A. S. (2008). Metrics of sequence constraint overlook regulatory sequences in an exhaustive analysis at phox2b. *Genome Res.*, *18*(2), 252–260. doi:10.1101/gr.6929408.1

McLean, C. Y., Bristor, D., Hiller, M., Clarke, S. L., Schaar, B. T., Lowe, C. B., … Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. *Nature Biotechnology*, *28*(5), 495–501. doi:10.1038/nbt.1630

Mefford, H. C., Muhle, H., Ostertag, P., von Spiczak, S., Buysse, K., Baker, C., … Eichler, E. E. (2010). Genome-wide copy number variation in epilepsy: novel susceptibility loci in idiopathic generalized and focal epilepsies. *PLoS Genet*, *6*(5), e1000962. doi:10.1371/journal.pgen.1000962

Missler, M., & Südhof, T. (1998). Neurexins: three genes and 1001 products. *Trends in Genetics*, *14*(1), 20–26. Retrieved from http://www.sciencedirect.com/science/article/pii/S0168952597013243

Molnar, C., & de Celis, J. F. (2013). Tay Bridge Is a Negative Regulator of EGFR Signalling and Interacts with Erk and Mkp3 in the Drosophila melanogaster Wing. *PLoS Genetics*, *9*(12), e1003982. doi:10.1371/journal.pgen.1003982

Mühleisen, T. W., Mattheisen, M., Strohmaier, J., Degenhardt, F., Priebe, L., Schultz, C. C., … Cichon, S. (2012). Association between schizophrenia and common variation in neurocan (NCAN), a genetic risk factor for bipolar disorder. *Schizophrenia Research*, *138*(1), 69–73. doi:10.1016/j.schres.2012.03.007

Nagamani, S. C. S., Erez, A., Ben-Zeev, B., Frydman, M., Winter, S., Zeller, R., … Wai Cheung, S. (2013). Detection of copy-number variation in AUTS2 gene by targeted exonic array CGH in patients with developmental delay and autistic spectrum disorders. *European Journal of Human Genetics : EJHG*, *21*(3), 1–4. doi:10.1038/ejhg.2012.157

Nagy A., Gertsenstein M. , Vintersten K., B. R. (2003). *Manipulating the Mouse Embryo: A Laboratory Manual (Third Edition)*. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.

Nasevicius, a, & Ekker, S. C. (2000). Effective targeted gene “knockdown” in zebrafish. *Nat Genet*, *26*(2), 216–20. doi:10.1038/79951

Navratilova, P., Fredman, D., Hawkins, T. a, Turner, K., Lenhard, B., & Becker, T. S. (2009). Systematic human/zebrafish comparative identification of cis-regulatory activity around vertebrate developmental transcription factor genes. *Dev Biol*, *327*(2), 526–40. doi:10.1016/j.ydbio.2008.10.044

Nusslein-Volhard, C. and R. D. (2002). *Zebrafish*. Oxford: Oxford University Press.

O’Leary, J., & Osborne, L. R. (2011). Global analysis of gene expression in the developing brain of Gtf2ird1 knockout mice. *PloS One*, *6*(8), e23868. doi:10.1371/journal.pone.0023868

O’Roak, B. J., Deriziotis, P., Lee, C., Vives, L., Schwartz, J. J., Girirajan, S., … Eichler, E. E. (2011). Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nature Genetics*, *43*(6), 585–9. doi:10.1038/ng.835

O’Roak, B. J., Vives, L., Girirajan, S., Karakoc, E., Krumm, N., Coe, B. P., … Eichler, E. E. (2012). Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature*, *485*(7397), 246–50. doi:10.1038/nature10989

Oksenberg, N., & Ahituv, N. (2013). The role of AUTS2 in neurodevelopment and human evolution. *Trends in Genetics : TIG*, 1–9. doi:10.1016/j.tig.2013.08.001

Oksenberg, N., Stevison, L., Wall, J. D., & Ahituv, N. (2013). Function and regulation of AUTS2, a gene implicated in autism and human evolution. *PLoS Genetics*, *9*(1), e1003221. doi:10.1371/journal.pgen.1003221

Pardo, C. a, & Eberhart, C. G. (2007). The neurobiology of autism. *Brain Pathol*, *17*(4), 434–47. doi:10.1111/j.1750-3639.2007.00102.x

Park, H. C., Kim, C. H., Bae, Y. K., Yeo, S. Y., Kim, S. H., Hong, S. K., … Huh, T. L. (2000). Analysis of upstream elements in the HuC promoter leads to the establishment of transgenic zebrafish with fluorescent neurons. *Dev Biol*, *227*(2), 279–93. doi:10.1006/dbio.2000.9898

Pennacchio, L. a, Ahituv, N., Moses, A. M., Prabhakar, S., Nobrega, M. a, Shoukry, M., … Rubin, E. M. (2006). In vivo enhancer analysis of human conserved non-coding sequences. *Nature*, *444*(7118), 499–502. doi:10.1038/nature05295

Penney, K. L., Pyne, S., Schumacher, F. R., Sinnott, J. a, Mucci, L. a, Kraft, P. L., … Freedman, M. L. (2010). Genome-wide association study of prostate cancer mortality. *Cancer Epidemiology, Biomarkers & Prevention : A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, *19*(11), 2869–76. doi:10.1158/1055-9965.EPI-10-0601

Philibert, R. a, Ryu, G.-Y., Yoon, J.-G., Sandhu, H., Hollenbeck, N., Gunter, T., … Madan, A. (2007). Transcriptional profiling of subjects from the Iowa adoption studies. *American Journal of Medical*

Genetics. Part B, Neuropsychiatric Genetics : The Official Publication of the International Society of Psychiatric Genetics, *144B*(5), 683–90. doi:10.1002/ajmg.b.30512

Philippi, A., Tores, F., Carayol, J., Rousseau, F., Letexier, M., Roschmann, E., … Hager, J. (2007). Association of autism with polymorphisms in the paired-like homeodomain transcription factor 1 (PITX1) on chromosome 5q31: a candidate gene analysis. *BMC Medical Genetics*, *8*, 74. doi:10.1186/1471-2350-8-74

Pinto, D., Pagnamenta, A. T., Klei, L., Anney, R., Merico, D., Regan, R., … Betancur, C. (2010). Functional impact of global rare copy number variation in autism spectrum disorders. *Nature*, *466*(7304), 368–72. doi:10.1038/nature09146

Poeck, B., Triphan, T., Neuser, K., & Strauss, R. (2008). Locomotor control by the central complex in Drosophila-An analysis of the tay bridge mutant. *Developmental Neurobiology*, *68*(8), 1046–58. doi:10.1002/dneu.20643

Pollard, K. S., Salama, S. R., King, B., Kern, A. D., Dreszer, T., Katzman, S., … Haussler, D. (2006). Forces shaping the fastest evolving regions in the human genome. *PLoS Genetics*, *2*(10), e168. doi:10.1371/journal.pgen.0020168

Pollard, K. S., Salama, S. R., Lambert, N., Lambot, M.-A., Coppens, S., Pedersen, J. S., … Haussler, D. (2006). An RNA gene expressed during cortical development evolved rapidly in humans. *Nature*, *443*(7108), 167–72. doi:10.1038/nature05113

Poot, M., Smagt, J. J. Van Der, Brilstra, E. H., Bourgeron, T., & van der Smagt, J. J. (2011). Disentangling the myriad genomics of complex disorders, specifically focusing on autism, epilepsy, and schizophrenia. *Cytogenetic and Genome Research*, *135*(3-4), 228–40. doi:10.1159/000334064

Prabhakar, S., Noonan, J. P., Pääbo, S., & Rubin, E. M. (2006). Accelerated evolution of conserved noncoding sequences in humans. *Science*, *314*(5800), 786. doi:10.1126/science.1130738

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)*, *26*(6), 841–2. doi:10.1093/bioinformatics/btq033

Redon, R., Ishikawa, S., Fitch, K. R., Feuk, L., Perry, G. H., Andrews, T. D., … Hurles, M. E. (2006). Global variation in copy number in the human genome. *Nature*, *444*(7118), 444–54. doi:10.1038/nature05329

Risch, N., Spiker, D., Lotspeich, L., Nouri, N., Hinds, D., Hallmayer, J., … Myers, R. M. (1999). A Genomic Screen of Autism : Evidence for a Multilocus Etiology. *Amer J Hum Genet*, *65*(3), 931.

Salichs, E., Ledda, A., Mularoni, L., Albà, M. M., & de la Luna, S. (2009). Genome-wide analysis of histidine repeats reveals their role in the localization of human proteins to the nuclear speckles compartment. *PLoS Genetics*, *5*(3), e1000397. doi:10.1371/journal.pgen.1000397

Sanders, S. J., Murtha, M. T., Gupta, A. R., Murdoch, J. D., Raubeson, M. J., Willsey, a J., … State, M. W. (2012). De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature*, *485*(7397), 237–41. doi:10.1038/nature10945

Sato, S., Hayashi, T., & Kobayashi, E. (2011). Characterization of porcine autism susceptibility candidate 2 as a candidate gene for the number of corpora lutea in pigs. *Animal Reproduction Science*, *126*(3-4), 211–20. doi:10.1016/j.anireprosci.2011.05.003

Schumann, G., Coin, L. J., Lourdusamy, A., Charoen, P., Berger, K. H., Stacey, D., … Elliott, P. (2011). Genome-wide association and genetic functional studies identify autism susceptibility candidate 2 gene (AUTS2) in the regulation of alcohol consumption. *Proc Natl Acad Sci U S A*, *108*(17), 7119–24. doi:10.1073/pnas.1017288108

Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., … Wigler, M. (2007). Strong association of de novo copy number mutations with autism. *Science*, *316*(5823), 445–9. doi:10.1126/science.1138659

Sedaghat, Y., Mazur, C., Sabripour, M., Hung, G., & Monia, B. P. (2012). Genomic analysis of wig-1 pathways. *PloS One*, *7*(2), e29429. doi:10.1371/journal.pone.0029429

Shen, Y., Yue, F., McCleary, D. F., Ye, Z., Edsall, L., Kuan, S., … Ren, B. (2012). A map of the cis-regulatory sequences in the mouse genome. *Nature*, *488*(7409), 116–20. doi:10.1038/nature11243

Siepel, A., Bejerano, G., Pedersen, J. S., Hinrichs, A. S., Hou, M., Rosenbloom, K., … Haussler, D. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Research*, *15*(8), 1034–50. doi:10.1101/gr.3715005

Slattery, C., Ryan, M. P., & McMorrow, T. (2008). E2A proteins: regulators of cell phenotype in normal physiology and disease. *The International Journal of Biochemistry & Cell Biology*, *40*(8), 1431–6. doi:10.1016/j.biocel.2007.05.014

Srinivasan, K., Leone, D. P., Bateson, R. K., Dobreva, G., Kohwi, Y., Kohwi-Shigematsu, T., … McConnell, S. K. (2012). A network of genetic repression and derepression specifies projection fates in the developing neocortex. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(47), 19071–8. doi:10.1073/pnas.1216793109

Stadler, Z. K., Esposito, D., Shah, S., Vijai, J., Yamrom, B., Levy, D., … Offit, K. (2012). Rare de novo germline copy-number variation in testicular cancer. *American Journal of Human Genetics*, *91*(2), 379–83. doi:10.1016/j.ajhg.2012.06.019

Stein, J. L., Parikshak, N. N., & Geschwind, D. H. (2013). Rare inherited variation in autism: beginning to see the forest and a few trees. *Neuron*, *77*(2), 209–11. doi:10.1016/j.neuron.2013.01.010

Sultana, R., Yu, C.-E., Yu, J., Munson, J., Chen, D., Hua, W., … Villacres, E. C. (2002). Identification of a Novel Gene on Chromosome 7q11.2 Interrupted by a Translocation Breakpoint in a Pair of Autistic Twins. *Genomics*, *80*(2), 129–134. doi:10.1006/geno.2002.6810

Talkowski, M. E., Minikel, E. V., & Gusella, J. F. (2014). Autism spectrum disorder genetics: diverse genes with diverse clinical outcomes. *Harvard Review of Psychiatry*, *22*(2), 65–75. doi:10.1097/HRP.0000000000000002

Talkowski, M. E., Rosenfeld, J. A., Blumenthal, I., Pillalamarri, V., Chiang, C., Heilbut, A., … Gusella, J. F. F. (2012). Sequencing chromosomal abnormalities reveals neurodevelopmental loci that confer risk across diagnostic boundaries. *Cell*, *149*(3), 525–537. doi:10.1016/j.cell.2012.03.028

Thisse, B., Heyer, V., Lux, A., Alunni, V., Degrave, A., Seiliez, I., … Thisse, C. (2004). Spatial and temporal expression of the zebrafish genome by large-scale in situ hybridization screening. *Methods Cell Biol*, *77*, 505–19. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15602929

Trapnell, C., Pachter, L., & Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics (Oxford, England)*, *25*(9), 1105–11. doi:10.1093/bioinformatics/btp120

Tropeano, M., Ahn, J. W., Dobson, R. J. B., Breen, G., Rucker, J., Dixit, A., … Collier, D. a. (2013). Male-Biased Autosomal Effect of 16p13.11 Copy Number Variation in Neurodevelopmental Disorders. *PLoS ONE*, *8*(4), e61365. doi:10.1371/journal.pone.0061365

Tropepe, V., & Sive, H. L. (2003). Can zebrafish be used as a model to study the neurodevelopmental causes of autism? *Genes Brain Behav*, 268–281. doi:10.1046/j.1601-183X.2003.00038.x

Visel, A., Blow, M. J., Li, Z., Zhang, T., Akiyama, J. a, Holt, A., … Pennacchio, L. a. (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*, *457*(7231), 854–8. doi:10.1038/nature07730

Visel, A., Minovitsky, S., Dubchak, I., & Pennacchio, L. a. (2007). VISTA Enhancer Browser--a database of tissue-specific human enhancers. *Nucleic Acids Res*, *35*(Database issue), D88–92. doi:10.1093/nar/gkl822

Visel, A., Thaller, C., & Eichele, G. (2004). GenePaint.org: an atlas of gene expression patterns in the mouse embryo. *Nucleic Acids Res*, *32*(Database issue), D552–6. doi:10.1093/nar/gkh029

Wagner, A. H., Anand, N., Wang, W.-H., Chatterton, J. E., Sun, D., Shepard, A., … Clark, A. F. (2013). Exon-level expression profiling of ocular tissues. *Experimental Eye Research*, (March), 1–7. doi:10.1016/j.exer.2013.03.004

Wang, G.-S., Hong, C.-J., Yen, T.-Y., Huang, H.-Y., Ou, Y., Huang, T.-N., … Hsueh, Y.-P. (2004). Transcriptional modification by a CASK-interacting nucleosome assembly protein. *Neuron*, *42*(1), 113–28. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15066269

Watanabe, E., Aono, S., Matsui, F., Yamada, Y., Naruse, I., & Oohira, a. (1995). Distribution of a brain-specific proteoglycan, neurocan, and the corresponding mRNA during the formation of barrels in the rat somatosensory cortex. *The European Journal of Neuroscience*, *7*(4), 547–54. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7620606

Weir, B., Woo, M., Getz, G., & Perner, S. (2007). Characterizing the cancer genome in lung adenocarcinoma. *Nature*, *450*(7171), 893–898. doi:10.1038/nature06358.Characterizing

Wen, R.-T., Zhang, M., Qin, W.-J., Liu, Q., Wang, W.-P., Lawrence, A. J., … Liang, J.-H. (2012). The phosphodiesterase-4 (PDE4) inhibitor rolipram decreases ethanol seeking and consumption in alcohol-preferring Fawn-Hooded rats. *Alcoholism, Clinical and Experimental Research*, *36*(12), 2157–67. doi:10.1111/j.1530-0277.2012.01845.x

Westerfield, M. (2007). *The Zebrafish Book* (5th ed.). Eugene, Oregon: University of Oregon Press.

White, R. M., Sessa, A., Burke, C., Bowman, T., LeBlanc, J., Ceol, C., … Zon, L. I. (2008). Transparent adult zebrafish as a tool for in vivo transplantation analysis. *Cell Stem Cell*, *2*(2), 183–9. doi:10.1016/j.stem.2007.11.002

William, D. a, Saitta, B., Gibson, J. D., Traas, J., Markov, V., Gonzalez, D. M., … Kusumi, K. (2007). Identification of oscillatory genes in somitogenesis from functional genomic analysis of a human mesenchymal stem cell model. *Developmental Biology*, *305*(1), 172–86. doi:10.1016/j.ydbio.2007.02.007

Wong, C., Meaburn, E., Ronald, A., Price, T., Jeffries, a R., Schalkwyk, L., … Mill, J. (2013). Methylomic analysis of monozygotic twins discordant for autism spectrum disorder and related behavioural traits. *Molecular Psychiatry*, (November 2012), 1–9. doi:10.1038/mp.2013.41

Xi, B., Chen, J., Yang, L., Wang, W., Fu, M., & Wang, C. (2011). GABBR1 gene polymorphism (G1465A) is associated with temporal lobe epilepsy. *Epilepsy Research*, *96*(1-2), 58–63. doi:10.1016/j.eplepsyres.2011.04.014

Xie, Q., Hao, Y., Tao, L., Peng, S., Rao, C., Chen, H., … Yuan, Z. (2012). Lysine methylation of FOXO3 regulates oxidative stress-induced neuronal cell death. *EMBO Reports*, *13*(4), 371–7. doi:10.1038/embor.2012.25

Yang, W., Liu, J., Zheng, F., Jia, M., Zhao, L., Lu, T., … Wang, L. (2013). The evidence for association of ATP2B2 polymorphisms with autism in Chinese Han population. *PloS One*, *8*(4), e61021. doi:10.1371/journal.pone.0061021

Young, E. J., Lipina, T., Tam, E., Mandel, a, Clapcote, S. J., Bechard, a R., … Osborne, L. R. (2008). Reduced fear and aggression and altered serotonin metabolism in Gtf2ird1-targeted mice. *Genes, Brain, and Behavior*, *7*(2), 224–34. doi:10.1111/j.1601-183X.2007.00343.x

Zhang, Y. E., Landback, P., Vibranovski, M. D., & Long, M. (2011). Accelerated Recruitment of New Brain Development Genes into the Human Genome. *PLoS Biol*, *9*(10), e1001179. doi:10.1371/journal.pbio.1001179

Zhang, Y., Liu, T., Meyer, C. a, Eeckhoute, J., Johnson, D. S., Bernstein, B. E., … Liu, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biology*, *9*(9), R137. doi:10.1186/gb-2008-9-9-r137

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

Author Signature

3-24-2014
Date