Summer 8-15-2019

Determining the Genetic Contributions of the Williams Syndrome Critical Region to Behavior Using Mouse Models and Human Genetics

Nathan David Kopp
Washington University in St. Louis

Recommended Citation

Kopp, Nathan David, "Determining the Genetic Contributions of the Williams Syndrome Critical Region to Behavior Using Mouse Models and Human Genetics" (2019). *Arts & Sciences Electronic Theses and Dissertations*. 1919.
https://openscholarship.wustl.edu/art_sci_etds/1919

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology and Biomedical Sciences
Human and Statistical Genetics

Dissertation Examination Committee:
Joseph Dougherty, Chair
Yehuda Ben-Shahar
Don Conrad
Harrison Gabel
Christina Gurnett
Beth Kozel

Determining the Genetic Contributions of the Williams Syndrome Critical Region to Behavior Using Mouse Models and Human Genetics

by
Nathan Kopp

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

August 2019
St. Louis, Missouri

Table of Contents

List of Figures

Chapter 2

Supplemental Figures

Chapter 3

Supplemental Figures

Chapter 4

Supplemental Figures

List of Tables

Chapter 2

Supplemental Tables

Chapter 3

Supplemental Tables

Chapter 4

Supplemental Tables

Acknowledgments

Many people have helped me achieve the work that is included in this thesis and I could not have done it without them. First, Dr. Joe Dougherty provided amazing support and mentorship. He helped guide me and taught me how to be a better scientist. Joe fostered an amazing lab culture where we worked hard but also had a good time. There are many people in the Dougherty Lab that also contributed to my scientific growth and my well-being while in the lab. I would especially like to thank Michael Rieger, Kristina Sakers, and Susan Maloney. They each taught me valuable lessons and always made sure I was producing good science while having a good time.

I have made many friends throughout graduate school that have encouraged me and pushed me. My Human and Statistical Genetics Cohort included Matt Bailey, Beth Ostrander, and Renee Sear. We all helped each other and our relationships made working hard in graduate school more fun. I was also fortunate enough to find people that enjoyed eating lunch with every day so we could talk about our work, but also utilize that time to take our minds off of what we were struggling with. I also want to thank the Open Reading Frames book club. It was great to share my thoughts on a wide variety books with such a great like-minded yet diverse group of people.

I would like to thank my family for all the support and love they have giving me. My Dad has been an unwavering example of how patience and hard work can pay off. My mother has taught me to be fearless and confident. My three older siblings have helped me mature and also made sure I was having a good time. My family's successes have always motivated me to be the best I can.

Finally, I need to thank Gregory Scheetz Jr. He has kept me smiling and as calm and optimistic as possible throughout graduate school. Throughout the late nights, weekend experiments, and the waiting in the car while I run into the mouse house, Greg has been there and encouraged me. He is a great caregiver not only to me, but our two kitties Atticus and Ophelia, who I also must thank for their cuddles and keeping me happy. Greg has helped me reach my greatest potential and for that I am very grateful

Nathan Kopp

Washington University in St. Louis

August 2019

Dedicated to my family and Gregory Scheetz Jr.

ABSTRACT OF THE DISSERTATION

Determining the Genetic Contributions of the Williams Syndrome Critical Region to Behavior Using Mouse Models and Human Genetics

by

Nathan Kopp

Doctor of Philosophy in Biology and Biomedical Sciences

Human and Statistical Genetics

Washington University in St. Louis, 2019

Professor Joseph D. Dougherty, Ph.D. Chair

Williams syndrome is a neurodevelopmental model caused by the deletion of 26-28 genes on chr7q11.23. The loss of these genes affects multiple organ systems resulting in severe cardiovascular disease, craniofacial dysmorphology, intellectual impairment, a specific Williams syndrome cognitive profile made up of deficits in visual-spatial processing with preserved language skills, and a characteristic hypersocial personality. The reciprocal duplication occurs at a lower frequency and manifests with diametric phenotypes to the deletion. This suggests that this locus harbors dosage sensitive genes that play a role in neurodevelopment. Large efforts have been taken to identify which genes are responsible for causing the different aspects of the disorder. Only the cardiovascular phenotype has been linked to the hemizgosity of the *ELN* gene. In order to incorporate the complexity of genetic contributions to complex traits, we synthesize genetic and behavioral analyses in both humans and mouse models. We performed whole exome sequencing on 85 individuals with Williams syndrome to test the hypothesis that genetic variation on the remaining chr7q11.23 allele contributes to variation in the social phenotype. We show that the social phenotype consists of deficits in several aspects of social behavior, but social motivation is preserved in Williams syndrome. Whole exome sequencing revealed that there is little common variation contribution to the variability of the social phenotype but did suggest involvement of SNPs in the *BAZ1B* and *GTF2IRD1* genes. Using mouse models, we generated three new mouse lines to test the hypothesis that two genes in the syntenic region, *Gtf2i* and *Gtf2ird1*, share overlapping DNA targets and both contribute to overlapping behavioral phenotypes suggesting an oligogenic contribution of these genes to phenotypes relevant to WS. Finally, we show that loss of function mutations in both *Gtf2i* and *Gtf2ird1* are not sufficient to reproduce the full phenotype that is produced by deleting the entire syntenic Williams syndrome critical region in mice. Taken together these data suggest an oligogenic pattern of contribution to the phenotypes seen in WS.

Chapter 1: Introduction

The aim of human genetics is to identify the genes that contribute to human biology. This approach will allow us to study the underlying mechanisms that manifest as interesting human phenotypes, such as our complex central nervous system, which gives rise to many diverse behaviors. Geneticists have developed and employed many approaches to elucidate genes that are important for specific human traits. These include linkage analysis, genome-wide association, whole-exome sequencing studies, and whole-genome sequencing studies. These tools have driven the progress of genotype-phenotype correlations and resulted in many important discoveries.

Along with sophisticated approaches, human genetics has been informed by identifying genes that cause human diseases. The underlying genetic causes of the disorder highlights the functional pathways in which the causal gene plays an important role. From these natural experiments the genetic search space is narrowed from the 3 billion base pairs that make up the human genome to a specific gene that can then be studied at different levels of genomic and biological organization. Some disorders are not caused by the disruption of one gene, but by a change in the dosage of many contiguous genes. These copy number variation disorders point to a region in the genome that affect multiple aspects of human development, such as neurodevelopment, cardiovascular development, and craniofacial development. However, copy number disorders offer a unique challenge, because while they emphasize the importance of a specific genomic region, there are still many genes and many phenotypes to disentangle. The question then becomes, which genes in the region are responsible for causing the specific phenotypes seen in the disorder.

Williams-Beuren syndrome (WS) is one such copy number variation disorder. It is caused by the deletion of chromosome 7q11.23, referred to as the Williams syndrome critical region (WSCR), and results in a constellation of phenotypes that include cardiovascular disease, craniofacial dysmorphology, a specific cognitive profile, and a characteristic hypersocial personality (1, 2). There are 26-28 genes that are commonly deleted in WS. Large efforts have been put forth to connect specific genes in the region to specific phenotypes in the syndrome. The only substantiated monogenic contribution of a causal gene in the WSCR is to the cardiovascular phenotype driven by the elastin gene (*ELN*) (3), leaving much more work to be done to understand how the genes in this region affect complex phenotypes such as cognition and social behavior.

The research presented in this thesis uses both human genetic techniques as well as mouse models to dissect the effect of genes in the WSCR on different aspects of behavior. I analyzed the whole-exome sequences of 85 individuals with WS to test if variation on the remaining chr7q11.23 allele, as well as exome-wide variation, contributes to the social phenotype, providing the largest genetic analysis of individuals with WS. I have also leveraged the experimental advantages of the mouse model organism to ask how two genes in the WSCR, *Gtf2i* and *Gtf2ird1*, interact in the developing mouse brain. I go on to show that in the mouse, these genes are not sufficient to produce the behavioral and transcriptional phenotypes of the full deletion. I have tested several longstanding hypotheses in the field of Williams syndrome genetics through my experiments and provide evidence that the genetic risk for the phenotypes observed in WS are not solely driven by these two transcription factors.

1.1 History and description of Williams syndrome

Williams-Beuren syndrome (OMIM #194050) was first recognized as a syndrome by two physicians in the early 1960's. First, in 1961 Williams *et al*. described four cases of children that were being treated for supravalvular aortic stenosis (SVAS). Williams observed that the children were "mentally deficient" and had similar facial features. He thought the similarities could be a part of a previously unrecognized syndrome (4). In 1962 Bueren *et al.* described three more patients that had SVAS, intellectual disability, and craniofacial features that were remarkably similar to the patients described by Williams *et al.* Beuren mentioned that all the children had a friendly nature and "loved everyone" (5). This observation is the first description of the gregarious personality that is now recognized as a hallmark of WS.

Since the association between the cardiovascular disease, intellectual disability, and craniofacial features made by Williams and Beuren, the genetic etiology of WS has been well-defined. The *ELN* gene on chromosome seven was discovered to be the cause of familial SVAS , in a linkage analysis of one kindred (6). Subsequently, it was shown that individuals with WS were hemizygous for the *ELN* gene and that the hemizygosity extended beyond the *ELN* locus, suggesting that WS is caused by a contiguous deletion on chromosome seven (3). These findings lead to the use of *ELN* FISH probes as the first clinical genetic test for WS (7). Using artificial chromosomes the 1.5Mbp region on chromosome seven that is deleted in WS has been delineated (8–10). The region contains 26 genes that are commonly deleted and two more genes that are deleted in the longer 1.8Mbp version of the deletion. The WSCR was found to be demarcated by three regions of low copy repeats: the centromeric, medial, and distal regions (9, 11). Within each region there are three blocks that consist of repeated genes. Block A contains the three pseudogenes of the *STAG3* gene, *PMS2L*, and *GATS*. The medial block B contains the functional genes *GTF2I*, *NCF1*, and *GTF2IRD2*, where the centromeric and telomeric block B contains the corresponding pseudogenes. Block C contains *POM121*, *NSUN5*, and *TRIM50* (11). The low copy repeat blocks themselves are demarcated by Alu repeats. These low copy repeats facilitate non homologous allelic recombination (NHAR), which leads to the recurrent deletion and duplication of the region (12). The most common 1.5Mbp deletion, which occurs in about 95% of cases, is caused by misalignement of the B centromeric and B medial blocks, which have 99.6% sequence identity. The less common larger 1.8Mbp deletion, with a prevalence of 3-5% of cases, occurs by the misalignment of the A centeromeric and A medial blocks, which have 98.2% sequence identity (13). This well-defined and common genetic cause of most cases of WS makes studying this disorder an excellent opportunity to make genotype-phenotype correlations.

Along with the well-characterized genetic cause of WS, the phenotypic spectrum of the constellation of symptoms in WS has been thoroughly described and reviewed by many researchers (1, 2, 13–16). The cardiovascular disease in WS manifests as SVAS as well as other focal artery stenoses and affects all elastic vessels. Other issues also relate to connective tissues such as lax skin and join hypermobility have been attributed to *ELN* haploinsufficiency. The facial dysmorphology consists of periorbital fullness, long philtrum, full lips, stellate irises, low nasal bridge, micrognathia, microcephaly, and dental problems. The deletion also affects the endocrine system and results in precocious puberty, subclinical hypothyroidism, and an increased prevalence of diabetes milletus. Neurological symptoms include poor balance and coordination, hypotonia, and hyperacusis.

Of particular interest to this thesis are the cognitive and behavioral phenotypes of WS. The deletion of the WSCR has a specific effect on cognition and this gestalt is termed the Williams syndrome cognitive profile (WSCP). Individuals with WS have a wide range of intellectual ability as measured by different tests for intelligence quotient (IQ). IQ scores span from severe intellectual disability (ID) to average scores of IQ. Despite overall lower levels of IQ the WSCP consists of relative strengths in auditory rote memory and verbal skills coupled with impairment in visual spatial construction. The definition of the WSCP was standardized by Mervis *et al.* (17). Along with a specific cognitive profile, WS is associated with a characteristic hypersocial personality (14). The social aspect of WS consists of increased attention to faces. Eye tracking studies have shown that individuals with WS fixate on eyes for longer periods of time compared to typically developing children (18). In observational studies, children with WS tend to focus on the experimenter rather than toys (14). Individuals with WS are more likely to approach strangers and have overall increased global sociability as measured by the Salk Institute Sociability Questionnaire (19). While there are prosocial aspects to the hypersocial phenotype of WS, it also consists of a maladaptive component. Individuals with WS have difficulties in social cognition and responding appropriately in social situations (20). Beyond differences in sociality, individuals with WS have other psychiatric comorbidities, that include anxiety, specifically non social anxiety, phobias, and attention deficit/hyperactivity disorder (ADHD) (21, 22). Thus, the constellation of symptoms that make up WS gives geneticists a unique window into the genetic underpinnings of many different aspects of human cognition and behavior.

The presence of the low copy repeats that are responsible for the recurrent deletion of the WSCR should also predispose the region to duplications. The first case of an individual with the duplication was described in 2005 (23). The duplication of the region results in dup7q11.23 syndrome (OMIM #609757). The symptoms of 7q11.23 have been described by Mervis and Morris (24, 25). The phenotypes are generally more mild than in the deletion of the region. Mild craniofacial dysmorphology has been reported but it is not as consistent as in WS. There are some cases that have cardiovascular anomalies that present as dilated blood vessels. However, the most consistent phenotype of the duplication is language delay. The duplication has been associated with autism spectrum disorders (ASD) (26), but in a rigorous study of ASD symptomology in 7q11.23 dup syndrome and WS, it was found there is a similar prevalence of ASD diagnosis in both disorders (27, 28). However, in contrast to the social fearlessness in WS, it was reported that there is a higher proportion of children with the duplication that have separation anxiety (29). The observation of some diametric phenotypes in 7q11.23dup syndrome compared to WS corroborates the idea that genes in this region are dosage sensitive and affect aspects of human behavior. One goal of the work I have done was to use human genetics to provide evidence for the role of specific genes in the WSCR to the behavioral phenotypes.

1.2 Genotype-phenotype correlations using human genetics

The knowledge that the WSCR causes WS and dup7q11.23 has launched many efforts to try and dissect the region to identify which genes are responsible for specific symptoms in each disorder. One avenue of research has been to make these genotype-phenotype correlations directly in humans. Human research in WS has employed three strategies: 1) compare individuals with atypically small deletions of the WSCR to individuals with the typical deletion to ask what the differences are when some genes are spared, 2) use iPSC lines derived from patients with WS, dup7q11.23, and atypical deletions to test molecular and cellular effects of the region, and 3) using classical human genetic strategies to identify variation in the general population in this region that is associated with phenotypes of interest. While each strategy has unique benefits and limitations, each has provided insight into the genetic contributions of the WSCR to different phenotypes seen in WS.

1.2.1 Atypical deletions

WS is caused by the deletion of 1.5Mbp, which covers 26 genes, in 95% of cases. In 3-5% of cases of WS, a 1.8Mbp deletion removes one copy of two more genes, *NCF1* and *GTF2IRD2*. In addition, there are a very small percentage of cases that are caused by atypically small deletions that maintain the normal copy number of a subset of genes. Researchers have explicated the different phenotypes of individuals with atypical deletions to understand the contribution of the spared genes to the phenotypes observed in typical cases of WS.

While most cases of WS are caused by *de novo* deletions, there are instances of smaller inherited deletions that allow the study of atypical deletions across several family members. Two families were ascertained based on the presence of SVAS and only a few clinical features of WS. These families were tested to show that they had smaller deletions that encompassed the *ELN* gene and the *LIMK1* gene. The phenotypes of the family members that had the deletion included cardiovascular disease, usually SVAS, a few of the craniofacial features of WS but not all of them, and deficits in their visual spatial cognition with auditory rote memory similar to the unaffected family members, consistent with the WSCP (30). This, along with expression data showing that *LIMK1* is present in the brain, led the authors to conclude that the *LIMK1* gene is important for the manifestation of the visual spatial impairment (30). Another study analyzed the two aforementioned families and three more kindred with inherited small deletions. The three new deletions all included *ELN* and *LIMK1*, and either extended centromerically or telomerically. All of the family members had two copies of the *GTF2I* gene. The affected members in each kindred had some craniofacial features, and fit the WSCP with poor visual spatial cognition. All affected family members had similar overall IQ that was in the normal range. These data gave further support that the *LIMK1* gene is sufficient to cause the visual spatial deficit, and since none of the deletions included the *GTF2I* gene, this gene was considered important for contributing to intellectual disability (31). In contrast to the above families that support the hypothesis that *LIMK1* is sufficient to cause the visual spatial phenotype, another study that described four new patients (including two brothers with the same inherited deletion) with small deletions that cover *LIMK1* showed that they had no visual spatial deficits (32). None of the individuals described in the study had the characteristic facial features or intellectual disability. These conflicting results highlight the complexity of using humans with atypical deletions to make conclusive genotype-phenotype correlations. The conflicting results could be due to confounds from incomplete penetrance of these genes, environmental factors, and contributions from other genetic loci in the genome.

Other atypical deletions in patients have led to the hypothesis that most of the genetic risk of the region is harbored in the telomeric end of the deletion. This is supported by the lack of any phenotypes besides SVAS in on of the patients described above that had the typical centromeric breakpoint that extended to *LIMK1* (32, 33), and three patients described by Botta *et al*. (34) and Heller *et al.* (35) that had the typical telomeric break point that extended through *ELN* but spared *STX1A*, who presented with the full phenotypic spectrum of WS. This pattern is also mentioned by Hirota *et al.* (36), who detailed the lack of the WSCP and most craniofacial features in three cases with typical centromeric breakpoints but telomeric breakpoints that extend through *ELN* but spare *GTF2I* in all cases. These findings, as well as others that are reported (2, 37–40) have lead the field to focus on two paralogous transcription factors in the telomeric end of the deletion, *GTF2IRD1* and *GTF2I* as major contributors to the WS profile.

Two case studies provide specific support for the role of *GTF2IRD1* in craniofacial development and *GTF2I* in the intellectual disability and social phenotypes. Tassabehji *et al.*

(39) analyzed the facial features of a patient with a typical centromeric breakpoint and telomeric breakpoint that falls within *GTF2IRD1* deleting its transcription start site, leaving *GTF2I* intact. The patient did not have the hypersocial phenotype, yet her language development was delayed, and she had visual spatial deficits, however, not to the same extent that is normally seen in typical WS. Her facial features were intermediate of what is typically seen in WS. Dai et al. (38) described another patient with the typical centromeric break point that extended through *GTF2IRD1* and spared only *GTF2I.* This patient had all the typical craniofacial features of WS and performed higher on verbal tasks but still had difficulty with some spatial tasks, but not as large of a deficit as seen in typical WS. Finally, the patient did not show the hypersocial phenotype, which led the authors to conclude that *GTF2I* plays an important role in this domain.

Larger deletions that delete the *NCF1* and *GTF2IRD2* as well as the typical genes in the WSCR, can provide insight into the contribution of these two genes. In general individuals with larger deletions tend to have more cognitive difficulties (37, 41). Comparing the larger deletion groups with a typical deletion group showed similar overall cognitive functioning, but specific areas of further deficit in the larger deletion group. These areas pertained to cognitive flexibility and spatial perception (41). Individuals with larger deletions also had more social cognition problems and obsessive behaviors than the typical deletion (42). The *GTF2IRD2* gene has been suggested to cause the slightly more severe phenotype because of its similarity to the other member of the GTF2I family and the evidence that it is expressed in the brain. The *NCF1*gene has been shown to modify the cardiovascular phenotype, and deletion of this gene is protective against hypertension in WS (43). These studies show that the larger deletion further exacerbates the cognitive phenotypes of the typical deletion and modifies the cardiovascular phenotype, suggesting that multiple genes contribute to multiple phenotypic domains in WS.

Studying atypical deletions in patients with WS has provided insight into the contribution of loci within the region to phenotypes in specific cases. This study design has several inherent limitations that should caution the field from making too strong of conclusions. First, the atypical deletions are rare events and each patient represents a unique deletion, except in the case of inherited deletions. This makes it impossible to generalize the conclusions from one case to the others and limits the potential to perform and make statistical inferences. Second, there is an ascertainment bias towards individuals with *ELN* deletions, which means the atypical deletions rarely affect just one of the genes in the region, making it difficult to test if one gene is sufficient to cause a specific phenotype. Third, these studies ignore the consequences of environmental and background genetic variation. It would be beneficial to be able to compare typical and atypical deletions to their parent's data to get an idea of the effect size of the deletion in the context of other inherited genetic variation. Finally, each of the cases is described by different clinicians with different and biased expertise for specific phenotypes. This makes it difficult to directly compare phenotypes across studies especially when some of the phenotypes weren't investigated. Overall, the study of atypical human deletions consistently shows that several genes can contribute in some degree to many phenotypes, such as craniofacial features, the WSCP, and overall cognitive ability. The telomeric end of the deletion seems to harbor the largest risk for most of the phenotypes observed in WS (2, 34, 36, 38).

1.2.2 Human induced pluripotent stem cell (iPSCs) studies

Patients with atypical deletions of WS allows for the study of the effects of specific genes or sets of genes on observable clinical phenotypes, but does not permit the study of underlying cellular or molecular changes. The advent of human derived induced pluripotent stem cells (iPSCs) as a model for human disease circumvents the need to obtain specific tissues from a human patient – particularly a challenge for the brain – and let's researchers query cellular morphology and function, and look at the disruption of different molecular pathways. The Williams syndrome field has adopted these approaches to study the effects of the deletion and duplication of the region at a cellular level in different affected tissues (44–49). This strategy has highlighted the roles of *GTF2I* (44) as well as other genes, such as *BAZ1B* (49) and *FZD9* (46).

Two early iPSC studies looked at the effect of the typical deletion on cardiovascular (47) and neuronal phenotypes, establishing this technique as a model for the study of WS (48). Kinnear *et al.* used iPSC to test the cardiovascular phenotype of cells with the WSCR. They showed that when the cells were differentiated into vascular smooth muscle cells, the WS cell lines were more immature based on lower expression of markers in mature smooth muscle cells. They went on to show that rapamycin can rescue this immaturity phenotype (47). Khattak *et al.* used the same patient's cells to investigate the functioning of iPSC derived neurons with the WS deletion. The main electrophysiological deficit was in the repolarization of the cells due to lower expression of potassium channels. This study also profiled the transcriptomes of the WS derived neurons and wild type (WT) derived neurons and found that synaptic genes were among the most differentially expressed (48). Since these studies used stem cells from the same patient that was selected for severe cardiovascular disease, they don't represent independent biological experiments. Further, the patient was also diagnosed with clinical autism, which has a higher prevalence in WS, but this could affect the interpretation of the neural phenotypes that are not generalizable to typical cases of WS. These studies show the potential for identifying physiological differences at the cellular level in cases of typical deletions, however they did not attempt to make specific genotype-phenotype correlations.

iPSCs can be used to make assertions about the contribution of specific genes to specific cellular phenotypes, which helps understand the functional roles of genes in the WSCR. Adamo *et al.* performed RNA-seq experiments in iPSCs from four separate patients with WS, two patients with 7q11.23dup syndrome, and three related normal controls and three external control cell lines, and showed that there were symmetrical changes in expression of genes in disease related pathways. They performed a similar experiment after differentiating the iPSCs into neurons, and observed enrichment of genes involved in axon guidance, cell polarity, and transmission of nerve impulses. To test the specific contributions of *GTF2I*, they performed RNAi knockdown of *GTF2I* in the 7q11.23dup and WT cell lines, and showed that about 10-20% of the transcriptional changes observed in the full WS deletion can be attributed to *GTF2I.* They went on to show that *GTF2I* interacts with the chromatin modifiers LSD1, a histone demethylase, and HDAC2, a histone deacetylase (44). They argue that most the transcriptional changes caused by decreased dosage of *GTF2I* are indirect, and propose that the dysregulation of the *GTF2I* target, *BEND4,* is a likely candidate that contributes to the downstream transcriptional changes. They remark that there is considerable variation between patient cell lines and the expression of *BEND4*, which highlights the importance of considering the genetic background. Overall, this study does suggest that *GTF2I* plays a role in the transcriptional phenotype, but does not account for all of it.

Additional iPSC studies provided evidence for functional roles of genes on the centromeric end of the deletion in neuronal phenotypes (46, 49), which the atypical deletion human studies have suggested do not contribute to the phenotypic spectrum of WS. Neural progenitor cells derived from typical deletion WS cases showed increased apoptosis that was not seen in cell lines derived from WT or an atypical case, whose deletion spanned from *CLDN3* to

RFC2. Reasoning that FZD9, which is not deleted in the atypical case, and regulates programmed cell death, the authors showed that knocking down *FZD9* in the WT cell lines could recapitulate the apoptosis phenotype and overexpressing *FZD9* in the WS cell lines could ameliorate the apoptosis phenotype (46). Transcriptional profiling of WS and WT derived neuronal cell lines along with *BAZ1B* knockdown showed as much as 42% of the transcriptional difference between WS and WT neurons were caused by decreased expression of *BAZ1B*. The transcriptional changes along with genes bound by *BAZ1B*, suggested a role for this gene in the regulation of Wnt signaling as well as synaptic development. Decreased expression of *BAZ1B* resulted in neural progenitor cells maintaining a proliferative state, which prevented proper differentiation into neurons. This phenotype could be rescued by antagonizing Wnt/Beta-catenin signaling (49). Together, these two iPSC studies strengthen the evidence for genes in the centromeric end of the deletion to play an important role in neural development, which could lead to the striking cognitive and behavioral phenotypes of WS. They also further implicate specific pathways such as Wnt signaling and synaptic functioning in the pathogenesis of WS.

It has been shown for WS that iPSCs are a valuable model to understand cellular and molecular phenotypes caused by the typical deletion as well as by specific genes in the region. While this model has its advantages it also has several limitations. iPSCs study designs allow for the testing of disease relevant tissues using human cells, however, the cells are artificially differentiated outside the context of the organ-specific microenvironment. This can lead to unforeseen changes in the biological functioning of the cells. Further, the study of cells *in vitro* precludes making associations with the cellular changes directly to behavior at the organismal level. For example, iPSC differentiated neurons do not form the complex anatomical circuits equivalent to what is seen in the brain *in vivo*. In spite of these limitations, in the case of WS these studies have provided further insight into genes such as *FZD9*, *BAZ1B*, and *GTF2I*, and suggest that they all contribute to neuronal phenotypes.

1.2.3 Human general population association studies

Another strategy that employs human genetics to identify genotype-phenotype correlations of genes in the WSCR, is to test variation in these genes for association with traits in the general population, both in samples of individuals with WS and in case-control designs. The duplication of the WSCR was found to be significantly associated with autism spectrum disorder (ASD) diagnosis in a case-control study design (26). Association analyses have further implicated the general transcription factor 2i family of genes in social and cognitive phenotypes (50–53). The advent of next generation sequencing technology offers new potential to implicate not only common variants, but also rare variants in the pathogenesis of WS (54).

Candidate gene associations have implicated two single nucleotide polymorphisms (SNPs) in the *GTF2I* in ASD as well as in neural phenotypes related to social cognition (50, 51). While these studies were not unbiased screens of the whole genome, the authors reasoned that the WSCR contains loci that affect social behavior. When variants in *STX1A, CLIP2*, and *GTF2I* were tested for association with ASD diagnosis in families with at least one affected child, only two SNPs in *GTF2I*, rs4717907 and rs13227433, were found to be over transmitted in the probands (50). Using this previous finding, these two SNPs were further associated with a metric that captures the low social anxiety and reduced social communicative skills of individuals with WS in a sample of 488 individuals attending university (53). The imputed rs13227433 genotype was also found to be associated with reduced amygdala reactivity to threatening stimuli, a neural phenotype that has been documented in WS (51) in a sample of 808 university students. Finally the SNP, rs2267824, located within the *GTF2IRD1* gene, was associated with a metric that captures the neuroanatomical gestalt of the WS brain in a sample of 1,863 people from the general population, suggesting that it contributes to brain anatomy that is specifically observed in individuals with WS (52). These candidate gene focused studies corroborate the role of both *GTF2I* and *GTF2IRD1* in behavior and brain development related to WS.

Association studies are valid study designs to identify genomic loci that correlate with a trait of interest, but they have several limitations. The detection of a significantly associated variant does not mean the causal variant has been detected. Rather, in most cases an association elucidates a region in the genome that contains the causal variant. In addition, association studies based on genotyping with SNP-chips are only able to test common SNPs, which are expected to have small effect sizes, so in order to detect these effects large sample sizes are required. To overcome this, next generation sequencing technologies can be used to query the role of rare variants in modifying the phenotypes of WS. Since WS is caused by the contiguous deletion of 1.5-1.8Mbp on chromosome seven, individuals with WS only have one remaining copy of the region, which could unmask the effects of recessive alleles (55). This hypothesis was tested for the cardiovascular phenotype, looking specifically at variants in the *ELN*. With a sample size of 55 individuals, no one specific variant associated with severity of the cardiovascular disease (56). This approach could be applied to other genes in the WSCR as well as other phenotypes in the region to understand how the genetic variation associates with different aspect of the disorder.

The human approaches taken to study the genotype-phenotype correlations within WS has, so-far, have highlighted the variability of the phenotypes and a complex relationship with the genes in the region. My work has focused on describing how genetic variation within the WSCR and in the whole exome can modify the social phenotype of WS. I analyzed the whole exome sequencing data of 85 individuals with WS to associate genetic variants with the social phenotype. I can use the variation across individuals with WS that have received the same standardized social questionnaire to ask how much does genetic variation contribute to the social phenotype. This allowed me, in an unbiased, way to test for genes in the WSCR and the whole exome that are important for modifying social behavior in WS, which could inform clinicians taking care of individuals with WS as well as inform genes involved social behavior in the general population.

1.3 Introduction to the general transcription factor 2I family

Performing gene associations in humans, while informative on what locations of the genome are important for different traits, are not conducive to conducting controlled experiments that could lead to a mechanistic understanding of how genes exert their effects on behavior. Along with the human studies I did, I leveraged the experimental advantages of the mouse model to focus on the interactions of two genes in the WSCR, *Gtf2i* and *Gtf2ird1*. I chose to investigate these genes to test the hypothesis that they contribute to the cognitive and behavior phenotypes as the human literature has suggested and to extend the current research by testing how they interact. This family is made up of three paralogous transcription factors that are located in the WSCR. *GTF2I* and *GTF2IRD1* are deleted in the 1.5Mbp deletion, and *GTF2IRD2* is deleted in the larger 1.8Mbp deletion. These transcription factors have been extensively studied in different model systems, including cell lines and mouse models, usually focusing either on *GTF2I* or *GTF2IRD1*. Since both seem to contribute to overlapping phenotypes and they share overlapping DNA binding targets, these transcription factors merit further investigation.

1.3.1 General background on the GTF2I family

Different groups discovered the GTF2I family of genes independently. *GTF2I* was discovered in several contexts, including a target of Bruton's tyrosine kinase in B-cells (BAP-135) (57), a protein that stabilizes the serum response factor complex (SPIN) (58), and as a transcription factor in the WSCR that can bind to the E-box and *Inr* element (59), which were all shown to be the same GTF2I protein. *GTF2IRD1* has a similar history in which it was discovered many independent times as a gene expressed in the muscle (MusTRD1) (60) as well as a transcription factor in the WSCR (WBSCR11) (61, 62).

All three are multiexonic genes that are subject to extensive alternative splicing. *GTF2I* is made up 35 exons, *GTF2IRD1* contains 27 (63), and *GTF2IRD2* has only 16 exons due to the replacement of the 3'prime end of the gene with a CHARLIE8 transposon (64). The sequence features that distinguish these genes as a family are the I repeats, of which *GTF2I* contains 6, *GTF2IRD1* contains 5, and *GTF2IRD2* has 2. These are helix-loop-helix domains that are thought to be important for protein-protein interactions and DNA binding (65). They also have a conserved N-terminal leucine zipper (66, 67), that is involved in homo and heterodimeriziation that can affect DNA binding function. The evolutionary history of these genes points to *GTF2IRD1* as the ancestral gene that was duplicated to produce *GTF2I.* These two genes are present in all land mammals with the duplication and inversion of *GTF2I* giving rise to *GTF2IRD2*, which is present in all placental mammals (68). This conserved evolutionary history in mammals makes studying these genes tractable in mouse models. The mouse Gtf2ird1 and human GTF2IRD1 share 87.9% amino acid identity and the mouse Gtf2i and human GTF2I share 97.3% amino acid identity (69). Given the similar evolutionary history of these genes, it is important to understand to what extent these genes share overlapping function, as well as how they differ.

GTF2I was the first gene discovered and has been the best studied, probably due to its higher abundance in many different tissues and due to the availability of effective antibodies. The expression of *GTF2I* is described as ubiquitous, with higher expression early in development. In the mouse, *Gtf2i* mRNA is maternally deposited by the mother in the fertilized egg and is highly expressed in the inner cell mass, and continues to be highly expressed throughout development (70, 71). *In situ* hybridization experiments in the mouse brain showed uniform expression of *Gtf2i* from embryonic day 18.5 to postnatal day seven, with enhanced expression of the mRNA in Purkinje cells, the hippocampus, and cerebral cortex in the adult brain, all of which was described as neuronal. The protein showed a similar expression pattern, with protein detected in both the nucleus and the cytoplasm, with enrichment in the hippocampus and cerebellum (72). The presence of GTF2I in both the nucleus and the cytoplasm suggests that this transcription factor has functions beyond regulating nuclear transcription.

Along with its roles as a basal transcription factor, GTF2I plays a role in the cytoplasm that allows it to convey cellular information to the nucleus. GTF2I was first discovered due to its ability to bind the *Inr* element at transcription start sites but also at upstream enhancers (73). It was shown that some of its transcriptional activity was due to tyrosine phosphorylation by *SRC* that allowed cytoplasmic GTF2I to translocate to the nucleus, suggesting that GTF2I can induce transcriptional changes based on signal transduction pathways (74). Interestingly *Src* knockout mouse models show phenotypes such as hyperactivity and hypersociability, suggesting that disruption of this gene and its downstream pathways can recapitulate some features of *Gtf2i* knockout models (75). Another effect that phosphorylation of GTF2I by SRC has is to inhibit agonist induced calcium entry (76). The cytoplasmic phosphorylated GTF2I competes with TRPC3 protein, a calcium channel, to bind PLC-γ, which prevents the localization of TRPC3 to the membrane and inhibits calcium entry into the cell. This was shown to affect neuron morphology and calcium electrophysiology in neurons that are missing one copy of *Gtf2i*. The neurons with less *Gtf2i* had more complex axons and increased calcium entry (77). These studies have elaborated the complex cellular role that GTF2I plays in both transcription and signal transduction and how it can affect neural phenotypes, which may contribute to phenotypes in WS. No studies have been done that show what happens to transcription genome-wide in the brain when *Gtf2i* is increased or decreased, which the work I present in chapter three describes. Also, given the dual role of this transcription factor the paucity of data concerning its effect on transcription makes it difficult to disentangle which functionality of *Gtf2i* is contributing to affect behavior.

In contrast to the extensive transcriptional roles and signal transduction function of *GTF2I*, *GTF2IRD1* has mostly been characterized as having a role in transcriptional regulation. The expression of this gene was described using a lacZ reporter in the mouse. Ubiquitous expression was seen at embryonic day 7.5 with more localized expression occurring after organogenesis. In the developing brain it is expressed most highly in the pituitary, developing hypothalamus and thalamus, and hindbrain with little expression in the telencephalon. The gene is expressed less in adulthood across all tissues, and within the brain it is the most highly expressed in the olfactory bulbs, Purkinje neurons, and neurons of the piriform cortex. It is highly expressed in adult brown adipose tissue (78). The low expression of this gene *in vivo* along with poor antibodies has made this protein difficult to study *in vivo.* However, work in cells that highly express this gene show that is mostly localized to the nucleus in a punctate pattern and in close proximity to other chromatin regulators such as SP1 and H3K27Me2/3 and H3K4Me3 marks. A yeast 2 hybrid screen further suggested that GTF2IRD1 interacts with chromatin modifiers such as ZMYM2 and ZMYM3 along with proteins involved in ubiquitin pathways such as USP20 and USP33 (79). These data suggest that it plays a role in transcriptional regulation.

Other studies have shown that GTF2IRD1 binds to specific genomic regions to affect transcription and the *Gtf2ird1* genes is under tight transcriptional and posttranslational regulation. In the mouse retina Gtf2ird1 binds to the LCR enhancer and promoter regions of opsin genes to promote transcription (80). Hasegawa *et al.* showed that Gtf2ird1 expression is induced in mouse brown adipose tissue in cold conditions and associates with the PRDM16 complex to repress fibrotic gene transcription (81). In addition, GTF2IRD1 has been shown to negatively autoregulate its own transcription. The N-terminal leucine zipper was proposed to increase binding to its own upstream regulatory element and mutating the leucine zipper resulted in a difference in bind affinity to the sequence (66). Finally, GTF2IRD1 is post translationaly modified by the addition of a SUMO group that alters its protein-protein interactions and targets the protein for degradation (82). The extensive roles of Gtf2ird1 in transcriptional regulation and its tightly regulated mRNA and protein expression suggest that this gene plays an important biological role that could contribute to the phenotypes of WS.

The DNA binding of these two transcription factors has been studied genome-wide in different model systems. The core binding motifs for the fourth I repeat of GTF2I and GTF2IRD1 was identified as RGATTR using the SELEX method (83). In a similar experiment the binding site of the full length GTF2IRD1 was determined to be GGGRSCWGCGAYAGCCSSH (65). Chip-Chip experiments in mouse embryonic stem cells revealed 5,744 binding peaks for Gtf2i and 625 binding peaks for Gtf2ird1, most of which were located in promoters of genes. When binding was investigated in embryonic craniofacial tissue they identified 1,181 Gtf2i binding peaks and 1,520 Gtf2ird1 binding peaks, again most were located in promoter regions. They showed examples of sites where both proteins were located at the same promoter regions suggesting they can overlap in the genes they regulate. Most of the binding sites were located in areas of bivalent chromatin marks (84). GTF2I binding has been assessed in human iPSC cells using ChIP-seq and was found to bind 1,554 genes at their promoters. About half of these binding sites were also targets of the LSD1 histone demethylase (44). Gtf2i has also been shown to help target CTCF to promoter regions. Genome-wide binding analysis of Gtf2i and Gtf2ird1 show that they have overlapping targets and cooperate with other chromatin regulators. Further study of the binding patterns of these proteins *in vivo* in other relevant tissues will continue to elucidate the role these genes play in transcription regulation and downstream affected pathways.

Given that both *GTF2I* and *GTF2IRD1* are transcription factors and they bind many genes in the genome, their affects on transcription genome wide have been minimally described and with contrasting results. Gtf2ird1 overexpression in mouse embryonic fibroblasts led to around 1,000 upregulated genes and 1,000 downregulated genes covering pathways such as ubiquitin cycle, RNA binding, and cell cycle (85). In contrast, Gtf2i overexpression in mouse embryonic fibrobalsts led to fewer changes with only 90 genes upregulated and 68 genes downregulated. These genes made up categories such as transcription regulation, immune response, and apoptosis (86). The effects of knocking out each transcription factor was assessed in embryonic day 9.5 mouse models. In the *Gtf2i* null embryos there were 217 upregulated and 2,356 downregulated genes spanning categories such as cytoskeleton remodeling, cell cycle, transcription, and the ubiquitin cycle. However, *Gtf2ird1* null embyros showed only 38 upregulated genes and 498 downregulated genes that did not show any enrichment for specific GO categories (87). These findings somewhat mirror the overexpression data.

Another *Gtf2ird1* mouse model profiled the transcriptomes of the developing brain at embryonic day 15.5 and postnatal day 0 and showed no significantly differentially expressed genes (88). Yet another *Gtf2ird1* model that showed overgrown lip epidermal tissue revealed 1,165 upregulated genes and 1,073 down regulated genes. Gene set enrichment analysis on the upregulated genes highlighted pathways such as cell cycle, the ribosome, proteasome, and ubiquitin mediated proteolysis. Down regulated genes showed enrichment in calcium signaling, oxidative phosphorylation, and cardiac muscle contraction (89). Finally, transcriptome profiling of the hippocampus in a mouse model that has the entire syntenic WSCR deleted showed down regulation of genes in the Pik3 kinase pathway as well as *Bdnf* (90, 91).

Overall, transcriptional studies of *Gtf2i* and *Gtf2ird1* seem to be dependent on many factors that include tissue type, stage in development, how the genes are mutated, and mouse strain. The transcriptome data generated in the E9.5 embryos should be cautiously interpreted since both the *Gtf2i* and *Gt2ird1* null mutants described were embryonic lethal and showed neural tube closure defects as well as vascular defects. Comparing these very severe embryos to the WT embryos show that many of the transcriptional changes detected are probably consequences of the disrupted development of the embryo, which make teasing out the direct and indirect effects of reducing the expression of *Gtf2i* and *Gtf2ird1* difficult. The discrepancy between the transcriptome findings of the brain and the lip tissue could arise for several reasons. Different mutants were used and *in vivo* analysis of the Gt2ird1 protein was lacking in both studies. It would be beneficial to know how the mutations are affecting the protein levels as well as the normal WT levels of protein expression between these two tissue types. The lip tissue also showed a clear morphological phenotype that specifically affected the epidermal tissue and not the dermal tissue, cartilage, or underlying muscle. The striking difference between genotypes in the lip tissue could be driven by a clear disruption of a specific cell type (92), while in the brain there could be more subtle effects in different cell types diluting the signal. Incorporating multiple levels of information such as ChIP-seq, RNA-seq, and tissue specific expression of these genes will aid in constructing a more complete understanding the role of these transcription factors.

1.3.2 Mouse models of *Gtf2i* and *Gtf2ird1*

Along with understanding what the molecular functions of these two transcription factors are, in order to provide useful insight into the etiology of WS, the affect these two transcription factors have on behavior should also be studied. Previously, I have described the evidence that supports the functioning of these genes in behavior, cognition, and physical attributes that we have gleaned from human studies. As mentioned, human studies come with their own limitations: in rare partial deletions one is making inferences based on single individuals. Likewise, one is unable to model behavioral consequences in iPSCs. Model organisms, specifically the mouse, have been instrumental in understanding both the functional roles of genes as described in the previous section and the consequences of dosage changes of genes on behavior. The mouse is an attractive model in which to model WS for several reasons: 1) a region of chromosome five in the mouse is syntenic to the WSCR in humans, 2) geneticists have a large tool kit in which to accurately modify the mouse genome to test specific mutations or sets of mutations, 3) mice are able to be bred so that the same mutation can be studied in a large, controlled sample allowing for statistical inferences, and 4) mice are social animals that display behaviors in domains that are disrupted in WS. Many different mouse models have been used to try and understand the behavioral consequences of genes in the WSCR, with varying degrees of face validity to WS. The strain of mouse and how the mutations were generated play a large role in the manifestation of phenotypes in mouse models. This makes synthesizing the data from different labs and experiments difficult, but consistent phenotypes across many different models can provide strong corroborative evidence for genotype-phenotype correlations.

Large deletion mouse models

The mouse model with the highest construct validity is a hemizygous deletion of the syntenic WSCR on the mouse chromosome five and is termed the complete deletion (CD) mouse (93). The mouse was generated using the cre-lox system with a loxP site situated in exon two of *Gtf2i* and the other loxP site in intron five of *Fkbp6* on the C57BL/6J background. This mouse model showed phenotypes that are consistent with most of the phenotypes of WS that can be tested in the mouse. The physical features include mild cardiovascular phenotypes, smaller skulls, reduced brain size, decreased volume of hippocampus, and more immature neurons in the dentate gyrus as determined by doublecortin immunostaining. A battery of behavior tests in the CD mice showed deficits in motor coordination, decreased motor tonicity strength, increased startle response to stimulus noise, and a decreased habituation to a social stimulus (93). Another study of the CD mice showed deficits in working memory as tested by the spontaneous alternating T-maze and novel object recognition, which was reported as normal in a previous study. The social phenotype was replicated as well as a decrease in the number of marbles buried in the marble burying task (91). Finally, the role of *Gtf2i* in the manifestation of the behavioral phenotypes in the CD animal was tested by delivering adeno-associated virus 9 (AAV9) that carried the mouse *Gtf2i* cDNA into the cisterna magna of CD mice. The addition of *Gtf2i* cDNA rescued the increased social phenotype, partially rescued the motor coordination, but did not affect the marble burying deficits (90). This mouse model showed deficits in motor coordination and increased startle to a stimulus noise. Humans with WS are known to have poor balance as well as hyperacusis. The mice also recapitulated the hypersocial phenotype of WS as tested in these behavioral tasks using only male mice. The CD model is a great tool to understand how the entire WSCR affects mouse behavior and the underlying mechanisms. However, the work done in the CD mouse should be expanded to include female mice to understand any sex or sex by genotype interactions. This would also inform how robust the phenotypes are. For instance the social phenotypes have only been tested in males using an unconventional method. Including social tasks that probe different aspects of sociality would help pinpoint the specific pathways involved in manifesting the disorder.

There are two other large deletion models that attempt to localize which genes are involved in specific mouse behaviors by splitting the WSCR into two halves and deleting each half (94). These mice were generated using the cre-lox system on the C57BL/6J background. The proximal deletion mice (PD) are hemizygous for *Gtf2i* through *Limk1*. The distal deletion mice (DD) are hemizygous for *Trim50* through *Limk1*. Breeding the PD and DD mice together results in four littermate genotypes, which include a mouse that is hemizygous for the whole region on two different chromosome and is homozygous null for *Limk1*, this is called the P/D mouse. The DD and P/D mice showed similar shortened skulls with more severe differences in the P/D mice. This indicates that genes on the distal half of the deletion contributing to the craniofacial phenotypes, but perhaps genes in the proximal half can exacerbate the phenotype. There were mixed results on a series of behavior tasks that probe social behavior. The partition task showed all three genotypes spent more time at a partition that held a social stimulus than WT littermates.

A direct social task showed only the PD mice spent more time investigating a conspecific compared to WT littermates. The three chamber social approach task showed a significant preference for the social stimulus in the PD and P/D mice but no such preference in the WT or DD animals. Finally in a test of social dominance the PD and P/D mice had a decrease win ratio suggesting reduced dominance behavior. The P/D mice showed decreased locomotor activity and poor balance, and the partial deletions had intermediate values. Altered response to sensory stimuli was tested using the acoustic startle response and pre-pulse inhibition. This was only altered in the PD mice with no phenotype in the P/D or DD genotypes. In a learning and memory task, the DD mice showed decreased freezing in contextual and cued fear memory. Studying the two half deletions can help further localize the genes involved in specific phenotypes. These studies suggest that the DD genes are involved in the craniofacial phenotypes and fear memory recall. The PD genes affect, in some tasks, social behavior and the response to sensory stimuli (94). Genes in both halves of the deletions may contribute to balance deficits, which is more affected when both halves are deleted. Overall, it seems like some phenotypes such as the balance and craniofacial differences are being influenced by multiple genes.

***Gtf2i* and *Gtf2ird1* mouse models**

The larger mouse models of WS test the affects of knocking out the entire region on mouse behavior. One of the advantages of mouse models is the wide range of tools geneticists have at their disposable to manipulate the genome, permitting the study of very specific mutations of single genes. Single gene knockout mice exist for several different genes in the WSCR, with many genes having multiple different mouse models (95). For *Gtf2i* there are two mouse models that decrease the expression of Gtf2i to varying degrees. One model has a gene trap cassette in intron 3 of *Gtf2i* ($\text{Gtf2i}^{\text{Gt(YTA365)Byg}/\beta}$) that has been characterized in (29, 87, 96,

97) and results in a null allele. The second *Gtf2i* model has a targeted deletion of exon 2 (Gtf2i$^{\Delta ex2}$), which contains the canonical translation start codon, and produces an N-terminally truncated protein that begins at a methionine in exon five and is described in (67, 90). The former model is embryonic lethal in the homozygous state (29, 87, 96) and the latter model produces viable homozygous animals at a lower than expected Mendelian ratio (67). There are four different mouse models of *Gtf2ird1* that have been described in the literature: 1) the Gtf2ird1^{XE465} model has a gene trap lacZ cassette located in intron 22 (87), which makes a fusion protein, 2) the *Gtf2ird1* Tg(Alb1-Myc)166.8 model has a myc transgene that randomly integrated into the locus replacing the transcription start site and the first exon of *Gtf2ird1*(39, 98–100), which has no detectable expression, 3) the Gtf2ird1^{tm1Hrd} model was made by homologous recombination removing exon 2, which contains the canonical translation start codon, and has increased expression of *Gtf2ird1* transcript but produces an N-terminally truncated protein at 3% of WT levels (66, 78, 92), and 4) the Gtf2ird1^{tm1LR} model was made by homologous recombination removing exons 2,3,4 and part of 5, which still makes an aberrant *Gtf2ird1* transcript but protein analysis was not done (101). All of the Gtf2ird1 models can produce viable homozygous animals except for Gtf2ird1^{XE465}, which expire embryonically. This more severe phenotype has been attributed to the production of a fusion protein whose function is unknown (88). While there are many mouse models of both of these genes that have been tested on different mouse backgrounds and on different behavioral tasks, synthesizing the data across all the experiments can provide strong evidence of the roles of these genes on behaviors.

The two mouse models of *Gtf2i* have shown hypersocial phenotypes (90, 96, 97). The specific social phenotypes queried by the specific tasks differ. In two experiments the *Gtf2i* heterozygous mutants display a lack of habituation to a social stimulus that is normally observed in WT littermates (96). Another experiment using the N-terminally truncated protein, and showed that heterozygous and homozygous mutants investigate a social stimulus for more time compared to WT littermates (90). The most convincing experiment employed a social operant learning paradigm, in which the heterozygous mutants will work harder for more social rewards (97). Besides the social phenotypes other behaviors have been documented such as impaired novel object recognition, increased anxiety, motor coordination marble burying in homozygous animals, and smaller craniums (67, 90, 96). The work done in single *Gtf2i* mutants supports its role in the social aspect of WS, and potentially in anxiety, motor ability, and the craniofacial features.

The several *Gtf2ird1* mutant mouse models show many behavioral and physical deficits, but in some models exhibited findings that contrast other models. Furthermore, some of the phenotypes are only seen in the homozygous knockouts, which don't reflect the gene dosage effects that are expected to be seen in humans with WS. One consistent phenotype seen in two models of *Gtf2ird1* is a motor coordination deficit, which was also seen in the larger deletions of the WSCR and in one *Gtf2i* model (92, 100). Other phenotypes such as activity levels and anxiety-like behaviors are discrepant across models. Some models report increased activity and decreased anxiety, while another reports the opposite (92, 100, 101). Social behavior has only been tested in one *Gtf2ird1* mouse model, using the resident intruder paradigm, which showed decreased aggression, but an increase is social investigation by the *Gtf2ird1* heterozygous and homozygous mutants (101). Two models have reported facial dysmorphisms in the mice, one which affects the cranium and the other affects the soft tissue of the face (39, 92). The contrasting evidence in these mouse models could be due to the mouse background on which each model was made or how the gene itself is disrupted. The evidence shows that this gene may also contribute to motor deficits, the social phenotype, and the craniofacial phenotype that is observed in WS.

Overall, single gene knock out models of both *Gtf2i* and *Gtf2ird1* show overlapping behavior phenotypes, specifically in social and motor deficits. However, these genes have not been studied in combination, which is what is expected in the deletion of WS. This leaves open the question if this two paralogous transcription factors can interact with each other to synergistically affect behavior? Understanding how these genes function together will give a more complete understanding of how genes in the WSCR interact to produce the full phenotypic spectrum of WS. These hypotheses are addressed in chapters three and four of this thesis.

1.4 Conclusions

Both human and mouse genetic experiments have demonstrated that the WSCR is an important genomic region for a variety of traits, such as craniofacial development, cardiovascular health, cognition, anxiety, and social behaviors. The field has employed many different strategies to further understand the genes responsible for causing the phenotypes of WS and thus providing insight on the biological mechanisms of different human characteristics. Still the only strong monogenetic contribution of a gene in the WSCR to a specific phenotype of WS is the role *ELN* plays in the cardiovascular disease. Even this monogenic contribution can be modified by another gene in the region *NCF1*. There is evidence for the role of several genes contributing to several different phenotypes. This oligogenic hypothesis may help the field further understand how the genes in the WSCR work together to produce the WS phenotypes, as has been shown for other copy number variation disorders (102, 103). The work I describe in this thesis uses both human genetics and mouse models to expand the knowledge of how genes in the WSCR affect behavior. I use whole-exome sequencing to analyze the largest genetic dataset of individuals with WS, to test the hypothesis that variation in the remaining WSCR allele and exome-wide can modify the social phenotype. I then use newly generated mouse models to understand where *Gtf2i* and *Gtf2ird1* bind genome-wide in the developing brain and what are the transcriptional and behavioral consequences on mutating these genes. I am able to test the hypothesis that these genes both affect the same phenotypes, testing the oligogenic contribution of these genes on behavior. Finally, I use mouse models to directly compare the affects of both *Gtf2i* and *Gtf2ird1* to the affects of the entire WSCR to test if these two genes, which have been highly speculated in the literature as driving the phenotypes of WS, are sufficient to replicate the phenotypes produced by all the genes in the WSCR. My data suggest that these genes do contribute to behavior, but other genes in the region or the effect of deleting the entire WSCR has more striking behavioral consequences. This leads me to conclude that the complex phenotypes that are disrupted in WS are caused by complex genetic interactions of genes in the region and require more than loss of just these two genes. Further testing of the oligogenic relationship of genes will highlight the complex biology of human traits and the pathobiology of WS.

Chapter 2: Exome sequencing of 85 Williams Beuren syndrome cases rules out coding variation as a major contributor to remaining variance in social behavior

Nathan D. Kopp, Phoebe C. R. Parrish, Michael Lugo, Joseph Dougherty, and Beth A. Kozel

From:

Exome sequencing of 85 Williams Beuren syndrome cases rules out coding variation as a major contributor to remaining variance in social behavior.

Kopp, Nathan D., Phoebe CR. Parrish, Michael Lugo, Joseph D. Dougherty, and Beth A. Kozel. “Exome sequencing of 85 Williams-Beuren syndrome cases rules out coding variation as a major contributor to remaining variance in social behavior”. *Molecular Genetics and Genomic Medicine* (2018):1-17

2.1 Abstract

Large, multigenic deletions at chromosome 7q11.23 result in a highly penetrant constellation of physical and behavioral symptoms known as Williams Beuren syndrome (WS). Of particular interest is the unusual social-cognitive profile evidenced by deficits in social cognition and communication reminiscent of autism spectrum disorders (ASD) that are juxtaposed with normal or even relatively enhanced social motivation. Interestingly, duplications in the same region also result in ASD-like phenotypes as well as social phobias. Thus, the region clearly regulates human social motivation and behavior, yet the relevant gene(s) have not been definitively identified. Here, we deeply phenotyped 85 individuals with WS and used exome sequencing to analyze common and rare variation for association with the remaining variance in social behavior as assessed by the Social Responsiveness Scale. We replicated the previously reported unusual juxtaposition of behavioral symptoms in this new patient collection, but we did not find any new alleles of large effect in the targeted analysis of the remaining copy of genes in the Williams syndrome critical region. However, we report on two nominally significant SNPs in two genes that have been implicated in the cognitive and social phenotypes of Williams syndrome, *BAZ1B* and *GTF2IRD1*. Secondary discovery driven explorations focusing on known ASD genes and an exome wide scan do not highlight any variants of a large effect. Whole exome sequencing of 85 individuals with WS did not support the hypothesis that there are variants of large effect within the remaining Williams syndrome critical region that contribute to the social phenotype. This deeply phenotyped and genotyped patient cohort with a defined mutation provides the opportunity for similar analyses focusing on noncoding variation and/or other phenotypic domains.

2.2 Introduction

Williams Beuren syndrome (WS) (OMIM #194050) is a neurodevelopmental disorder caused by a 1.5 to 1.8 Mbp deletion on chromosome 7q11.23. The deletion causes a constellation of symptoms that include cardiovascular pathology, craniofacial dysmorphology, and a unique cognitive and personality profile(4, 14, 17). The well-defined genetic lesion that causes WS is an opportunity to assess genotype-phenotype correlations. To date, only the cardiovascular phenotype has been convincingly linked to the haploinsufficiency of a single gene - the *ELN* gene(6, 104). Studying rare events that result in atypical deletions sparing different genes in the Williams syndrome critical region (WSCR), as well as single gene knock out studies in mouse models, have suggested that *GTF2IRD1* and *BAZ1B* play a role in the craniofacial abnormalities(39, 105). Likewise, the genes *STX1A*, *LIMK1*, *CYLN2*, *BAZ1B, GTF2IRD1*, and *GTF2I* (31, 38, 49, 96, 99, 106–109) have been implicated in the cognitive and behavioral phenotypes.

Understanding contributions to social phenotypes in particular for WS may define genes that regulate human social behavior, providing insight not only into WS, but also in other disorders as well as possible modifiers of social behavior in the general population. Deleting one copy of the genes in the WSCR produces the personality profile observed in WS, which consists of prosocial behaviors such as gregariousness, empathy, retained expressive language skills, and low levels of social anxiety, in spite of high anxiety in other domains(14, 19, 110–112). Despite the high social motivation of individuals with WS, they exhibit deficits in social cognition and communication(20, 113, 114). The Williams syndrome critical region duplication, 7q11.23 duplication syndrome (Dup7) (OMIM#609757), conversely, is characterized by diametric social behaviors to those seen in WS, including separation anxiety, poor eye contact, and language impairment, as well as overlapping phenotypes such as restricted and repetitive behavior and poor social communication (27). It has also been shown that the prevalence of ASD in WS and Dup7 is higher than in the general population and the male sex bias for ASD diagnosis is present among individuals with Dup7(27, 115). The similarities and differences in the social communication domains of WS and ASD have been described, and suggest that while both disorders show deficits in social communication, the WS group was not as impaired as the ASD group (113, 114). Unlike ASD, there is no sex bias in the frequency of WS and severity of social and cognitive phenotypes are similar across both sexes (21, 116).

As in many diseases of haploinsufficiency, within WS there remains considerable variability in expressivity of the phenotypes, despite the very homogeneous genetic cause. It is thought that both genetic background and the environment introduce variation in the expression of a phenotype. The fact that individuals with WS are hemizygous for 26-28 genes has led to the assertion that variation in the remaining allele could contribute to the severity of symptoms in WS(13, 56). The presence of only one copy of genes in the WSCR could unmask the effects of recessive alleles in the region that are more difficult to detect in a diploid setting. Indeed, this logic has been applied to investigate the variability in the cardiovascular phenotype. Delio et al. 2013 sequenced the exons that make up the *ELN* gene in a sample of 55 individuals with WS, but found no clear link between severity of phenotype and remaining genetic variation. However, no similar studies have investigated the social profile of WS, in spite of the fact that there is some evidence that common variation in the region can influence social behavior in the general population. For example, variation in the *GTF2I* gene has been associated with the WS cognitive profile, autism, oxytocin reactivity, amygdala activity, and social anxiety(53, 117, 117, 118). Furthermore, genes outside of the WSCR are also likely to affect aspects of social behavior. In particular genes that are associated with ASD have a profound effect on social interaction and could harbor variants that modify the phenotype of individuals with WS.

Here, we employ whole exome sequencing to understand how genetic variation within the WSCR, and other protein coding genes, impacts the severity of the WS social phenotype. We generate a rich catalogue of genetic variants identified from 85 individuals with the typical WS deletions; each individual has also been assessed with the Social Responsiveness Scale-2 (SRS) questionnaire, a quantitative measure of reciprocal social behavior. The SRS was first developed to quantify autistic traits in both the general and clinical populations(119, 120). SRS scores have also been used to describe different aspects of the social phenotype in WS (20). We then employ a three-tiered approach to screen for the existence of alleles that contribute to SRS scores in the context of a potentially sensitizing WSCR deletion, ordering the analyses to conserve statistical power. First, we describe the genetic variants observed in the remaining WSCR and test if they can explain the variance in the SRS scores. We find little evidence that these common or rare variants in the region are associated with SRS scores. Next, we go beyond the WSCR and test variants in 71 genes known to be associated with ASD (121), reasoning variation that contributes to autistic features in non-WS children may modify autistic features in the WS cohort as well. Finally, we test variants throughout the whole exome. We find no genetic variants of sufficient effect size to support the hypothesis that they contribute to the social phenotype in this sample of individuals with WS. However, we have more thoroughly described the variation in the WSCR region as it relates to social behavior and provide the largest genetic dataset to date of individuals with typical WS deletions for future analyses of other phenotypic domains.

2.3 Results

2.3.1 SRS variability in Williams syndrome

The unique social profile of Williams syndrome includes increased social motivation (e.g. indiscriminate approach to strangers), strong eye contact, use of affective language, emotional sensitivity as well as poor social judgment and restricted interests(19, 110–113, 122). Many comorbidities, such as specific phobias, ADHD, and anxiety, have been commonly reported in WS as well(21, 22, 123–125). To quantify social features in our WS cohort, we used a standard instrument for assessing social reciprocity, parent-reported SRS scores from 85 individuals with WS.

We examined the SRS and its subscores in depth. In our sample, the SRS T-scores are continuously distributed in the WS population with a male mean T-score±SD of 64.58±12.28 (mean male raw score±SD 74.53±32.03) and female mean T-score±SD of 62.94±11.04 (mean female raw score±SD 67.08±26.04) (**Figure 1**). There is no significant difference in SRS T-scores ($t_{70.76}$=0.6365, p=0.52) or raw scores ($t_{65.907}$=1.1445, p=0.257) between sexes. To benchmark the WS values, Constantino and Todd, 2003 measured raw SRS scores in 788 twin pairs from the general population ranging in ages between 7 and 15 and estimated the mean male raw score±SD as 35.3±22.0 and the female mean raw score±SD as 27.5±18.4; males and females were significantly different. In our analysis, we show that individuals with WS have SRS scores that are shifted towards the more impaired end of the spectrum, and we do not detect any significant sex differences in WS, which has been observed in the general population.

Our results largely replicate the results seen in Klein-Tasman et al. 2010. The overall T-score distribution reveals that 40% of our samples fall into the no clinically significant impairment range, followed by 41.1% with mild to moderate deficits, and 18.9% with severe deficits. The number of individuals showing no clinical signs in our sample is higher than the 13.4% observed when the parents completed the SRS in Klein-Tasman et al. 2010, but more similar to the teacher reported results of 38.8%. The sub scores also follow a similar pattern to what has been reported previously (20). There is a significant effect of sub scale on the T-scores ($F_{4,420} = 24.759$, $p < 0.001$)(**Figure 1B**). Post hoc Tukey all-pairwise comparisons show that social motivation has significantly better T-scores than all other sub scales, consistent with Klein-Tasman et al. 2010. The social awareness and communication scales are not different from each other, but both show less impairment than social cognition and restricted and repetitive behaviors. Social cognition and restricted and repetitive behaviors were significantly more impaired than all other sub scales, but not each other.

The distribution of SRS scores in WS point to the possibility of additional genetic variants that modify the social phenotype. First, we see a larger standard deviation in the SRS data in our sample compared to that of the norming population from Constantino and Todd 2003. The extra variance suggests individuals with WS are more sensitive to genetic or environmental factors that modify social behavior. Second, in our sample there are only two individuals that show severe social motivation deficits, and these individuals also show severe deficits in the total SRS T-score as well as all other sub scales. These outliers also suggest some individuals may harbor additional rare variants of large effect size resulting in a phenotype that is more frankly autistic. To test these two hypotheses, we generated and analyzed exome sequence from this cohort of WS patients.

2.3.2 Identification of variants in the Williams syndrome critical region

Williams syndrome individuals are hemizygous for 1.5-1.8Mbp on chromosome 7q11.23. Since they only have one remaining allele, our primary hypothesis was that second hits in genes believed to impact social phenotypes within the WSCR would produce more extreme social phenotypes. We performed whole exome sequencing on 85 individuals, all of whom have an SRS score. We called 120 variants in the remaining WSCR and annotated them with the allele frequency in our sample, ExAC allele frequency, mutation consequence, clinical significance as assessed by ClinVar, and scores that assess deleteriousness of missense variants catalogued in dbNSFP. (**Supplemental Table S1**). Table I shows the 55 exonic variants discovered in the region. For display purposes we have only included the CADD PHRED score and the MetaLR score, which is a composite score that incorporates information from nine other measures of deleteriousness and has been shown to have more predictive power than the individual component scores(126).

We first examined this set of variants to determine if any loss-of-function variants might be present in individuals with particularly severe SRS scores in our sample. Upon inspection of the exonic variants, we notice no severe likely protein truncating variants. As homozygous nulls for at least two genes in this region*(ELN* and *GTF2I*) are expected to be lethal(96, 127), we also assessed missense mutations in these genes that might alter function. Based upon the predictions of MetaLR all the missense mutations called are expected to be tolerated. None of the variants were reported as pathogenic in ClinVar. The highest CADD scores observed are a novel variant and SNP rs35607697, both located in the *TBL2* gene. Another novel variant was identified as a synonymous change in the *BAZ1B* gene. Similar results are found for non-exonic variants in the region (**Supplemental Table S1**). This suggests that beyond the reduced copy number of the entire WSCR, neither a second rare deleterious coding variant nor any common missense mutations in the WSCR explain individuals with outlier SRS scores. It should be noted that we did not identify any variants in *GTF2I*, one of the primary candidates for mediating the social cognitive profile.

2.3.3 Association analyses

To test the hypothesis that individual variants in the WSCR can explain the variance in the SRS scores in our sample, we perform classic quantitative trait loci associations. Rare disease populations by definition will have small sample sizes such as in this study. We calculated the power of our current study to be able to detect variants with different effect sizes and also calculated the number of samples that would be needed to reach a certain power given an effect size (**Figure 2**). We calculated the power for analyzing variants in the WSCR, variants in 71 ASD genes, and the remaining variants identified throughout the exome. Since we are conducting fewer tests in the WSCR, we have the most power in this analysis, however we are still only powered to detect very large effect sizes that might be unmasked by the hemizygosity of the region, such variants would need to explain more than 10% of the heritability of the trait to achieve 80% power. Most effect sizes for common variants in diploid regions of the genome typically assessed by GWAS for complex traits explain around 1% of the heritability of the trait(128). In order to be able to detect variants that explain 5% of the variance of the trait with 80% power using only variants in the WSCR would require 312 individuals (**Figure 2B**).

We then performed a quantitative trait association analysis of common variants in the WSCR on the SRS T-scores from the whole cohort. We used PLINK to test for an association on each of the 34 common variants in the WSCR, defined as MAF > 0.05, which corresponds to an allele count of at least four in the WSCR due to the hemizygosity of the region. We adjusted for age, sex, and ancestry. We found no association between any SNP and SRS that survived multiple comparison corrections (**Figure 3A**). The top five SNPs are displayed in Table II. Interestingly, the most significant SNP, rs2074754, is located in the *BAZ1B* gene, which has been previously implicated in contributing to the cognitive phenotypes in WS (49). Furthermore, the next most nominally significant SNP is rs61438591, an intronic variant in the *GTIF2RD1* gene, another gene highly implicated in the cognitive and social phenotypes seen in WS(92, 99–101).

Since the common variants in WSCR showed no association, we wanted to test for the possibility that rare variants could contribute to the variability in SRS T-scores. To test this, we used SKAT-O, which tests all variants in the region at once and weights each variant by its minor allele frequency. Similarly, we included age, sex, and ancestry as covariates. We tested each gene in the WSCR independently, because we hypothesized only certain genes in the region, such as *STX1A*, *LIMK1*, *CYLN2*, *BAZ1B, GTF2IRD1* (31, 38, 49, 96, 99, 106–109) that have been implicated in the cognitive phenotypes would contribute to the social phenotype rather than the entire region. While no gene p-value survives multiple testing corrections, the *ELN* gene has the most nominally significant p-value of 0.013

The results of our analysis of variation in the WSCR suggest that common and rare variants in the remaining allele do not strongly influence social behavior in WS. This does not exclude the possibility that a second deleterious hit or common variation in other genes outside the region contributes to the variation in the SRS T-scores. To test this, we next examined variation in 71 genes known to be associated with autism spectrum disorders(121). These genes should be enriched for loci that affect social behavior and genetic variation in these genes could contribute to variability seen in WS. We called 1,367 variants in the 71 genes (**Supplemental Table S2**). We annotated the variants as above, with clinical significance and measures of deleteriousness compiled in dbNSFP. There are 313 (22.9%) variants that had at least one submission to ClinVar. None of these variants had previous evidence to support pathogenicity. There are 33 missense variants predicted to be deleterious by MetaLR that are seen in 36 individuals in our sample. Despite having a putatively deleterious variant the distribution of SRS T-scores is similar between individuals either carrying or lacking deleterious variants in these genes ($t_{82.999}$=0.6878, p-value=0.4935). There are seven variants that should result in a truncated protein, one stop gain in the USP45 gene and six frameshift mutations. Only one sample harboring one of these mutations has a severe SRS T-score of 77. All of these protein-truncating mutations are also observed in the ExAC cohort.

We next tested for associations of each of the 381 common variants (MAF> 0.05) in these genes. No SNP was significant after multiple testing corrections (**Figure 3B**). The top five SNPs are located in Table II. Since each of these genes has been associated with ASD, we hypothesized that rare and common variants in each of the genes could contribute to SRS. We performed SKAT-O on the variants located in the autosomal ASD genes altogether, which also showed that there is little evidence to support variants in these 68 ASD genes have a strong effect on SRS T-scores, p=0.431

While it would be underpowered for any but the largest effect sizes (**Figure 2A**), for thoroughness we did an unbiased scan of the whole exome. We also examined the polygenic contribution of common variants to the SRS. The common variant analysis was performed on

66,620 variants (**Figure 3C**). The most nominally significant single SNP is rs527221 located in the *DMPK* gene, which is responsible for causing type 1 myotonic dystrophy (129) (**Table II**). While there is suggestive evidence for single variants such as rs527221, we calculated the polygenic risk scores (PRS) for each of the individuals in our sample to test if exome wide there are many SNPs of small effect that contribute to the social phenotype in WS. We used the summary statistics from the most recent PGC GWAS on autism spectrum disorders to calculate the PRS for our sample(130). We reasoned the polygenic risk of autism would be correlated with the SRS because this is a questionnaire used to assess behaviors that are affected by autism. Variants from the PGC GWAS were included if the p-value for the variant was under the threshold determined by the high resolution screen in the PRSice software(131). Interestingly, only the PRS for the motivation sub score was nominally significant (p=0.033), but after permutation to determine an empirical p-value it was not significant (p=0.308). The correlations of the PRS for each of the samples and the sub score as well as total SRS are shown in supplemental figure 1. Counterintuitively, there is a negative correlation between the PRS and motivation sub score. While this is the largest correlation between the PGS and sub scores it implies that more genetic risk for autism leads to a lower and less impaired social motivation T-score. However, given the small sample size and small number of SNPs available from whole exome sequencing compared to whole genome genotyping we are wary of making strong conclusions from this analysis.

We and others (20) have shown that individual sub scores of the SRS are affected differently by the deletion of the WSCR. Therefore, we wanted to rule out the possibility that variants are indeed affecting specific sub scales of social behavior, but that testing the total SRS score is masking those effects. Thus, in an exploratory manner, we repeated the quantitative trait loci associations for each of the sub scores of the SRS using the variants in the WSCR, 71 ASD genes, and the remaining whole exome variants. Since the sample size is small we conducted these associations for exploratory and hypothesis generating purposes. The top five SNPs from each association are reported in supplemental tables 3-5. For each of the analyses we see similar variants showing the highest association as were associated with the total SRS, likely due to the high correlation between the SRS and the sub scores (**Supplemental Figure 2**). Thus, an analysis of the total SRS was not masking independent genetic effects on each sub scale.

2.4 Discussion

Phenotypic variability has been appreciated in many of the symptom domains of WS including the cardiovascular phenotypes, the unique cognitive profile, and in social behavior(132–134). Here, we have described the variability of reciprocal social behavior in a sample of 85 individuals with the typical WS deletion using the SRS-2. Our results replicate the findings of Klein-Tasman et al. 2010, revealing that overall individuals with WS have SRS scores that are shifted to the more socially impaired end of the distribution, with most problems relating to the social cognition and restricted and repetitive behavior sub scales of the SRS while social motivation is spared.

We also note that sex differences in the general population have been reported previously in the literature for SRS. These sex differences were not consistent with different genetic factors contributing to the SRS in boy and girls, but due to discrepant effects of common genetic and environmental factors on SRS, such as differences in sensitivity to environmental factors or the X-inactivation phenomenon (119). However, we do not see evidence of sex effects in our sample of individuals with WS. The magnitude of the difference between males and females in our sample is similar to what was reported in the general population, so our lack of a significant finding could be due to our small sample size. The standard deviation of the SRS is large in both the general population and still larger in the WS population, so it may also be that larger sample sizes are needed to overcome the considerable variance in the data. The fact that the WS population has a larger standard deviation could also suggest that individuals with the deletion are sensitized to other factors that contribute to variation in the SRS such as background genetic variation or environmental factors.

We performed whole exome sequencing on our sample of 85 individuals to test for additional genetic contributions to the variability seen in social behavior in individuals with WS. We used the identified variants to test the hypothesis that genetic variation in the remaining WSCR allele can explain some of the variability in SRS T-scores. Genes in this region have a dosage sensitive effect on social behavior evidenced from the contrasting social phenotypes of the WS deletion and the reciprocal duplication, suggesting that variants in the remaining WSCR allele that affect expression or function of the genes could further contribute to the social phenotype(13). We called 120 variants in the WSCR with 55 variants being exonic. We used evidence such as the amino acid change, clinical significance suggested by the ClinVar database, and multiple algorithms to predict the consequences of the variants. Within the WSCR we do not find any variants that cause protein truncation. None of the missense variants are predicted to be deleterious based on the MetaLR composite score. Of the nine variants that have been submitted to ClinVar, all were described as benign or likely benign. A quantitative trait association analysis using the common variants in the region resulted in no SNP that survived multiple testing corrections. The most significant SNP, rs2074754, is a synonymous SNP in the *BAZ1B* gene. This gene encodes for a protein product in the bromodomain protein family that modifies chromatin to affect transcription and has been implicated in the cognitive phenotypes in WS.

Knocking down this gene in human derived induced pluripotent stem cells upregulates genes involved in mitosis as well as downregulating genes that are involved in the development of the nervous system(49) The second most nominally significant SNP, rs61438591, is an intronic variant in *GTF2IRD1*, which encodes for a transcription factor that has been suggested to contribute to the cognitive and social behavior deficits (38, 39, 49, 92, 100, 101). If future studies with increased power replicate this association, it would suggest that noncoding variation, perhaps controlling the expression of this gene, might contribute to variation is social behavior. We also tested the association of all variants in the WSCR using SKAT-O. This test indicated no variants with sufficient effect size were detected in the WSCR.

While we have not shown evidence that variants in the remaining WSCR contribute to the social phenotype in WS, we cannot conclusively discard this hypothesis. However, our study does clearly indicate that the alleles genotyped here are either not causative or exert too small an effect size on SRS for our current power (**Figure 2**), but it does not rule out variants of small effect on social behavior in the region. Research on other copy number variants associated with ASDs showed that larger CNVs tended to have genes of smaller individual effect size and suggests the phenotype of the overall CNV is due to the cumulative effect of each of those genes(121). Further we did not detect any variants in the gene *GTF2I*, which has been highly suspected of contributing to the social behaviors in WS(31, 38, 53, 90, 96). The lack of variant calls in our sample could be due to the fact that *GTF2I* is under stringent purifying selection. Indeed, looking at the ExAC data covering this gene, they show that there are fewer missense variants than expected by chance. ExAC discovered 62 synonymous and 56 missense mutations in 60,706 people(135). In our sample of 85 individuals we would expect to see variants in ExAC that have an allele frequency of greater than 0.0059, which is an allele count of one in our sample. There are ten variants with an allele frequency greater than 0.0059 detected in ExAC, only three of which are exonic. Thus, we would need a much larger sample size to investigate coding variants in *GTF2I.* The two linked variants in *GTF2I* that have previously been associated with oxytocin responsiveness and amygdala reactivity, rs1322743 and rs4717907, are intronic and were not covered in our sequencing(51, 118).

We further used the genetic data to investigate the role of variation in 71 genes that have been associated with ASD. WS and ASD do show phenotypic overlap(114, 136), and we reasoned that these genes should be enriched for functional roles in social behaviors. Likewise, the presence of outlier scores on the SRS that indicated severe impairment, suggested there could be possible second deleterious hits on top of the WS deletion in our dataset. Second hits are expected to be rare but have been observed in WS to explain a case of a child with comorbid seizures(54). Inspecting the 1,367 variants discovered in the ASD genes, 313 variants have been previously submitted to ClinVar, none of which show evidence for any pathogenicity. We observed seven protein-truncating mutations that do not associate with severe SRS T-scores. Several missense mutations were predicted to be deleterious, but there was no association between individuals that had a putative deleterious variant and a more impaired SRS score. Testing the common and rare variants in these genes showed no associations with the social phenotype. Similar results were found when we performed the association analyses on all of the variants discovered in the cohort. The most significant SNP was rs527221, a nonsynonymous variant in the *DMPK* gene, which is responsible for causing type 1 myotonic dystrophy, severe childhood forms of which have been associated with ASD(137). We also tested if polygenic risk for increased ASD liability is associated with the SRS T-score and sub scores. This boosts our ability to detect the impact of many loci with small effects. The largest correlation was between the PRS and the social motivation sub score, although this was not significant.

WS seems to affect specific domains of social behavior as evidenced by significant differences between the sub scores of the SRS. This observation led us to an exploratory examination of associations with the sub scores of the SRS and test if different genetic variants contribute to each sub score. Overall using variants from the WSCR, ASD genes, or the whole exome identified the same variants as nominally significant. The SRS and the sub scores are very correlated, but the social motivation in the WS sample is the least correlated to all other scores. This reflects that fact that social motivation tends to be rated within the normal range in WS, while the other scores are often higher. Interestingly, the whole exome association on the motivation T score leads to the lowest FDR values compared to the other scores, suggesting that there may be more genetic signal when using this sub scale. Indeed, this decoupling of the social motivation subscale from other SRS items highlights the possibility that the social motivation subscale might provide useful clinical information going forward; individuals carrying the WSCR deletion yet not showing a spared social motivation might warrant a deeper examination for additional factors impacting their presentation.

There are several limitations to our current study that should be addressed in future research. First the current study genotyped and assessed only the probands and not their parents. Having genetic information from trios would allow us to distinguish between variants that are inherited or *de novo*, which would aid in interpretation and prioritization of variants. Further, being able to compare the SRS score of the individual with WS to biparental SRS mean would let us control for effects of background genetic variation(120). Second, we are limited to investigating exonic variation. While interpretation of exonic variants is more straightforward because they potentially disrupt coding sequences, and can aid in the detection of deleterious rare variants, we could be missing important regulatory information that is located in promoters or introns of genes. Third, we were not able to control for intellectual functioning of the individuals with WS. The SRS has been reported to not correlate with intellectual functioning(138), but Klein-Tasman et al. 2010 found significant negative correlations between intellectual functioning and the total SRS T-score when parents completed the report, but not when teachers completed the report. SRS values have been shown to be dependent on levels of expressive language, nonverbal IQ, and behavioral problems. A subset of SRS questions was selected to ameliorate these dependences(139). The short form of the SRS as well as other questionnaires that assess adaptive skills and social behaviors could be used in the future to provide supporting information about the social phenotype and underlying genetics in WS. Finally, while our study represents the largest single collection of WS samples reported to date, it is only powered to detect strong effects of common variants due to our small sample size. This is challenging to overcome due to the low prevalence of WS.

In conclusion, we have tested the hypothesis that variation in the remaining WSCR allele affects the social phenotype of individuals with WS, by applying whole exome sequencing to a sample of 85 individuals with typical WS deletions. We show that common and rare variants in the region do not associate with SRS T-scores in our sample. Further, we show that variation outside of the region does not account for the social variability. This is not to say that genetic variation does not play a significant role in phenotypic variability in WS, but that it will require larger sample size to detect. In the future, applying whole genome sequencing to a sample of individuals with WS might elucidate the roles of genetic variation in the regulatory elements. Whole genome data could also allow for more accurate breakpoint determination. Redundant sequences in the low copy number repeat areas at either end of the WS deletion prevent accurate end point detection by CMA. This will be an interesting avenue to pursue in order to investigate how deletion size variation among individuals with typical 1.5 to 1.8 MB deletions contributes to social behavior. For example, Porter *et al.* showed that those with larger (1.8Mb deletions) had decreased executive functions(41). It is also worth noting that the current genetic data set has additional clinical data available, which can be queried in the future for the presence of more substantial associations with other WS related phenotypes.

2.5 Materials and Methods

Ethical Compliance and samples

This study was conducted with approval of the IRBs at Washington University School of Medicine and the National Institutes of Health. Consent was obtained prior to inclusion in the study. Once enrolled, participants provided a DNA sample by blood or saliva and their care-givers filled out health related questionnaires. The 85 individuals that make up our sample have ages that range from 2.5 to 65.5 years with a mean of 16.1 years. Caregivers provided a self-reported ethnicity. The majority of the sample was reported as white (77 individuals). There are two individuals that are African American, three Chinese, and three others.

Confirmation of diagnosis

WS diagnosis and typical deletion size was confirmed using either chromosomal microarray or quantitative PCR. In some cases, clinical microarray results were derived from the medical record. Array type varied by individual. For the remaining individuals, some received a research array (Cytoscan HD, Applied Biosystems) with analysis using the accompanying ChAS software. Others underwent deletion size assessment using quantitative PCR for genes within and outside of the Williams region using Taqman copy number probes (Thermo-Fisher, AUTS2: Hs04984177_cn, CALN1: Hs04946916_cn, FZD9: Hs03649975_cn, CLIP2: Hs00899301_cn, HIP1: Hs00052426_cn, POM121C: Hs07529820_cn). Copy number analysis was done according to the manufacturer's instructions and output data analyzed using their Copy Caller software. All individuals were confirmed to have deletions that included the WSCR genes ELN, FZD9 and CLIP2, but did not include genes external to the typical deletion such as CALN, AUTS2, POM121C or HIP1 (data not shown).

Social Responsiveness Scale

The social responsiveness scale-2 (SRS) is a 65-item questionnaire that measures aspects of social interaction that make up the core symptoms of autism spectrum disorders. The output is a total raw score as well as a T-score that is adjusted for sex, age, and the relationship of the reporter to the proband. The total score is made up of the scores of five subcategories that are impaired in ASDs: social awareness (AWR), social cognition (COG), social motivation (MOT), social communication (COM), and behaviors typical of autism such as restricted interests and repetitive behaviors (RRB). The response to each question ranges from 1 (not true) to 4 (almost always true). The T-scores are binned into four groups: normal < 59, mild between 60 and 65, moderate between 66 and 75, and severe > 76. For this study, the age-specific (pre-school, school age, or adult) SRS-2 was completed by the participant's caregiver and analyzed as a T-score that is adjusted for sex, age, and the relationship of the reporter. We provide values from the general population that have been previously reported for comparison (119, 138).

Sequencing and Variant calling

Whole exome sequencing and alignment was performed at Washington University in St. Louis by the McDonnell Genome Institute on 85 DNA samples from individuals with WS. Exomes were captured using Nimblegen SeqCap EZ Choice HGSC Library version 2.1, which targets 45.1 Mbp covering 23,585 genes and 189,028 non-overlapping exons. Exomes were aligned to the GRCh37-lite genome using bwa –mem v0.7.10(140) default settings, samtools v0.1.19(141) was used to assign mate pairings, sort, and index the bam files. Duplicates were marked using Picard MarkDuplicates v1.113.

Variant calling was done following GATK best practices on the aligned exomes (142). Briefly, using GATK v3.6.0 indels were realigned and the base quality scores recalibrated. Variants were initially called per sample using the haplotype caller tool, followed by jointly calling variants. To improve variant calls, we recalibrated variants and used a truth sensitivity tranche of 97 for SNPs, and a truth sensitivity tranche of 94 for indels. These thresholds were chosen to maximize the number of known and novel variants while still being stringent enough to limit the number of false positive variant calls. To further filter the variants we used the VariantFiltration tool to filter variant sites that had lower than a 10x average coverage or an inbreeding coefficient less than -0.20 to remove sites with excess heterozygosity. Genotype calls were filtered and considered to be missing if they had a genotype quality score of less than 20, which refers to a 99% probability that the call is correct. Finally, using vcftools v0.1.14(143), we removed sites that had a genotype missing rate of greater than 10%, as well as sites that no longer showed any variation. This produced a call set of 202,820 variant sites. The final call set has a Ti/Tv ratio of 2.76 and a dbSNP rate of 88.5%. These metrics are consistent with quality variant calls and a low false positive rate.

Variant annotation

The variant call set was split into three groups using vcftools: 1) variants in the Williams syndrome critical region (WSCR) defined by hg19 coordinates chr7:72,395,660-74,267,841 2) variants located in 71 genes associated with ASD(121), and 3) the remaining non-overlapping variants. All sets include exonic variants as well as variants located in introns that are pulled down by the capture reagents. Bcftools v1.2(141) was used to split multiallelic sites into separate lines for each allele and left normalized so positions would be compatible with ANNOVAR annotation files version 2016-02-01(144). The ANNOVAR table_annovar.pl function was used to annotate all three variant call sets with the RefSeq gene annotation, variant consequence, ExAC allele frequency(135), sample specific allele frequency, dbsnp147 name, clinical significance assessed by ClinVar(145). Missense variants were also annotated with measures of deleteriousness compiled in dbNSFPv3.3a(146). We highlight the CADD PHRED score and MetaLR as two measures of variant deleteriousness. CADD scores are defined at each base in the genome and for every possible single nucleotide change(147). CADD scores compare 65 annotations, including functional data as well as conservation scores, between fixed human derived alleles and simulated variants. Deleterious variants should be depleted in the observed fixed alleles and not in the simulated variants. CADD PHRED scores represent the relative rank of a CADD score compared to all other possible allele CADD scores; a CADD score of 10 means this allele is ranked as the top 10% of all possible CADD scores. Larger CADD PHRED score indicates an increased predication of deleteriousness. MetaLR uses logistic regression to incorporate information from 9 other variant annotations that consider function as well as conservation(126). The model was trained on true deleterious variants and true neutral variants described in the Uniprot database. The composite MetaLR score was found to have greater predictive ability than any of the single scores that make up MetaLR.

Power Analysis

We performed a power analysis to provide the limits of genetic effects that we would be able to detect given our cohort size. For future studies we also calculate the sample sizes that would be needed to detect different magnitudes of genetic effects. We used the Genetic Power Calculator (148). We calculated the predicted power of the current sample size n=85 using a p-value threshold corresponding to the Bonferroni corrected alpha for each set of analyses (WSCR 34 variants, alpha=0.00147, ASD 381 variants, alpha=0.000131, WEX 66620 variants, alpha=$7.5x10^{-7}$. Our main hypothesis is variants on the remaining WSCR allele affect the social phenotype; we wanted to calculate the sample sizes that would be required to detect different size genetic effects in the WSCR at different levels of power. We again used the alpha threshold based on the 34 common variants we identified in the exons of the WSCR and report the sample size required to achieve a specific power.

Association analyses

Common variant analysis

The variant call files were converted to plink binary bed format using the GATK tool VariantToBinaryPed. We used PLINK v1.9(149) --linear option to conduct a quantitative trait association using the SRS T-score as the quantitative trait. Ancestry was controlled for by including the first four principle components, determined by the --pca function in PLINK, as covariates along with sex and age. We used alleles that had a minor allele frequency (MAF) of 0.05 or greater. We performed the association analyses on the three separate groups of variants described in the previous section. It should be noted that allele frequency in the Williams syndrome critical region is inflated because of the hemizygous state of the region in individuals with WS. A MAF of 0.05 in this region corresponds to an allele count of four. In all cases we report the effect size of a variant under an additive model. Though the small sample size of this study limits power, in an exploratory fashion we also performed the same quantitative trait analysis on each of the sub scores of the SRS using variants in the WSCR, ASD genes, and the whole exome.

SKAT-O

SKAT-O (150)was implemented in the R v3.1.3 environment. SKAT-O fits a multiple linear regression of all SNPs located in a user provided region. The framework in SKAT-O allows for correlation between SNPs in a region, where if all SNPs are perfectly correlated this would become a burden test, but also allows SNPs in the same region to have effects in opposite directions. Significance is assessed by region rather than by SNP. We considered each gene that harbors a variant in the WSCR as a separate region for a total of 26 regions. To test for an overall effect of variants in the ASD genes we collapsed the 61 autosomal genes into one region. We used the beta function shape parameters (1,50) to put more weight on SNPs that have lower minor allele frequency, reasoning that rare causal alleles potentially have a greater effect size. We again controlled for age, sex, and the first four principal components.

Polygenic Risk Score

Polygenic Risk Scores (PRS) can be used to test if there is a contribution of many loci of small effect on the phenotype of interest by summing the effects of variants that may have not reached genome-wide significance. For a discovery set, we used the publically available summary statistics from the most recent Psychiatric Genome Consortium genome wide association study (GWAS) of autism spectrum disorder (130), reasoning that genetic risk for autism would contribute to SRS scores. The best-fit PRS was determined using the high-resolution functionality in the PRSice software(131). All of the variants identified throughout the exome with a MAF >0.05 and that are also present the in the discovery set were used to calculate the PRS. Sex, age, and the first four PCs were included as covariates. After clumping there were a total of 23,191 variants used to calculate the PRS. PRSice was used to calculate the significance of the PRS at the best-fit p-value threshold using 10000 permutation to determine an empirical p-value. PRS for each of the samples was calculated for the total SRS T-score as well as the sub scores.

Other statistical analyses

All remaining statistical tests were done in the R v3.1.3 environment. Two sample t-tests were used to compare the means of two groups. ANOVA was used to test differences in mean of sub scales of SRS. TukeyHSD post hoc comparison was performed using the multcomp package. The qqman(151) package was used to generate manhattan and qq plots.

2.6 Acknowledgments

None of the authors have any conflicts of interest to report. We would like to thank Drs. Natasha Marrus and John Constantino for helpful discussions of the SRS, and Dr. Don Conrad for discussion of association analyses. We acknowledge the McDonnell Genome Institute at Washington University who performed the exome sequencing performed in this study. NDK was supported by the Autism Science Foundation and the NSF (DGE-1143954), This work was supported by the Children’s Discovery Institute(MD-II-2013-269 to JDD and BAK and CH-FR-2011-169 to BAK) and the NIH (1R01MH107515-01A1, 5R01HG008687-02, 9R01MH100027-06 for JDD and HL006212-01 for BAK). JDD is an investigator of the Brain and Behavior

Research Foundation. BAK received funding for this project from the Heartland Genetic Services Collaborative. We also thank the Williams syndrome association and the families who participated in this study.

2.7 Figures

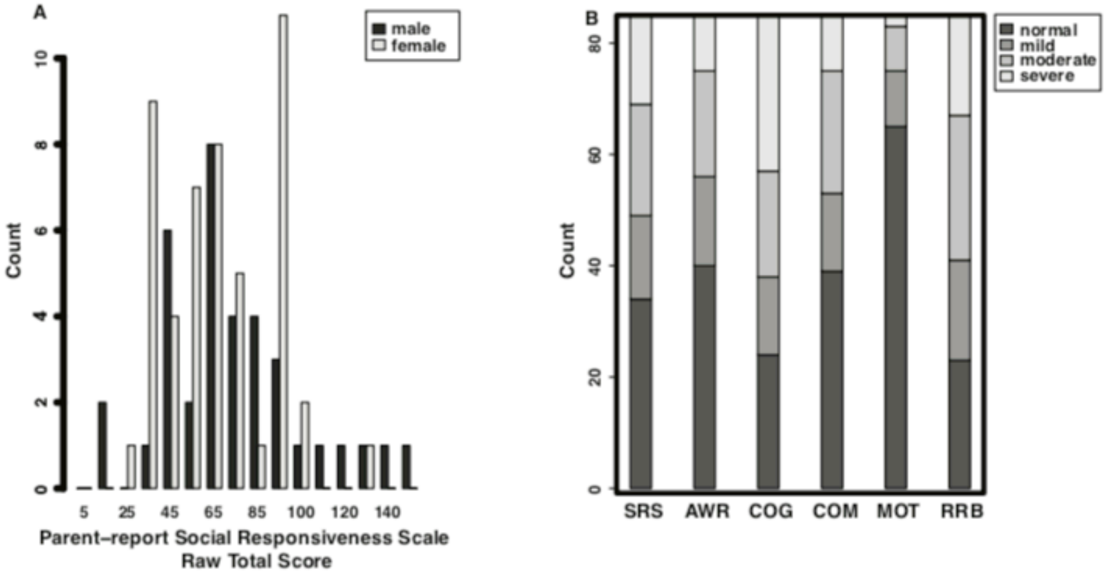


Figure 1: Distribution of Social Responsiveness in 85 individuals with typical WS deletion. A Distribution of the raw SRS scores **B** Severity bins of SRS and subcategory scores.

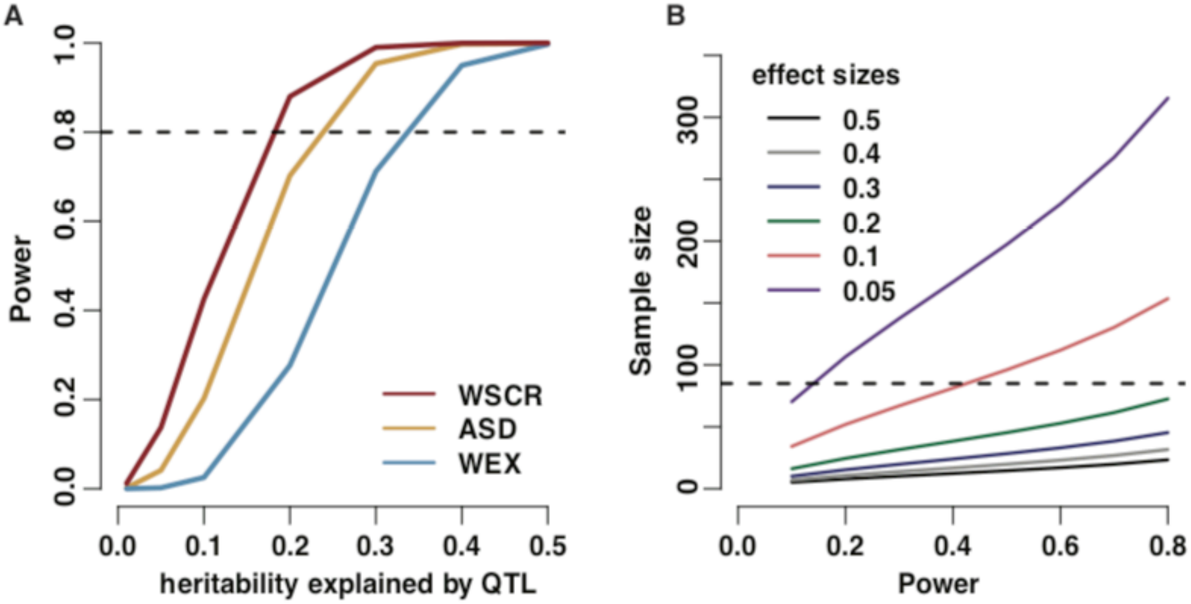


Figure 2: Power analysis. **A**The power to detect variants of different effect sizes for the current study. The alpha for the three different sets of analyses was determined by using the Bonferroni correction based on the number of SNPs tested in each analysis. (WSCR: variants in the WSCR, ASD: variants in the 71 ASD genes, WEX: all remaining variants exome wide). **B** The predicted sample sizes that would be required to achieve different levels of power for detecting variants of different effect sizes. The sample size predictions were only done using the alpha for the number of SNPs tested in the WSCR. The horizontal dashed line indicates the sample size of the current study.

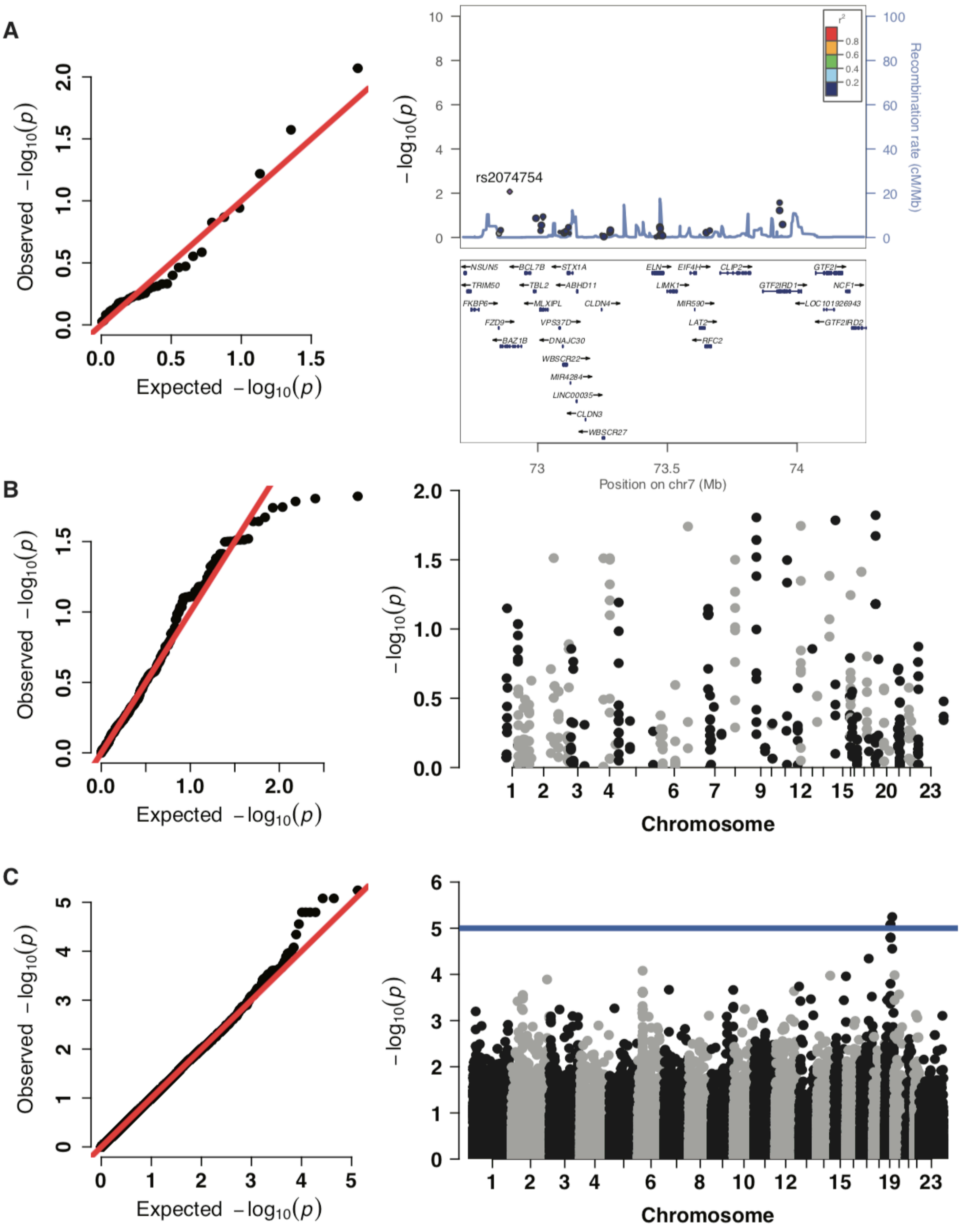
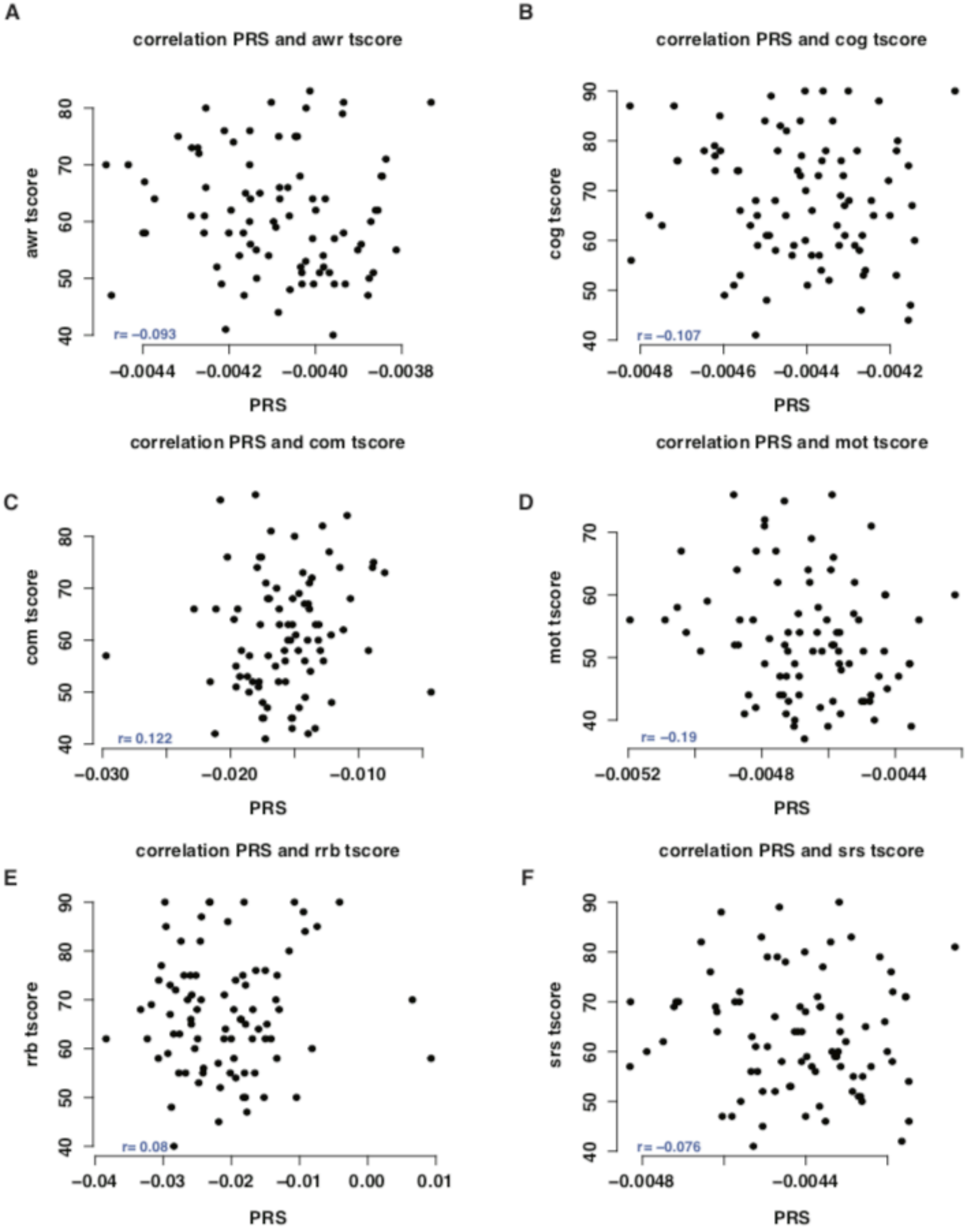
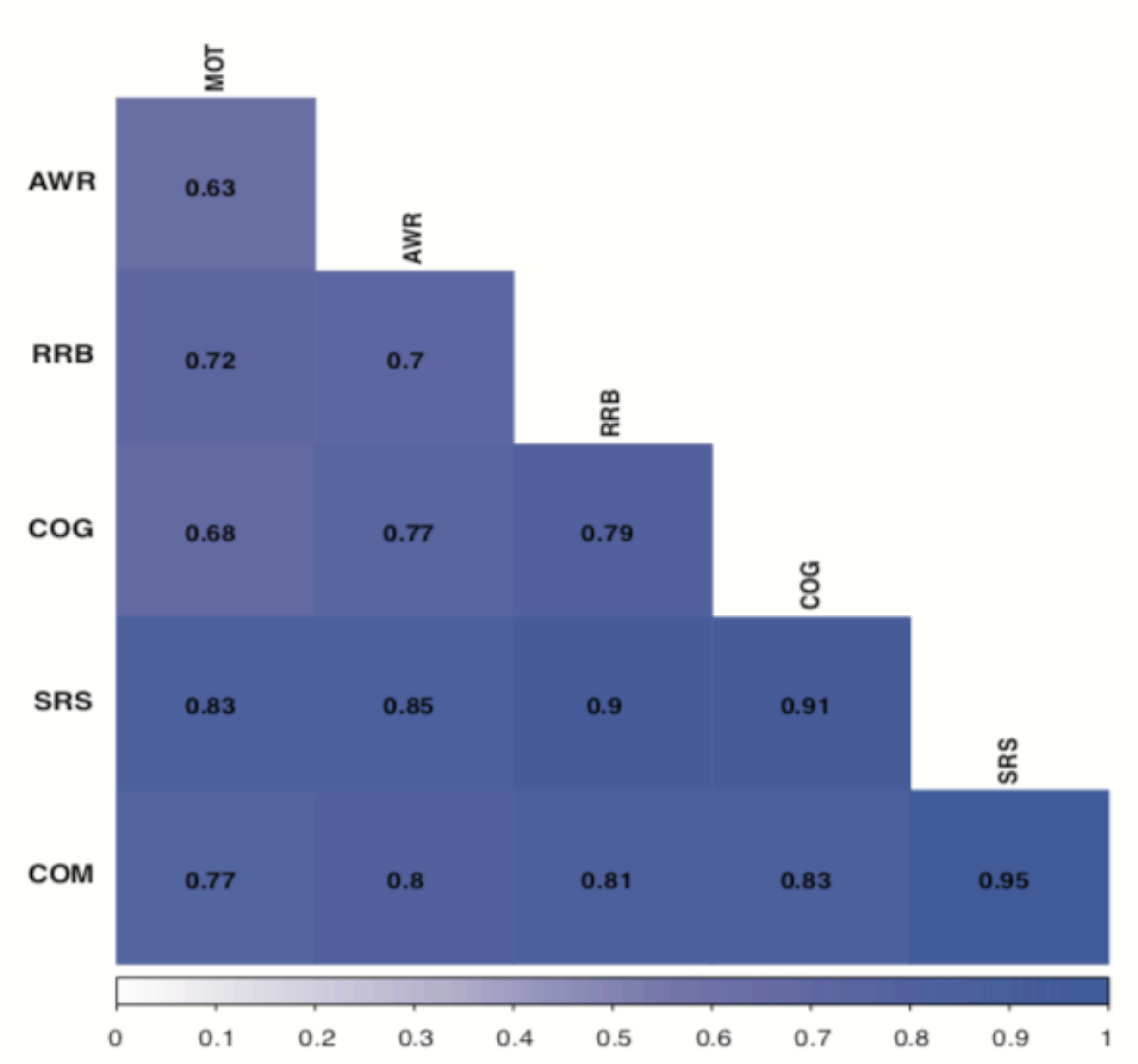


Figure 3: Variants in the WSCR, ASD genes, or whole exome do not contribute to SRS variability in a sample of WS with typical deletions. A qq plot showing distribution of p-values for common variants in the WSCR. Locus zoom plot showing the SNPs tested in the WSCR, highlighting the most nominally significant SNP in *BAZ1B*. **B** qq

and manhattan plot for variants called in 71 genes associate with ASD from Sanders *et al.* 2015. **C** qq and manhattan plot for variants exome wide. Blue line demarcates a suggestive p value threshold of $1x10^{-5}$.



Supplemental Figure 1: Polygenic Risk Score correlation with SRS and SRS subscores. A-F Panels show the correlation between the polygenic risk score (PRS) for the sub scores of the SRS calculated using variants from the PGC ASD GWAS that fall below the p-value threshold calculated from the best-fit PRS. Pearson correlation values between the samples PRS and the SRS subscore shown as the inset.



Supplemental Figure 2: SRS and sub scales are correlated. Heatmap display of the Pearson correlation values of the SRS and sub scale T-scores in 85 individuals with WS. Values of the correlation are labeled in the plot.

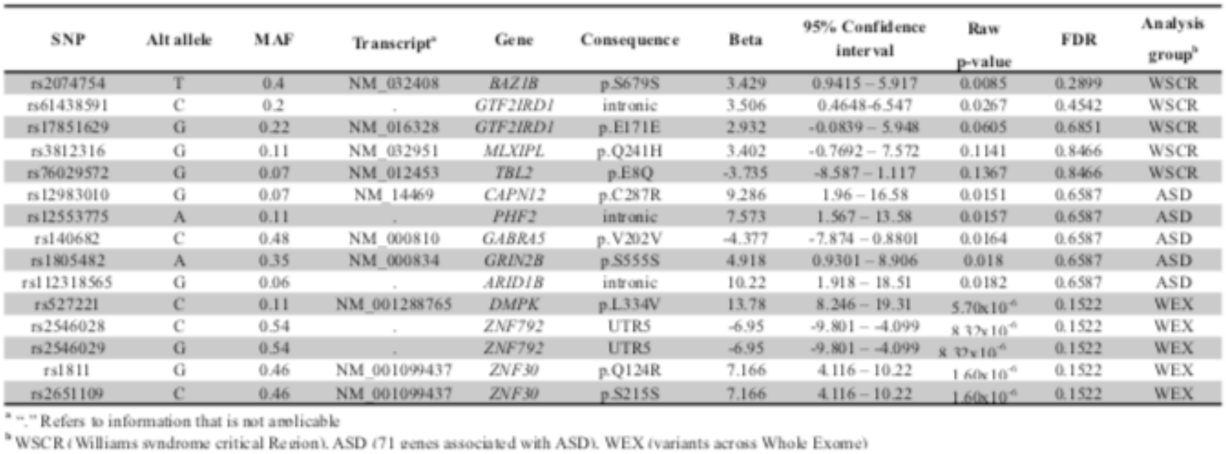
Table 1: Annotation of 55 exonic variants discovered in the WSCR

Chr	Start	avsnp147[a]	Alt	MAF	Transcript	Gene	Consequence	MetaLR score	MetaLR Prediction[b]	CADD PHRED
7	72413057	rs782618986	A	0.005882	NM_172020	POM121	p.S577N	0.011	T	0.006
7	72717686	rs145622470	T	0.01176	M_00116834	NSUN5	p.P399P	.	.	8.726
7	72719048	rs34913552	A	0.01176	M_00116834	NSUN5	p.P183S	0	T	0.002
7	72738534	rs371073794	T	0.01176	M_0012814	TRIM50	p.P84P	.	.	15.11
7	72738561	rs61741334	T	0.04706	M_0012814	TRIM50	p.I75I	.	.	11.01
7	72738762	rs6980258	T	0.9882	M_0012814	TRIM50	p.L8L	.	.	0.46
7	72738763	rs6980124	G	0.9882	M_0012814	TRIM50	p.L8P	.	.	0.001
7	72744246	rs200493820	T	0.01176	M_0012813	FKBP6	p.T90M	0.492	T	13.74
7	72754645	rs56301507	A	0.01176	M_0012813	FKBP6	p.L168L	.	.	3.802
7	72856676	rs1178978	T	0.01176	NM_032408	BAZ1B	p.Q1434Q	.	.	11.69
7	72857130	rs150115317	T	0.01176	NM_032408	BAZ1B	p.R1340K	0.025	T	23.6
7	72891754	rs2074754	T	0.4	NM_032408	BAZ1B	p.S679S	.	.	10.13
7	72936183	.	A	0.01176	NM_032408	BAZ1B	p.H27H	.	.	2.032
7	72951640	rs142166738	G	0.01176	M_00119724	BCL7B	p.A142A	.	.	7.437
7	72985148	rs35607697	T	0.03529	NM_012453	TBL2	p.V345I	0.014	T	26.3
7	72987758	.	C	0.01176	NM_012453	TBL2	p.F164V	0.154	T	27.3
7	72992858	rs76029572	G	0.07059	NM_012453	TBL2	p.E8Q	0.054	T	9.196
7	73010754	rs61738649	T	0.05882	NM_032951	MLXIPL	p.L626L	.	.	2.706
7	73013901	rs13235543	T	0.1294	NM_032951	MLXIPL	p.P342P	.	.	6.53
7	73020301	rs799157	C	0.9647	NM_032951	MLXIPL	p.S253S	.	.	2.151
7	73020337	rs3812316	G	0.1059	NM_032951	MLXIPL	p.Q241H	0.001	T	19.07
7	73020439	rs12539160	T	0.01176	NM_032951	MLXIPL	p.A207A	.	.	12.68
7	73083889	rs61743139	T	0.02353	M_00107762	VPS37D	p.A93A	.	.	18.4
7	73097082	rs79849491	G	0.02353	NM_032317	DNAJC30	p.F224F	.	.	0.66
7	73097238	rs1569062	A	0.3294	NM_032317	DNAJC30	p.Y172Y	.	.	11.69
7	73122977	rs2229854	A	0.05882	M_00116590	STX1A	p.N50N	.	.	11.25
7	73150934	rs138932141	A	0.01176	M_00114536	ABHD11	p.D244D	.	.	1.115
7	73245591	rs142910620	T	0.01176	NM_001305	CLDN4	p.A20A	.	.	17.87
7	73254812	rs13241921	C	0.7882	NM_152559	WBSCR27	p.Q107R	0	T	0.001
7	73275565	rs11770052	A	0.7647	NM_182504	WBSCR28	p.I14N	0	T	15.45
7	73279361	rs61742124	T	0.1294	NM_182504	WBSCR28	p.L37L	.	.	14.82
7	73279413	rs118088869	T	0.03529	NM_182504	WBSCR28	p.R55W	0.01	T	15.49
7	73280020	rs1136647	T	0.7176	NM_182504	WBSCR28	p.T205T	.	.	12.3
7	73466285	rs6979788	G	0.01176	M_00127891	ELN	p.A271A	.	.	1.511
7	73470714	rs2071307	A	0.4706	M_00127891	ELN	p.G412S	0	T	6.674
7	73474268	rs200512332	T	0.01176	M_00127891	ELN	p.V408V	.	.	9.149
7	73474367	rs61734584	A	0.01176	M_00127891	ELN	p.G441G	.	.	1.008
7	73474825	rs17855988	C	0.07059	M_00127891	ELN	p.G500R	0.007	T	23.2
7	73477524	rs140425210	A	0.01176	M_00127891	ELN	p.G529S	0.131	T	23.7
7	73631177	rs144269935	G	0.02353	NM_014146	LAT2	p.I39M	0.013	T	25.9
7	73651743	rs3135688	C	0.01176	M_0012787	RFC2	p.V160V	.	.	8.01
7	73663362	rs1805395	C	0.05882	M_0012787	RFC2	p.E3E	.	.	7.454
7	73731906	rs148561130	T	0.02353	NM_003388	CLIP2	p.P10P	.	.	18.78
7	73811479	rs76865959	C	0.01176	NM_003388	CLIP2	p.R897R	.	.	4.969
7	73814702	rs17145468	A	0.03529	NM_003388	CLIP2	p.D926E	0.006	T	17.3
7	73814749	rs2522943	C	0.9647	NM_003388	CLIP2	p.R942P	0	T	18.33
7	73929826	rs111256098	T	0.01176	M_00119920	GTF2IRD1	p.G139G	.	.	12.93
7	73932488	rs112098981	G	0.01176	M_00119920	GTF2IRD1	p.A179A	.	.	9.272
7	73932494	rs145535993	T	0.02353	M_00119920	GTF2IRD1	p.V181V	.	.	10.27
7	73932560	rs17851629	G	0.2118	M_00119920	GTF2IRD1	p.E203E	.	.	9.058
7	73933793	rs148463467	T	0.01176	M_00119920	GTF2IRD1	p.V252V	.	.	14.93
7	73944095	rs61744518	T	0.02353	M_00119920	GTF2IRD1	p.P406P	.	.	16.53
7	73944185	rs2240357	C	0.2353	M_00119920	GTF2IRD1	p.Y436Y	.	.	0.434
7	73953017	rs55634982	T	0.01176	M_00119920	GTF2IRD1	p.S517S	.	.	14.02
7	74211576	rs587728502	C	0.01176	NM_173537	GTF2IRD2	p.M759V	0.021	T	0.893

[a] "." Refers to information that is not applicable

[b] "T" the missense mutation is predicted to be Tolerated

Table 2: Top five SNPs from quantitative trait locus associations



SNP	Alt allele	MAF	Transcript[a]	Gene	Consequence	Beta	95% Confidence interval	Raw p-value	FDR	Analysis group[b]
rs2074754	T	0.4	NM_032408	*BAZ1B*	p.S679S	3.429	0.9415 – 5.917	0.0085	0.2899	WSCR
rs61438591	C	0.2	.	*GTF2IRD1*	intronic	3.506	0.4648-6.547	0.0267	0.4542	WSCR
rs17851629	G	0.22	NM_016328	*GTF2IRD1*	p.E171E	2.932	-0.0839 – 5.948	0.0605	0.6851	WSCR
rs3812316	G	0.11	NM_032951	*MLXIPL*	p.Q241H	3.402	-0.7692 – 7.572	0.1141	0.8466	WSCR
rs76029572	G	0.07	NM_012453	*TBL2*	p.E8Q	-3.735	-8.587 – 1.117	0.1367	0.8466	WSCR
rs12983010	G	0.07	NM_14469	*CAPN12*	p.C287R	9.286	1.96 – 16.58	0.0151	0.6587	ASD
rs12553775	A	0.11	.	*PHF2*	intronic	7.573	1.567 – 13.58	0.0157	0.6587	ASD
rs140682	C	0.48	NM_000810	*GABRA5*	p.V202V	-4.377	-7.874 – 0.8801	0.0164	0.6587	ASD
rs1805482	A	0.35	NM_000834	*GRIN2B*	p.S555S	4.918	0.9301 – 8.906	0.018	0.6587	ASD
rs112318565	G	0.06	.	*ARID1B*	intronic	10.22	1.918 – 18.51	0.0182	0.6587	ASD
rs527221	C	0.11	NM_001288765	*DMPK*	p.L334V	13.78	8.246 – 19.31	5.70×10^{-6}	0.1522	WEX
rs2546028	C	0.54	.	*ZNF792*	UTR5	-6.95	-9.801 – -4.099	8.32×10^{-6}	0.1522	WEX
rs2546029	G	0.54	.	*ZNF792*	UTR5	-6.95	-9.801 – -4.099	8.32×10^{-6}	0.1522	WEX
rs1811	G	0.46	NM_001099437	*ZNF30*	p.Q124R	7.166	4.116 – 10.22	1.60×10^{-6}	0.1522	WEX
rs2651109	C	0.46	NM_001099437	*ZNF30*	p.S215S	7.166	4.116 – 10.22	1.60×10^{-6}	0.1522	WEX

[a] "." Refers to information that is not applicable

[b] WSCR (Williams syndrome critical Region), ASD (71 genes associated with ASD), WEX (variants across Whole Exome)

Supplemental Table S1: Annotation of 120 variants discovered in the Williams syndrome critical region

Chr	Start	Alt	avsnp147	Genic location	Gene	Consequence[a]	AA change	Sample freq	ExAC freq	MetaLR[b] prediction	CADD PHRED
7	72409868	G	rs189678402	intronic	POM121	.	.	0.01176	0.0034	.	.
7	72413057	A	rs782618986	exonic	POM121	nonsynonymous SNV	p.S577N	0.005882	0.0625	T	0.006
7	72717686	T	rs145622470	exonic	NSUN5	synonymous SNV	p.P399P	0.01176	0.0021	.	.
7	72718187	A	rs147531105	intronic	NSUN5	.	.	0.02353	0.0025	.	.
7	72719048	A	rs34913552	exonic	NSUN5	nonsynonymous SNV	p.P183S	0.01176	0.0105	T	0.002
7	72722565	G	rs199740800	intronic	NSUN5	.	.	0.01176	0.0006	.	.
7	72722836	C	rs142091726	UTR5	NSUN5	.	.	0.01176	0.0053	.	.
7	72732712	C	rs192182316	intronic	TRIM50	.	.	0.01176	0.0022	.	.
7	72732754	C	rs532548355	intronic	TRIM50	.	.	0.01176	0.0004	.	.
7	72732785	T	rs183981056	intronic	TRIM50	.	.	0.02353	0.0175	.	.
7	72738534	T	rs371073794	exonic	TRIM50	synonymous SNV	p.P84P	0.01176	4.96E-05	.	.
7	72738561	T	rs61741334	exonic	TRIM50	synonymous SNV	p.I75I	0.04706	0.0204	.	.
7	72738762	T	rs6980258	exonic	TRIM50	synonymous SNV	p.L8L	0.9882	0.9993	.	.
7	72738763	G	rs6980124	exonic	TRIM50	nonsynonymous SNV	p.L8P	0.9882	0.9993	.	.
7	72743316	A	rs73131580	intronic	FKBP6	.	.	0.03529	0.0244	.	.
7	72744143	G	rs3950375	intronic	FKBP6	.	.	0.01176	0.0108	.	.
7	72744246	T	rs200493820	exonic	FKBP6	nonsynonymous SNV	p.T90M	0.01176	0.0001	T	13.74
7	72754645	A	rs56301507	exonic	FKBP6	synonymous SNV	p.L168L	0.01176	0.0622	.	.
7	72756785	A	rs55704260	intronic	FKBP6	.	.	0.01176	0.0171	.	.
7	72850178	C	rs1178947	UTR3	FZD9	.	.	0.2	.	.	.
7	72850295	T	rs1178946	UTR3	FZD9	.	.	0.01176	.	.	.
7	72850305	A	rs113683726	UTR3	FZD9	.	.	0.01176	.	.	.
7	72856676	T	rs1178978	exonic	BAZ1B	synonymous SNV	p.Q1434Q	0.01176	0.0038	.	.
7	72857049	G	rs1178977	intronic	BAZ1B	.	.	0.1882	0.1678	.	.
7	72857130	T	rs150115317	exonic	BAZ1B	nonsynonymous SNV	p.R1340K	0.01176	0.0027	T	23.6
7	72874088	A	rs799215	intronic	BAZ1B	.	.	0.01176	0.0039	.	.
7	72891754	T	rs2074754	exonic	BAZ1B	synonymous SNV	p.S679S	0.4	0.4452	.	.
7	72925046	A	rs377098092	intronic	BAZ1B	.	.	0.01176	8.24E-06	.	.
7	72936183	A	.	exonic	BAZ1B	synonymous SNV	p.H27H	0.01176	.	.	.
7	72951640	G	rs142166738	exonic	BCL7B	synonymous SNV	p.A142A	0.01176	0.0029	.	.
7	72985148	T	rs35607697	exonic	TBL2	nonsynonymous SNV	p.V345I	0.03529	0.04	T	26.3
7	72987758	C	.	exonic	TBL2	nonsynonymous SNV	p.F164V	0.01176	.	T	27.3
7	72992858	G	rs76029572	exonic	TBL2	nonsynonymous SNV	p.E8Q	0.07059	0.0552	T	9.196
7	73008330	A	rs72649011	intronic	MLXIPL	.	.	0.01176	0.003	.	.
7	73010754	T	rs61738649	exonic	MLXIPL	synonymous SNV	p.L626L	0.05882	0.0453	.	.
7	73011163	G	rs782188633	intronic	MLXIPL	.	.	0.01176	1.14E-05	.	.
7	73013901	T	rs13235543	exonic	MLXIPL	synonymous SNV	p.P342P	0.1294	0.1089	.	.
7	73019975	G	rs61010704	intronic	MLXIPL	.	.	0.2353	0.2325	.	.
7	73020301	C	rs799157	exonic	MLXIPL	synonymous SNV	p.S253S	0.9647	0.9643	.	.
7	73020337	G	rs3812316	exonic	MLXIPL	nonsynonymous SNV	p.Q241H	0.1059	0.1352	T	19.07
7	73020439	T	rs12539160	exonic	MLXIPL	synonymous SNV	p.A207A	0.01176	0.0404	.	.
7	73021654	T	rs200438567	intronic	MLXIPL	.	.	0.01176	0.0036	T	11.28
7	73030530	A	rs187002831	intronic	MLXIPL	.	.	0.02353	0.0027	.	.
7	73083889	T	rs61743139	exonic	VPS37D	synonymous SNV	p.A93A	0.02353	0.0144	.	.
7	73084309	C	rs7795181	intronic	VPS37D	.	.	0.2235	0.2649	.	.
7	73084316	A	rs185557423	intronic	VPS37D	.	.	0.01176	0.004	.	.
7	73097082	G	rs79849491	exonic	DNAJC30	synonymous SNV	p.F224F	0.02353	0.0326	.	.
7	73097238	A	rs1569062	exonic	DNAJC30	synonymous SNV	p.Y172Y	0.3294	0.3043	.	.
7	73101137	G	rs11769825	intronic	WBSCR22	.	.	0.3765	0.3342	.	.
7	73107003	A	rs2293490	intronic	WBSCR22	.	.	0.3412	0.3022	.	.
7	73108310	C	rs2293487	intronic	WBSCR22	.	.	0.3353	0.3024	.	.
7	73114829	A	rs45549734	intronic	STX1A	.	.	0.02353	0.0147	.	.
7	73118033	A	rs35459363	intronic	STX1A	.	.	0.4706	.	.	.
7	73122977	A	rs2229854	exonic	STX1A	synonymous SNV	p.N50N	0.05882	0.0762	.	.
7	73150934	A	rs138932141	exonic	ABHD11	synonymous SNV	p.D244D	0.01176	0.0017	.	.
7	73245591	T	rs142910620	exonic	CLDN4	synonymous SNV	p.A20A	0.01176	0.0007	.	.
7	73246461	G	rs1127155	UTR3	CLDN4	.	.	0.7765	.	.	.
7	73246496	T	rs1127156	UTR3	CLDN4	.	.	0.7647	.	.	.
7	73246555	G	rs11316	UTR3	CLDN4	.	.	0.7647	.	.	.
7	73246727	-	.	UTR3	CLDN4	.	.	0.005882	.	.	.
7	73254812	C	rs13241921	exonic	WBSCR27	nonsynonymous SNV	p.Q107R	0.7882	0.6764	T	0.001
7	73275501	C	rs11770024	UTR5	WBSCR28	.	.	0.7765	0.6655	.	.
7	73275509	C	rs111714725	UTR5	WBSCR28	.	.	0.01176	0.0048	.	.
7	73275565	A	rs11770052	exonic	WBSCR28	nonsynonymous SNV	p.I14N	0.7647	0.6563	T	15.45
7	73279361	T	rs61742124	exonic	WBSCR28	synonymous SNV	p.L37L	0.1294	0.2019	.	.
7	73279413	T	rs118088869	exonic	WBSCR28	nonsynonymous SNV	p.R55W	0.03529	0.0257	T	15.49
7	73280020	T	rs1136647	exonic	WBSCR28	synonymous SNV	p.T205T	0.7176	0.5139	.	.
7	73449750	A	.	intronic	ELN	.	.	0.01176	.	.	.
7	73450948	C	rs186990808	intronic	ELN	.	.	0.02353	0.0021	.	.
7	73452140	A	rs2301995	intronic	ELN	.	.	0.02353	.	.	.
7	73457255	A	rs28763981	intronic	ELN	.	.	0.05882	.	.	.
7	73457506	T	rs55868272	intronic	ELN	.	.	0.02353	0.016	.	.
7	73466285	G	rs6979788	exonic	ELN	synonymous SNV	p.A271A	0.01176	0.001	.	.
7	73470714	A	rs2071307	exonic	ELN	nonsynonymous SNV	p.G412S	0.4706	0.3262	T	6.674
7	73470782	T	rs2856728	intronic	ELN	.	.	0.8353	0.7895	.	.
7	73472050	T	rs28763986	intronic	ELN	.	.	0.6	0.5933	.	.

7	73474268	T	rs200512332	exonic	ELN	synonymous SNV	p.V408V	0.01176	0.0002	.	.
7	73474367	A	rs61734584	exonic	ELN	synonymous SNV	p.G441G	0.01176	0.003	.	.
7	73474825	C	rs17855988	exonic	ELN	nonsynonymous SNV	p.G500R	0.07059	0.0712	T	23.2
7	73477524	A	rs140425210	exonic	ELN	nonsynonymous SNV	p.G529S	0.01176	0.0011	T	23.7
7	73477922	T	rs142870606	intronic	ELN	.	.	0.01176	0.0015	.	.
7	73480258	G	rs45618836	intronic	ELN	.	.	0.01176	0.0279	.	.
7	73480332	C	rs111866046	intronic	ELN	.	.	0.01176	0.0065	.	.
7	73481028	T	rs3757587	intronic	ELN	.	.	0.1059	0.1084	.	.
7	73530295	A	rs222996	intronic	LIMK1	.	.	0.02353	0.0108	.	.
7	73588650	A	rs531201818	upstream	EIF4H	.	.	0.01176	.	.	.
7	73588782	A	rs113057898	intronic	EIF4H	.	.	0.01176	0.0075	.	.
7	73605599	T	rs6971711	ncRNA_exonic	MIR590	.	.	0.01176	0.0067	.	.
7	73631177	G	rs144269935	exonic	LAT2	nonsynonymous SNV	p.I39M	0.02353	0.0113	T	25.9
7	73636045	C	rs201410958	intronic	LAT2	.	.	0.01176	0.0167	.	.
7	73636048	G	rs200399943	intronic	LAT2	.	.	0.01176	0.0167	.	.
7	73638035	G	rs112055519	intronic	LAT2	.	.	0.04706	0.0433	.	.
7	73649825	C	rs3135698	intronic	RFC2	.	.	0.05882	0.1042	.	.
7	73651743	C	rs3135688	exonic	RFC2	synonymous SNV	p.V160V	0.01176	0.0372	.	.
7	73653244	T	rs41552517	intronic	RFC2	.	.	0.01176	0.0258	.	.
7	73654225	T	rs73129384	intronic	RFC2	.	.	0.04706	0.0385	.	.
7	73657626	C	rs146804166	intronic	RFC2	.	.	0.02353	0.005	.	.
7	73663362	C	rs1805395	exonic	RFC2	synonymous SNV	p.E3E	0.05882	0.0306	.	.
7	73663451	A	rs41548312	UTR5	RFC2	.	.	0.01765	0.012	.	.
7	73664115	A	rs1805391	intronic	RFC2	.	.	0.01176	0.0237	.	.
7	73666835	G	rs1805393	intronic	RFC2	.	.	0.01176	0.0234	.	.
7	73666853	T	rs77326053	intronic	RFC2	.	.	0.01176	0.0234	.	.
7	73731906	T	rs148561130	exonic	CLIP2	synonymous SNV	p.P10P	0.02353	0.0402	.	.
7	73811479	C	rs76865959	exonic	CLIP2	synonymous SNV	p.R897R	0.01176	0.0057	.	.
7	73814702	A	rs17145468	exonic	CLIP2	nonsynonymous SNV	p.D926E	0.03529	0.03	T	17.3
7	73814749	C	rs2522943	exonic	CLIP2	nonsynonymous SNV	p.R942P	0.9647	0.9822	T	18.33
7	73929826	T	rs111256098	exonic	GTF2IRD1	synonymous SNV	p.G139G	0.01176	0.0193	.	.
7	73932488	G	rs112098981	exonic	GTF2IRD1	synonymous SNV	p.A179A	0.01176	0.0017	.	.
7	73932494	T	rs145535993	exonic	GTF2IRD1	synonymous SNV	p.V181V	0.02353	0.0103	.	.
7	73932560	G	rs17851629	exonic	GTF2IRD1	synonymous SNV	p.E203E	0.2118	0.2026	.	.
7	73932683	C	rs61438591	intronic	GTF2IRD1	.	.	0.2	0.1998	.	.
7	73933793	T	rs148463467	exonic	GTF2IRD1	synonymous SNV	p.V252V	0.01176	0.0001	.	.
7	73944095	T	rs61744518	exonic	GTF2IRD1	synonymous SNV	p.P406P	0.02353	0.0304	.	.
7	73944185	C	rs2240357	exonic	GTF2IRD1	synonymous SNV	p.Y436Y	0.2353	0.2462	.	.
7	73949411	C	rs59656369	intronic	GTF2IRD1	.	.	0.02353	0.0745	.	.
7	73953017	T	rs55634982	exonic	GTF2IRD1	synonymous SNV	p.S517S	0.01176	0.0064	.	.
7	73954167	C	rs782323873	intronic	GTF2IRD1	.	.	0.01176	0.0002	.	.
7	73971959	T	rs76184137	intronic	GTF2IRD1	.	.	0.04706	0.0321	.	.
7	73973183	A	rs73702616	intronic	GTF2IRD1	.	.	0.02353	0.0122	.	.
7	74211576	C	rs587728502	exonic	GTF2IRD2	nonsynonymous SNV	p.M759V	0.01176	0.001	T	0.893

[a] **“.”** data is not available or not applicable

[b] “T” Missense variants is predicted to be tolerated

Supplemental Table S2: Genetic variants in 71 genes associated with autism spectrum disorder

Chr	Start	Alt	avsnp147	genic location	gene	consequence	AA change	Sample_freq	ExAC freq	MetaLRpred	CADD PHRED
1	150239478	G	rs2275778	exonic	APH1A	synonymous SNV	p.G145G	0.01176	0.0223	.	.
1	150239722	A	rs140561586	intronic	APH1A	.	.	0.01176	0.0012	.	.
1	150240104	A	rs202225606	intronic	APH1A	.	.	0.005882	2.17E-05	.	.
1	150241230	T	rs2275780	UTR5	APH1A	.	.	0.09412	0.1662	.	.
1	151377407	C	rs1571294	exonic	POGZ	synonymous SNV	p.T1273T	0.05882	0.1051	.	.
1	151378214	T	rs116755407	exonic	POGZ	synonymous SNV	p.E1004E	0.005882	0.0014	.	.
1	151378274	A	rs149003420	exonic	POGZ	synonymous SNV	p.H984H	0.005882	0.0055	.	.
1	151379137	G	rs754254486	intronic	POGZ	.	.	0.005882	4.14E-05	.	.
1	151379337	A	rs559037025	intronic	POGZ	.	.	0.005882	8.46E-06	.	.
1	151379699	T	rs112834709	intronic	POGZ	.	.	0.005882	0.003	.	.
1	151379818	C	rs150592542	intronic	POGZ	.	.	0.01176	0.0074	.	.
1	151380736	G	rs762774439	intronic	POGZ	.	.	0.005882	3.32E-05	.	.
1	151381321	-	rs3831142	intronic	POGZ	.	.	0.7235	0.6753	.	.
1	151384258	AGG	rs201882243	intronic	POGZ	.	.	0.005882	0.0029	.	.
1	151384733	A	rs3748550	intronic	POGZ	.	.	0.7824	0.8001	.	.
1	151384734	A	rs184678605	intronic	POGZ	.	.	0.005882	0.0062	.	.
1	151395829	T	rs368660854	intronic	POGZ	.	.	0.005882	0.0001	.	.
1	151396037	C	rs2274534	intronic	POGZ	.	.	0.7235	0.6746	.	.
1	151396583	T	.	exonic	POGZ	synonymous SNV	p.E360E	0.005882	.	.	.
1	151400771	A	rs749391687	exonic	POGZ	synonymous SNV	p.T134T	0.005882	0.0002	.	.
1	151402045	G	rs6587577	intronic	POGZ	.	.	0.7824	0.7931	.	.
1	151413367	C	rs201418770	intronic	POGZ	.	.	0.005882	9.98E-05	.	.
1	151413613	G	rs115951766	intronic	POGZ	.	.	0.02941	0.0457	.	.
1	153636466	-	.	intronic	ILF2	.	.	0.07059	.	.	.
1	153636465	0	.	intronic	ILF2	.	.	0.005882	.	.	.
1	153636469	-	rs768585289	intronic	ILF2	.	.	0.05294	0.0278	.	.
1	153636469	-	rs369602600	intronic	ILF2	.	.	0.2941	0.2104	.	.
1	153636468	0	.	intronic	ILF2	.	.	0.005882	.	.	.
1	153636472	C	rs4515830	intronic	ILF2	.	.	0.05294	0.1647	.	.
1	153636472	ACC	rs527872479	intronic	ILF2	.	.	0.005882	0.0142	.	.
1	153636472	0	.	intronic	ILF2	.	.	0.3294	.	.	.
1	153636860	G	rs4351684	intronic	ILF2	.	.	0.5059	0.5939	.	.
1	153636865	A	rs138777641	intronic	ILF2	.	.	0.01176	0.0045	.	.
1	153638078	C	rs116679182	intronic	ILF2	.	.	0.005882	0.0001	.	.
1	153641045	A	rs11265624	UTR5	ILF2	.	.	0.005882	0.0034	.	.
1	153641058	C	rs114292408	intronic	ILF2	.	.	0.01176	0.0096	.	.
1	153642372	T	rs79913857	intronic	ILF2	.	.	0.1706	0.129	.	.
1	155313481	T	rs748779793	exonic	ASH1L	synonymous SNV	p.P2678P	0.005882	2.47E-05	.	.
1	155316129	A	rs60981924	intronic	ASH1L	.	.	0.005882	0.015	.	.
1	155319323	T	rs185392232	intronic	ASH1L	.	.	0.005882	0.0042	.	.
1	155327091	A	rs41264233	intronic	ASH1L	.	.	0.01765	0.0117	.	.
1	155327559	G	rs60211142	intronic	ASH1L	.	.	0.005882	0.0036	.	.
1	155340435	T	rs41264237	exonic	ASH1L	synonymous SNV	p.R2182R	0.005882	0.0048	.	.
1	155348199	G	rs139363488	intronic	ASH1L	.	.	0.01765	0.008	.	.
1	155365388	C	rs149644746	intronic	ASH1L	.	.	0.01176	0.0055	.	.
1	155408636	T	rs61732805	exonic	ASH1L	synonymous SNV	p.V1770V	0.02353	0.0096	.	.
1	155429548	T	rs113404715	intronic	ASH1L	.	.	0.005882	0.0036	.	.
1	155429725	C	rs10908466	intronic	ASH1L	.	.	0.3	0.3301	.	.
1	155448461	A	rs112530764	exonic	ASH1L	synonymous SNV	p.Y1400Y	0.005882	0.0034	.	.
1	155451719	T	rs115209829	exonic	ASH1L	synonymous SNV	p.A314A	0.005882	0.0021	.	.
1	155452285	T	rs72993486	intronic	ASH1L	.	.	0.01176	0.019	.	.
1	202700209	-	rs745309787	intronic	KDM5B	.	.	0.005882	8.81E-06	.	.
1	202701097	C	rs111793412	intronic	KDM5B	.	.	0.005882	0.0071	.	.
1	202703053	A	rs4310498	intronic	KDM5B	.	.	0.8118	0.8126	.	.
1	202705401	G	rs150737727	exonic	KDM5B	synonymous SNV	p.A1068A	0.005882	0.0023	.	.
1	202705455	C	rs1141109	exonic	KDM5B	synonymous SNV	p.P1050P	0.7529	0.7394	.	.
1	202705562	G	rs55802892	intronic	KDM5B	.	.	0.005882	0.0038	.	.
1	202710776	T	rs111464225	exonic	KDM5B	synonymous SNV	p.A888A	0.005882	0.0023	.	.
1	202711743	A	rs55798081	intronic	KDM5B	.	.	0.02941	0.0259	.	.
1	202711778	A	rs56042155	intronic	KDM5B	.	.	0.02353	0.0235	.	.
1	202715244	A	rs3216061	intronic	KDM5B	.	.	0.7529	0.7388	.	.
1	202718028	G	rs1892163	intronic	KDM5B	.	.	0.7	0.67	.	.
1	202718034	G	rs368775532	intronic	KDM5B	.	.	0.005882	6.94E-05	.	.
1	202718069	C	rs61751237	intronic	KDM5B	.	.	0.01176	0.0058	.	.
1	202718202	A	rs1892164	exonic	KDM5B	synonymous SNV	p.H629H	0.7059	0.6691	.	.
1	202718310	C	rs10920472	intronic	KDM5B	.	.	0.04706	0.0779	.	.
1	202727475	G	rs115162047	intronic	KDM5B	.	.	0.005882	0.0015	.	.
1	202729496	C	rs369961856	intronic	KDM5B	.	.	0.005882	9.39E-05	.	.
1	202729636	G	rs61750265	exonic	KDM5B	synonymous SNV	p.D328D	0.005882	0.0023	.	.
1	202729678	C	rs149504096	exonic	KDM5B	synonymous SNV	p.L314L	0.005882	0.0007	.	.
1	202731986	CTCAA	rs140328663	intronic	KDM5B	.	.	0.7529	0.7432	.	.
1	202733178	G	rs61749325	exonic	KDM5B	synonymous SNV	p.N269N	0.03529	0.0253	.	.
1	202733238	T	rs3196669	exonic	KDM5B	synonymous SNV	p.T249T	0.7529	0.7371	.	.
1	202743892	C	rs17497253	intronic	KDM5B	.	.	0.2588	0.2471	.	.
1	202743898	G	.	intronic	KDM5B	.	.	0.005882	.	.	.
1	202777215	A	rs12028388	intronic	KDM5B	.	.	0.05294	0.0805	.	.

2	162274680	C	rs568889589	intronic	TBR1	.	.	0.005882	0.0026	.	.
2	162274847	C	rs116175783	intronic	TBR1	.	.	0.07059	0.038	.	.
2	162276712	G	rs79294493	exonic	TBR1	synonymous SNV	p.T378T	0.005882	0.0066	.	.
2	162280028	CGGGCG	rs780720807	exonic	TBR1	nonframeshift insertion	p.P447delinsPGA	0.005882	0.0002	.	.
2	162280748	A	rs890076	UTR3	TBR1	.	.	0.8176	0.8231	.	.
2	166152297	C	.	UTR5	SCN2A	.	.	0.005882	.	.	.
2	166152651	A	rs111535588	intronic	SCN2A	.	.	0.005882	0.0253	.	.
2	166153499	G	rs7593568	intronic	SCN2A	.	.	0.7882	0.825	.	.
2	166164348	A	rs2304015	intronic	SCN2A	.	.	0.005882	0.0073	.	.
2	166166789	A	rs112877649	intronic	SCN2A	.	.	0.005882	0.0025	.	.
2	166168503	G	rs2304016	intronic	SCN2A	.	.	0.01176	0.0127	.	.
2	166170127	C	rs2121371	intronic	SCN2A	.	.	0.7882	0.8016	.	.
2	166172313	T	rs189735691	intronic	SCN2A	.	.	0.005882	0.0026	.	.
2	166172317	G	rs1838846	intronic	SCN2A	.	.	0.7882	0.8014	.	.
2	166179650	T	rs1867864	intronic	SCN2A	.	.	0.6059	0.5205	.	.
2	166179779	C	rs141815642	exonic	SCN2A	synonymous SNV	p.D595D	0.005882	0.0092	.	.
2	166180061	G	.	intronic	SCN2A	.	.	0.005882	.	.	.
2	166183379	G	rs147891446	exonic	SCN2A	synonymous SNV	p.T678T	0.005882	0.0069	.	.
2	166223900	G	rs142439830	intronic	SCN2A	.	.	0.01765	0.0034	.	.
2	166229695	C	rs150568699	intronic	SCN2A	.	.	0.005882	0.001	.	.
2	166234076	G	rs1864885	intronic	SCN2A	.	.	0.1588	0.2493	.	.
2	166243206	T	rs150453735	intronic	SCN2A	.	.	0.005882	0.0079	.	.
2	183791498	T	rs113570654	intronic	NCKAP1	.	.	0.01176	0.0046	.	.
2	183793653	G	rs74942055	intronic	NCKAP1	.	.	0.05882	0.0606	.	.
2	183795520	T	rs544480697	intronic	NCKAP1	.	.	0.005882	0.0003	.	.
2	183799558	C	rs9288088	exonic	NCKAP1	synonymous SNV	p.V927V	0.7588	0.7903	.	.
2	183799570	G	rs139260477	intronic	NCKAP1	.	.	0.01176	0.0122	.	.
2	183800060	G	rs144374101	exonic	NCKAP1	synonymous SNV	p.I913I	0.01176	0.0039	.	.
2	183806894	-	rs140820523	intronic	NCKAP1	.	.	0.1059	0.1039	.	.
2	183806922	T	rs145294024	intronic	NCKAP1	.	.	0.005882	8.36E-05	.	.
2	183817473	G	rs188251808	intronic	NCKAP1	.	.	0.005882	0.0037	.	.
2	183817957	A	rs66829551	exonic	NCKAP1	synonymous SNV	p.S752S	0.08824	0.0734	.	.
2	183818074	C	rs74840499	intronic	NCKAP1	.	.	0.005882	0.0066	.	.
2	183818091	T	rs72886576	intronic	NCKAP1	.	.	0.005882	0.0097	.	.
2	183822374	C	rs17265866	intronic	NCKAP1	.	.	0.05294	0.0589	.	.
2	183827018	A	rs2271671	intronic	NCKAP1	.	.	0.6353	0.7185	.	.
2	183829568	A	rs554193334	intronic	NCKAP1	.	.	0.005882	0.0002	.	.
2	183848053	G	rs35142583	exonic	NCKAP1	synonymous SNV	p.A354A	0.04706	0.0292	.	.
2	183848114	C	rs1400130	intronic	NCKAP1	.	.	0.2412	0.2276	.	.
2	183850877	T	rs41270217	intronic	NCKAP1	.	.	0.05882	0.0613	.	.
2	183867004	A	rs141187393	intronic	NCKAP1	.	.	0.1118	0.0996	.	.
2	183889791	-	rs201661388	intronic	NCKAP1	.	.	0.005882	.	.	.
2	225339134	C	.	intronic	CUL3	.	.	0.005882	.	.	.
2	225346646	C	rs2070127	exonic	CUL3	synonymous SNV	p.Q598Q	0.1471	0.1845	.	.
2	225346804	G	rs112387056	intronic	CUL3	.	.	0.005882	0.0003	.	.
2	225365056	G	rs3768889	intronic	CUL3	.	.	0.1471	0.1841	.	.
2	225365109	A	rs41373148	exonic	CUL3	synonymous SNV	p.A461A	0.01765	0.0143	.	.
2	225367669	T	rs3754629	intronic	CUL3	.	.	0.1471	0.1857	.	.
2	225368321	G	rs3738951	intronic	CUL3	.	.	0.4471	0.518	.	.
2	225376032	C	rs6743816	intronic	CUL3	.	.	0.2412	0.2188	.	.
2	225376034	A	rs11686067	intronic	CUL3	.	.	0.1471	0.1855	.	.
2	225378199	T	rs111771397	intronic	CUL3	.	.	0.005882	0.002	.	.
2	225378383	-	rs3830376	intronic	CUL3	.	.	0.2471	0.2363	.	.
2	225379325	C	rs200164153	intronic	CUL3	.	.	0.005882	0.0002	.	.
2	225400395	G	.	intronic	CUL3	.	.	0.005882	.	.	.
2	225431630	-	.	intronic	CUL3	.	.	0.005882	.	.	.
2	225431679	T	rs142512545	intronic	CUL3	.	.	0.01176	.	.	.
2	225431758	A	rs146410838	intronic	CUL3	.	.	0.005882	.	.	.
2	230632268	A	rs6687	UTR3	TRIP12	.	.	0.1118	0.1409	.	.
2	230634093	A	rs772984820	intronic	TRIP12	.	.	0.005882	0.0002	.	.
2	230643350	AAAACAAA	rs777010656	intronic	TRIP12	.	.	0.005882	2.50E-05	.	.
2	230653540	C	rs149642198	exonic	TRIP12	synonymous SNV	p.T1259T	0.005882	0.0023	.	.
2	230656014	A	rs147135813	intronic	TRIP12	.	.	0.005882	0.0068	.	.
2	230656550	A	.	intronic	TRIP12	.	.	0.005882	.	.	.
2	230663576	T	rs6720868	intronic	TRIP12	.	.	0.3471	0.3183	.	.
2	230663950	C	rs190642694	intronic	TRIP12	.	.	0.005882	0.0017	.	.
2	230668858	C	rs13018957	exonic	TRIP12	synonymous SNV	p.T567T	0.4765	0.4262	.	.
2	230668961	T	rs199509519	intronic	TRIP12	.	.	0.005882	0.0023	.	.
2	230668968	A	rs4972915	intronic	TRIP12	.	.	0.8118	0.7918	.	.
2	230668982	G	rs4973229	intronic	TRIP12	.	.	0.4706	0.4249	.	.
2	230670409	-	rs557309405	intronic	TRIP12	.	.	0.005882	0.0022	.	.
2	230670510	G	rs148271689	exonic	TRIP12	synonymous SNV	p.N517N	0.01176	0.0009	.	.
2	230672404	T	rs373864596	intronic	TRIP12	.	.	0.005882	4.12E-05	.	.
2	230695855	C	rs7595730	intronic	TRIP12	.	.	0.005882	0.0037	.	.
2	230695872	C	rs150511774	intronic	TRIP12	.	.	0.005882	0.0007	.	.
2	230723777	C	rs544480	exonic	TRIP12	synonymous SNV	p.S204S	0.1882	0.2003	.	.
2	230724301	G	rs370527992	intronic	TRIP12	.	.	0.005882	0.0004	.	.
3	9470602	G	rs189957277	UTR5	SETD5	.	.	0.005882	0.0003	.	.
3	9478642	T	rs17050336	intronic	SETD5	.	.	0.02353	.	.	.

2	1805430	A	rs2288178	intronic	MYT1L	.	.	0.06471	0.0597	.	.
2	1805489	G	rs2288179	exonic	MYT1L	synonymous SNV	p.D1085D	0.06471	0.0601	.	.
2	1842879	A	rs117052831	intronic	MYT1L	.	.	0.005882	0.0064	.	.
2	1842892	A	rs111984953	intronic	MYT1L	.	.	0.03529	0.0295	.	.
2	1842968	C	rs6728368	exonic	MYT1L	synonymous SNV	p.G1011G	0.3412	0.367	.	.
2	1842968	A	rs6728368	exonic	MYT1L	synonymous SNV	p.G1011G	0.02941	0.0553	.	.
2	1893049	A	rs71442304	exonic	MYT1L	synonymous SNV	p.P828P	0.02941	0.0344	.	.
2	1893133	T	rs75247762	exonic	MYT1L	synonymous SNV	p.Q800Q	0.06471	0.0562	.	.
2	1906828	C	rs192624064	intronic	MYT1L	.	.	0.005882	0.0018	.	.
2	1913997	C	rs180992065	intronic	MYT1L	.	.	0.005882	0.0068	.	.
2	1915749	G	rs12988500	intronic	MYT1L	.	.	0.02941	0.0626	.	.
2	1921017	A	rs375847105	exonic	MYT1L	synonymous SNV	p.S526S	0.005882	.	.	.
2	1921083	A	rs148988262	exonic	MYT1L	synonymous SNV	p.S504S	0.005882	0.0012	.	.
2	1926437	T	rs1529667	exonic	MYT1L	synonymous SNV	p.P368P	0.9706	0.9802	.	.
2	1926488	T	rs13399855	exonic	MYT1L	synonymous SNV	p.P351P	0.04118	0.0445	.	.
2	1926617	A	rs2241686	exonic	MYT1L	synonymous SNV	p.N308N	0.03529	0.0258	.	.
2	1946830	A	rs372013056	exonic	MYT1L	synonymous SNV	p.D143D	0.005882	.	.	.
2	1946857	-	rs781642397	exonic	MYT1L	nonframeshift deletion	p.134_134del	0.005882	0.0008	.	.
2	1946914	G	rs3748988	exonic	MYT1L	synonymous SNV	p.D115D	0.3176	0.4054	.	.
2	1946968	T	rs3748989	exonic	MYT1L	synonymous SNV	p.E97E	0.08235	0.1165	.	.
2	1982877	G	rs2304007	intronic	MYT1L	.	.	0.1412	.	.	.
2	1982916	-	rs3214602	intronic	MYT1L	.	.	0.1412	0.1868	.	.
2	1983241	A	rs17338581	intronic	MYT1L	.	.	0.005882	.	.	.
2	1983374	C	rs2304008	intronic	MYT1L	.	.	0.2824	0.3254	.	.
2	1983395	T	rs2304009	intronic	MYT1L	.	.	0.1235	.	.	.
2	1983606	A	.	intronic	MYT1L	.	.	0.005882	.	.	.
2	25458546	T	rs2304429	intronic	DNMT3A	.	.	0.5706	0.5098	.	.
2	25462327	G	rs72810046	intronic	DNMT3A	.	.	0.09412	0.0916	.	.
2	25463483	A	rs2289195	intronic	DNMT3A	.	.	0.3647	0.4061	.	.
2	25466888	T	rs2289093	intronic	DNMT3A	.	.	0.7235	0.7275	.	.
2	25469502	T	rs2276598	exonic	DNMT3A	synonymous SNV	p.L233L	0.1471	0.1913	.	.
2	25469628	T	rs77345627	exonic	DNMT3A	synonymous SNV	p.A191A	0.005882	0.0052	.	.
2	25469886	A	.	intronic	DNMT3A	.	.	0.005882	.	.	.
2	25469913	T	rs2276599	intronic	DNMT3A	.	.	0.7294	0.7167	.	.
2	25471002	A	rs77558739	exonic	DNMT3A	synonymous SNV	p.P64P	0.005882	0.0022	.	.
2	25474992	C	rs556043659	intronic	DNMT3A	.	.	0.005882	.	.	.
2	25536772	T	rs370534287	intronic	DNMT3A	.	.	0.005882	0.0015	.	.
2	25536827	A	rs41284843	exonic	DNMT3A	synonymous SNV	p.P9P	0.1059	0.155	.	.
2	32312700	C	rs7561519	intronic	SPAST	.	.	0.4294	.	.	.
2	32340724	A	rs76399353	intronic	SPAST	.	.	0.05294	0.0195	.	.
2	32340779	A	rs145264166	exonic	SPAST	synonymous SNV	p.P261P	0.01176	0.0089	.	.
2	32341334	-	rs758517424	intronic	SPAST	.	.	0.005882	0.0001	.	.
2	32351982	G	rs755277514	intronic	SPAST	.	.	0.005882	3.62E-05	.	.
2	32361710	G	.	intronic	SPAST	.	.	0.005882	.	.	.
2	32362079	TATA	rs10627985	intronic	SPAST	.	.	0.3529	0.394	.	.
2	32368361	G	rs760953889	intronic	SPAST	.	.	0.005882	6.10E-05	.	.
2	32370105	C	rs112410719	intronic	SPAST	.	.	0.005882	0.0025	.	.
2	50149352	C	rs55923848	exonic	NRXN1	synonymous SNV	p.P353P	0.005882	0.001	.	.
2	50201110	G	rs9636391	intronic	NRXN1	.	.	0.8647	0.8501	.	17.05
2	50201255	A	rs138066456	intronic	NRXN1	.	.	0.01765	0.0016	.	.
2	50201382	T	rs200527832	intronic	NRXN1	.	.	0.005882	0.0005	.	.
2	50318648	C	rs756210249	intronic	NRXN1	.	.	0.005882	8.26E-06	.	.
2	50318661	C	rs74387895	intronic	NRXN1	.	.	0.005882	0.0142	.	.
2	50434741	T	rs17040210	intronic	NRXN1	.	.	0.2765	.	.	.
2	50434759	G	rs1715984	intronic	NRXN1	.	.	0.8471	.	.	.
2	50464065	T	rs80094872	exonic	NRXN1	synonymous SNV	p.T101T	0.005882	0.0025	.	.
2	50692560	G	rs3213756	intronic	NRXN1	.	.	0.3176	0.3834	.	.
2	50699479	A	rs75275592	exonic	NRXN1	synonymous SNV	p.S1067S	0.005882	0.0006	.	.
2	50699638	G	rs116737278	intronic	NRXN1	.	.	0.005882	0.0072	.	.
2	50758356	C	.	intronic	NRXN1	.	.	0.005882	.	.	.
2	50758612	T	rs201210484	intronic	NRXN1	.	.	0.005882	2.13E-05	.	.
2	50780119	G	rs201727684	exonic	NRXN1	synonymous SNV	p.L455L	0.005882	0.0005	.	.
2	50847139	A	rs199753235	intronic	NRXN1	.	.	0.005882	0.0001	.	18.8
2	51149072	C	rs7423296	intronic	NRXN1	.	.	0.09412	.	.	.
2	51149089	G	rs370178631	intronic	NRXN1	.	.	0.005882	.	.	.
2	51149102	C	rs4327263	intronic	NRXN1	.	.	0.09412	.	.	.
2	51253477	C	rs139064548	intronic	NRXN1	.	.	0.01176	0.0058	.	.
2	51254901	A	rs1045874	exonic	NRXN1	synonymous SNV	p.L171L	0.09412	0.1306	.	.
2	60679832	-	rs555661121	intronic	BCL11A	.	.	0.005882	0.0099	.	.
2	60687959	G	rs7569946	exonic	BCL11A	synonymous SNV	p.S696S	0.6059	0.6977	.	.
2	60688331	A	rs114252508	exonic	BCL11A	synonymous SNV	p.R572R	0.005882	0.001	.	.
2	60689441	C	rs61749494	exonic	BCL11A	synonymous SNV	p.E202E	0.01765	0.0217	.	.
2	60689522	A	rs61748090	exonic	BCL11A	synonymous SNV	p.C175C	0.02353	0.018	.	.
2	60768978	A	rs2665668	intronic	BCL11A	.	.	0.5882	0.6417	.	.
2	149225888	A	rs192677180	intronic	MBD5	.	.	0.005882	0.0014	.	.
2	149241448	A	rs568615665	intronic	MBD5	.	.	0.005882	0.0004	.	.
2	149243260	A	rs2121344	intronic	MBD5	.	.	0.6647	0.6501	.	.
2	149246906	T	rs111476224	intronic	MBD5	.	.	0.05882	0.0907	.	.
2	149259949	A	rs73003524	intronic	MBD5	.	.	0.06471	.	.	.

3	9478653	A	.	intronic	SETD5	.	.	0.005882	.	.	.
3	9483315	A	rs17050347	exonic	SETD5	synonymous SNV	p.R283R	0.01765	0.016	.	.
3	9487415	C	rs117265865	exonic	SETD5	synonymous SNV	p.H505H	0.005882	0.0045	.	.
3	9488715	A	rs74787011	intronic	SETD5	.	.	0.04118	0.0698	.	.
3	9489334	TTTC	rs201539707	intronic	SETD5	.	.	0.05294	0.0678	.	.
3	9489340	A	rs200418836	intronic	SETD5	.	.	0.05294	0.0681	.	.
3	9490352	T	rs187073560	intronic	SETD5	.	.	0.005882	0.0023	.	.
3	9495385	G	rs2648580	intronic	SETD5	.	.	0.7294	0.7048	.	.
3	9515054	C	rs73026975	exonic	SETD5	synonymous SNV	p.T1110T	0.02941	0.0233	.	.
3	9517337	A	rs199989217	exonic	SETD5	synonymous SNV	p.T1297T	0.005882	0.0008	.	.
3	9517577	T	rs17747739	exonic	SETD5	synonymous SNV	p.N1377N	0.06471	0.0385	.	.
3	11059035	A	rs183069336	exonic	SLC6A1	synonymous SNV	p.T46T	0.005882	0.0064	.	.
3	11060430	C	rs3817585	ncRNA_intronic	SLC6A1-AS1	.	.	0.1471	0.0905	.	.
3	11061907	T	rs116620331	exonic	SLC6A1	synonymous SNV	p.P160P	0.005882	0.0002	.	.
3	11061919	T	rs151263329	exonic	SLC6A1	synonymous SNV	p.C164C	0.005882	0.0001	.	.
3	11064091	T	rs6344	exonic	SLC6A1	synonymous SNV	p.T217T	0.03529	0.0594	.	.
3	11067583	T	rs35736058	intronic	SLC6A1	.	.	0.005882	0.0093	.	.
3	11068038	A	rs113953	exonic	SLC6A1	synonymous SNV	p.A357A	0.005882	4.17E-05	.	.
3	11070543	-	rs11290120	intronic	SLC6A1	.	.	0.1059	0.1279	.	.
3	11070864	T	rs6770472	intronic	SLC6A1	.	.	0.005882	0.0075	.	.
3	11070870	A	rs36034065	intronic	SLC6A1	.	.	0.01176	0.0124	.	.
3	11070883	T	rs150889270	intronic	SLC6A1	.	.	0.005882	0.0053	.	.
3	11076171	C	rs150938735	intronic	SLC6A1	.	.	0.01176	0.0037	.	.
3	11076336	T	rs2272401	exonic	SLC6A1	synonymous SNV	p.P549P	0.005882	0.0033	.	.
3	11076400	G	rs35957531	intronic	SLC6A1	.	.	0.1588	0.1098	.	.
3	20113779	T	rs41285055	intronic	KAT2B	.	.	0.005882	0.0046	.	.
3	20113830	A	rs3021408	exonic	KAT2B	synonymous SNV	p.E103E	0.4529	0.4074	.	.
3	20113989	A	rs3749180	intronic	KAT2B	.	.	0.4529	0.4075	.	.
3	20141301	C	rs11128939	intronic	KAT2B	.	.	0.1706	.	.	.
3	20141301	G	rs11128939	intronic	KAT2B	.	.	0.4353	.	.	.
3	20141308	G	rs62243131	intronic	KAT2B	.	.	0.1706	0.1869	.	.
3	20141356	T	rs35424474	exonic	KAT2B	synonymous SNV	p.L193L	0.1588	0.1271	.	.
3	20142983	G	rs199900176	intronic	KAT2B	.	.	0.005882	8.31E-06	.	.
3	20161043	T	rs17006623	intronic	KAT2B	.	.	0.1706	0.1655	.	.
3	20168875	G	rs41285061	intronic	KAT2B	.	.	0.1647	0.1258	.	.
3	20181676	C	rs3762633	intronic	KAT2B	.	.	0.1824	0.2404	.	.
3	20181904	T	rs115014362	intronic	KAT2B	.	.	0.03529	0.0169	.	.
3	20189565	T	rs116785376	intronic	KAT2B	.	.	0.01176	.	.	.
3	20193948	T	rs150884035	exonic	KAT2B	synonymous SNV	p.Y810Y	0.005882	0.0002	.	.
3	71008557	A	rs7639736	intronic	FOXP1	.	.	0.01176	0.0284	.	.
3	71015021	T	rs7638391	intronic	FOXP1	.	.	0.9	0.9452	.	.
3	71026047	C	rs72960080	intronic	FOXP1	.	.	0.005882	0.009	.	.
3	71026809	C	rs144080925	exonic	FOXP1	synonymous SNV	p.A371A	0.01176	0.0023	.	.
3	71027199	T	rs75214049	intronic	FOXP1	.	.	0.005882	0.0105	.	.
3	71161838	A	rs17008224	intronic	FOXP1	.	.	0.01765	0.0298	.	.
3	71247257	G	rs939845	intronic	FOXP1	.	.	0.03529	0.0891	.	.
3	71247304	G	rs2037474	intronic	FOXP1	.	.	0.1294	0.2009	.	.
4	85612737	G	rs3747680	intronic	WDFY3	.	.	0.2941	0.315	.	.
4	85618002	-	rs143604772	intronic	WDFY3	.	.	0.04706	0.0364	.	.
4	85625548	T	rs199518075	exonic	WDFY3	synonymous SNV	p.I2795I	0.005882	0.0006	.	.
4	85634264	A	rs61624886	intronic	WDFY3	.	.	0.005882	0.0191	.	.
4	85636427	-	rs764374772	intronic	WDFY3	.	.	0.005882	0.0005	.	.
4	85638024	T	rs17368157	intronic	WDFY3	.	.	0.04706	0.038	.	.
4	85642558	G	rs187897046	intronic	WDFY3	.	.	0.005882	0.002	.	.
4	85645557	T	rs200971792	intronic	WDFY3	.	.	0.005882	0.0003	.	.
4	85657280	T	rs77543638	intronic	WDFY3	.	.	0.08824	0.1178	.	.
4	85658525	-	rs150542176	intronic	WDFY3	.	.	0.005882	0.0035	.	.
4	85661615	G	rs17009235	intronic	WDFY3	.	.	0.005882	0.011	.	.
4	85663111	G	rs114507812	intronic	WDFY3	.	.	0.02353	0.0041	.	.
4	85672894	T	rs375868219	intronic	WDFY3	.	.	0.005882	8.24E-05	.	.
4	85687368	A	rs192153973	intronic	WDFY3	.	.	0.005882	0.0036	.	.
4	85695956	-	rs754521914	intronic	WDFY3	.	.	0.005882	7.88E-05	.	.
4	85704254	C	rs201177530	intronic	WDFY3	.	.	0.005882	4.16E-05	.	.
4	85708792	A	rs34548715	exonic	WDFY3	synonymous SNV	p.Y1248Y	0.04706	0.0348	.	.
4	85730994	C	rs781545971	intronic	WDFY3	.	.	0.005882	7.00E-05	.	.
4	85762385	C	rs2046402	exonic	WDFY3	synonymous SNV	p.L112L	0.3706	0.3985	.	.
4	113825611	C	rs775400814	UTR5	ANK2	.	.	0.01176	0.0035	.	.
4	114028104	T	rs148339577	ncRNA_exonic	MIR1243	.	.	0.005882	0.0007	.	.
4	114028149	A	rs17625836	intronic	ANK2	.	.	0.1824	0.1923	.	.
4	114120284	T	rs45502093	intronic	ANK2	.	.	0.03529	0.0303	.	.
4	114161619	C	rs45616931	intronic	ANK2	.	.	0.01765	0.0072	.	.
4	114163399	G	rs362504	intronic	ANK2	.	.	0.005882	0.0066	.	.
4	114179262	G	rs34145832	exonic	ANK2	synonymous SNV	p.E415E	0.005882	0.0017	.	.
4	114186005	T	rs29413	intronic	ANK2	.	.	0.02353	0.0217	.	.
4	114199722	G	rs74976371	intronic	ANK2	.	.	0.005882	0.0017	.	.
4	114203944	G	rs45442693	exonic	ANK2	synonymous SNV	p.P665P	0.005882	0.0017	.	.
4	114209667	G	rs767977681	intronic	ANK2	.	.	0.005882	8.46E-06	.	.
4	114209691	T	rs29357	intronic	ANK2	.	.	0.06471	0.1062	.	.
4	114213551	T	rs55706912	intronic	ANK2	.	.	0.02353	0.0177	.	.

5	65346599	A	rs56216143	intronic	ERBB2IP	.	.	0.02941	0.0199	.	.
5	65350002	T	rs34528338	exonic	ERBB2IP	synonymous SNV	p.P952P	0.005882	0.002	.	.
5	65350044	G	rs35278406	exonic	ERBB2IP	synonymous SNV	p.Q966Q	0.02941	0.028	.	.
5	65350173	T	rs77719384	exonic	ERBB2IP	synonymous SNV	p.L1009L	0.005882	0.002	.	.
5	65350374	G	rs36303	exonic	ERBB2IP	synonymous SNV	p.R1076R	0.1882	0.1915	.	.
5	65367974	G	rs76755657	intronic	ERBB2IP	.	.	0.005882	0.0052	.	.
5	65372805	T	rs73763088	intronic	ERBB2IP	.	.	0.01176	0.0167	.	.
5	170305139	G	rs747930513	exonic	RANBP17	frameshift insertion	p.Y19fs	0.005882	8.36E-06	.	.
5	170319575	T	rs79372097	intronic	RANBP17	.	.	0.005882	0.0085	.	.
5	170338163	A	rs62621882	intronic	RANBP17	.	.	0.02941	0.0381	.	.
5	170345708	G	rs112213519	intronic	RANBP17	.	.	0.005882	0.0086	.	.
5	170597190	A	rs36104512	exonic	RANBP17	synonymous SNV	p.E589E	0.04118	0.066	.	.
5	170610111	A	rs546355758	intronic	RANBP17	.	.	0.005882	.	.	.
5	170610273	A	rs6555936	intronic	RANBP17	.	.	0.7412	0.7345	.	.
5	170610349	T	rs774753172	exonic	RANBP17	synonymous SNV	p.G651G	0.005882	8.24E-06	.	.
5	170610452	C	rs150075974	intronic	RANBP17	.	.	0.005882	0.0008	.	.
5	170626735	G	rs35724654	exonic	RANBP17	synonymous SNV	p.V700V	0.1176	0.1173	.	.
5	170720853	A	rs189929475	intronic	RANBP17	.	.	0.005882	0.0018	.	.
6	33391270	C	rs142359891	exonic	SYNGAP1	synonymous SNV	p.S28S	0.005882	0.0061	.	.
6	33400061	A	rs114505996	intronic	SYNGAP1	.	.	0.01176	0.0112	.	.
6	33403422	T	rs453590	intronic	SYNGAP1	.	.	0.4471	0.3955	.	.
6	33406300	C	rs72887798	exonic	SYNGAP1	synonymous SNV	p.Y497Y	0.01765	0.0048	.	.
6	33406556	G	rs7759963	exonic	SYNGAP1	synonymous SNV	p.E512E	0.02353	0.0379	.	.
6	33408542	A	rs411136	exonic	SYNGAP1	synonymous SNV	p.S571S	0.4353	0.4861	.	.
6	33408612	T	rs761973471	exonic	SYNGAP1	synonymous SNV	p.L595L	0.005882	8.27E-06	.	.
6	33408764	-	rs572578854	intronic	SYNGAP1	.	.	0.1529	0.0969	.	.
6	33409329	C	rs778673739	intronic	SYNGAP1	.	.	0.005882	0.0008	.	.
6	33411299	T	rs61421477	exonic	SYNGAP1	synonymous SNV	p.S990S	0.01176	0.0268	.	.
6	33411602	G	rs145707539	exonic	SYNGAP1	synonymous SNV	p.L1091L	0.005882	0.0005	.	.
6	33411653	T	rs139841529	exonic	SYNGAP1	synonymous SNV	p.S1108S	0.005882	0.0016	.	.
6	33412197	T	rs9969005	intronic	SYNGAP1	.	.	0.01176	0.0091	.	.
6	43097443	C	rs45624340	intronic	PTK7	.	.	0.05294	0.0463	.	.
6	43098182	A	rs77231564	intronic	PTK7	.	.	0.03529	0.0482	.	.
6	43100156	T	rs141307288	intronic	PTK7	.	.	0.01176	0.0046	.	.
6	43106824	A	rs45484692	intronic	PTK7	.	.	0.05294	0.032	.	.
6	43109564	A	rs199816335	intronic	PTK7	.	.	0.005882	0.0006	.	.
6	43109751	A	rs6905948	exonic	PTK7	synonymous SNV	p.G487G	0.3706	0.3606	.	.
6	43110003	T	rs45453593	exonic	PTK7	synonymous SNV	p.N541N	0.005882	0.0079	.	.
6	43110055	G	rs6933124	intronic	PTK7	.	.	0.1	0.1835	.	.
6	43113190	G	rs200495132	intronic	PTK7	.	.	0.005882	4.69E-05	.	.
6	43127577	G	rs55921533	exonic	PTK7	synonymous SNV	p.P845P	0.005882	0.0053	.	.
6	43128434	G	rs61021888	intronic	PTK7	.	.	0.1294	0.187	.	.
6	99883694	C	rs9402791	exonic	USP45	synonymous SNV	p.A781A	0.2059	0.252	.	.
6	99885246	T	rs118066385	exonic	USP45	stopgain	p.Y730X	0.005882	0.0016	.	36
6	99891391	C	rs34401990	intronic	USP45	.	.	0.1824	0.1331	.	.
6	99891396	-	rs140108249	intronic	USP45	.	.	0.1824	0.1343	.	.
6	99893830	A	rs111592029	exonic	USP45	synonymous SNV	p.T606T	0.005882	0.0058	.	.
6	99893878	C	rs12203426	exonic	USP45	synonymous SNV	p.L590L	0.1941	0.1516	.	.
6	99912454	T	rs748532477	intronic	USP45	.	.	0.005882	2.51E-05	.	.
6	99912558	G	rs143913725	exonic	USP45	synonymous SNV	p.D410D	0.005882	0.0009	.	.
6	99930593	T	rs117666915	intronic	USP45	.	.	0.02353	0.0164	.	.
6	99951750	A	rs6934692	intronic	USP45	.	.	0.7941	0.8489	.	.
6	99951768	C	rs11751123	intronic	USP45	.	.	0.005882	0.0023	.	.
6	157099872	-	rs766956053	exonic	ARID1B	nonframeshift deletion	p.270_271del	0.01176	0.0017	.	.
6	157405761	A	rs3734440	intronic	ARID1B	.	.	0.5118	0.5227	.	.
6	157405930	A	rs3734441	exonic	ARID1B	synonymous SNV	p.A711A	0.5118	0.519	.	.
6	157431564	G	rs75599866	intronic	ARID1B	.	.	0.005882	0.0074	.	.
6	157470134	T	rs766787725	intronic	ARID1B	.	.	0.005882	0.0003	.	.
6	157496163	T	rs562037567	intronic	ARID1B	.	.	0.005882	0.0015	.	.
6	157502279	T	rs61736269	exonic	ARID1B	synonymous SNV	p.Y1091Y	0.005882	0.0006	.	.
6	157505418	T	rs142391292	exonic	ARID1B	synonymous SNV	p.T1120T	0.005882	0.0029	.	.
6	157507598	A	rs566480933	intronic	ARID1B	.	.	0.005882	0.0007	.	.
6	157507670	A	rs148720121	intronic	ARID1B	.	.	0.005882	0.0027	.	.
6	157519938	G	rs112318565	intronic	ARID1B	.	.	0.06471	0.0859	.	.
6	157520060	A	rs142594004	intronic	ARID1B	.	.	0.005882	0.0066	.	.
6	157522252	C	rs61747988	exonic	ARID1B	synonymous SNV	p.H1495H	0.005882	0.0053	.	.
6	157522360	A	rs61738955	exonic	ARID1B	synonymous SNV	p.P1531P	0.01765	0.0269	.	.
6	157527495	T	rs759473238	exonic	ARID1B	synonymous SNV	p.D1727D	0.005882	2.49E-05	.	.
6	157527498	T	rs775201232	exonic	ARID1B	synonymous SNV	p.D1728D	0.005882	1.65E-05	.	.
6	157528077	T	rs142499766	exonic	ARID1B	synonymous SNV	p.I1921I	0.005882	0.0036	.	.
6	157528197	T	rs112703040	exonic	ARID1B	synonymous SNV	p.I1961I	0.01176	0.0101	.	.
6	157528251	A	rs151317970	exonic	ARID1B	synonymous SNV	p.P1979P	0.005882	3.30E-05	.	.
6	157528356	A	rs753595827	exonic	ARID1B	synonymous SNV	p.T2014T	0.005882	4.12E-05	.	.
7	91623897	A	rs558415055	intronic	AKAP9	.	.	0.005882	0.0016	.	.
7	91624931	G	rs78515732	exonic	AKAP9	synonymous SNV	p.E249E	0.005882	0.0031	.	.
7	91632306	T	rs1989779	exonic	AKAP9	synonymous SNV	p.T1025T	0.8941	0.8924	.	.
7	91641928	G	rs13245393	exonic	AKAP9	synonymous SNV	p.E1168E	0.3765	0.3838	.	.
7	91645454	C	rs142610139	exonic	AKAP9	synonymous SNV	p.S1208S	0.005882	0.0004	.	.
7	91652178	AAC	rs10644111	exonic	AKAP9	nonframeshift insertion	p.K1335delinsKQ	0.3765	0.3989	.	.

4	114213631	T	rs29341	exonic	ANK2	synonymous SNV	p.V779V	0.04118	0.0363	.	.
4	114213705	T	rs72556366	intronic	ANK2	.	.	0.005882	0.0104	.	.
4	114213713	C	rs762049614	intronic	ANK2	.	.	0.005882	8.47E-05	.	.
4	114214588	T	rs139893914	intronic	ANK2	.	.	0.01176	0.0062	.	.
4	114244834	C	.	intronic	ANK2	.	.	0.005882	.	.	.
4	114257132	G	rs35336373	exonic	ANK2	synonymous SNV	p.V1170V	0.005882	0.0097	.	.
4	114257201	T	rs3736575	exonic	ANK2	synonymous SNV	p.R1193R	0.09412	0.146	.	.
4	114257706	T	rs59906453	intronic	ANK2	.	.	0.005882	0.0162	.	.
4	114257955	A	rs66792339	intronic	ANK2	.	.	0.005882	0.0072	.	.
4	114260492	T	rs2272231	intronic	ANK2	.	.	0.08824	0.1377	.	.
4	114262877	G	rs567570285	exonic	ANK2	synonymous SNV	p.E1309E	0.005882	0.0008	.	.
4	114267035	C	rs72556368	intronic	ANK2	.	.	0.005882	0.02	.	.
4	114269501	C	rs189883943	intronic	ANK2	.	.	0.01176	0.0034	.	.
4	114269509	A	rs2272234	intronic	ANK2	.	.	0.01765	0.0403	.	.
4	114275243	T	rs33966911	exonic	ANK2	synonymous SNV	p.P1823P	0.1118	0.0796	.	.
4	114275942	G	rs3796928	exonic	ANK2	synonymous SNV	p.L2056L	0.02353	0.0492	.	.
4	114276422	G	rs140926982	exonic	ANK2	synonymous SNV	p.G2216G	0.005882	0.0023	.	.
4	114276884	G	rs3733615	exonic	ANK2	synonymous SNV	p.Q2370Q	0.1647	0.1902	.	.
4	114279422	G	rs10013743	exonic	ANK2	synonymous SNV	p.E3216E	0.1176	0.0982	.	.
4	114284645	T	rs35728190	intronic	ANK2	.	.	0.1118	0.092	.	.
4	114286289	C	rs544195596	exonic	ANK2	synonymous SNV	p.S1576S	0.005882	5.77E-05	.	.
4	114294308	C	rs2293324	exonic	ANK2	synonymous SNV	p.H1806H	0.1706	0.1974	.	.
4	114302634	T	rs35446871	UTR3	ANK2	.	.	0.03529	0.0289	.	.
4	140281609	A	rs148484204	intronic	NAA15	.	.	0.005882	0.0013	.	.
4	140283103	-	rs3217605	intronic	NAA15	.	.	0.1647	0.2191	.	.
4	140299862	T	rs76253596	intronic	NAA15	.	.	0.005882	0.0018	.	.
5	14270947	T	rs55664610	exonic	TRIO	synonymous SNV	p.N57N	0.01765	0.0079	.	.
5	14280385	A	rs184686375	intronic	TRIO	.	.	0.005882	0.0012	.	.
5	14290805	A	rs30629	intronic	TRIO	.	.	0.4765	0.5072	.	.
5	14291259	A	rs61737132	exonic	TRIO	synonymous SNV	p.L325L	0.005882	0.0119	.	.
5	14293254	T	rs373817005	intronic	TRIO	.	.	0.005882	0.0001	.	.
5	14316642	A	rs55920001	exonic	TRIO	synonymous SNV	p.S507S	0.06471	0.0418	.	.
5	14330869	G	rs16903402	intronic	TRIO	.	.	0.005882	0.0017	.	.
5	14330940	C	rs2277045	exonic	TRIO	synonymous SNV	p.G595G	0.06471	0.0429	.	.
5	14359665	A	rs200438805	intronic	TRIO	.	.	0.01176	0.005	.	.
5	14359666	T	rs199633885	intronic	TRIO	.	.	0.01176	0.005	.	.
5	14363830	T	rs55751460	intronic	TRIO	.	.	0.04706	0.0482	.	.
5	14364726	T	rs114634082	intronic	TRIO	.	.	0.02941	0.0157	.	.
5	14368975	T	rs13189406	exonic	TRIO	synonymous SNV	p.N1011N	0.2647	0.2508	.	.
5	14369548	A	rs764660542	exonic	TRIO	synonymous SNV	p.A1044A	0.005882	8.27E-06	.	.
5	14374397	T	rs145133665	exonic	TRIO	synonymous SNV	p.S1092S	0.005882	0.0014	.	.
5	14378088	T	rs10866507	intronic	TRIO	.	.	0.08824	0.0921	.	.
5	14378269	G	rs147151301	intronic	TRIO	.	.	0.02941	0.0164	.	.
5	14387543	A	rs16903450	intronic	TRIO	.	.	0.005882	0.0053	.	.
5	14387615	C	rs140850570	exonic	TRIO	synonymous SNV	p.H1213H	0.01176	0.001	.	.
5	14387772	G	rs142970888	intronic	TRIO	.	.	0.01765	0.0058	.	.
5	14387813	C	rs2289849	intronic	TRIO	.	.	0.01176	0.0156	.	.
5	14388686	G	rs11949756	intronic	TRIO	.	.	0.08824	0.0671	.	.
5	14389469	T	rs7715916	exonic	TRIO	synonymous SNV	p.I1340I	0.08824	0.0696	.	.
5	14394283	A	rs256412	intronic	TRIO	.	.	0.2647	0.2805	.	.
5	14405920	T	rs141492551	intronic	TRIO	.	.	0.005882	0.0008	.	.
5	14420027	C	rs30612	exonic	TRIO	synonymous SNV	p.T1700T	0.8353	0.8298	.	.
5	14420150	A	rs56207939	intronic	TRIO	.	.	0.1	0.1006	.	.
5	14461168	G	rs55812347	exonic	TRIO	synonymous SNV	p.A1748A	0.02941	0.0451	.	.
5	14461454	A	rs62345056	intronic	TRIO	.	.	0.01176	0.0059	.	.
5	14462816	T	rs27100	intronic	TRIO	.	.	0.4294	0.419	.	.
5	14462890	T	rs17500731	exonic	TRIO	synonymous SNV	p.S1841S	0.05294	0.0217	.	.
5	14465628	T	rs79312837	intronic	TRIO	.	.	0.01765	0.0121	.	.
5	14465787	A	rs111451616	intronic	TRIO	.	.	0.005882	0.0027	.	.
5	14476968	T	rs40490	intronic	TRIO	.	.	0.1	0.1057	.	.
5	14479323	A	rs62345860	intronic	TRIO	.	.	0.04118	0.0168	.	.
5	14479333	C	rs60286979	intronic	TRIO	.	.	0.005882	0.0122	.	.
5	14479991	T	rs32527	intronic	TRIO	.	.	0.005882	0.013	.	.
5	14481685	A	.	exonic	TRIO	synonymous SNV	p.R2141R	0.005882	.	.	.
5	14485386	C	rs26098	intronic	TRIO	.	.	0.1235	0.1391	.	.
5	14487530	A	rs116296447	intronic	TRIO	.	.	0.005882	0.014	.	.
5	14487568	A	rs758053482	intronic	TRIO	.	.	0.005882	0.0002	.	.
5	14488194	-	rs140308852	exonic	TRIO	nonframeshift deletion	p.2486_2490del	0.005882	0.0025	.	.
5	14498000	C	rs80292429	intronic	TRIO	.	.	0.005882	0.0028	.	.
5	14498624	A	rs115607794	intronic	TRIO	.	.	0.01765	0.0228	.	.
5	14498727	A	rs26107	exonic	TRIO	synonymous SNV	p.S2770S	0.1176	0.1334	.	.
5	65290543	T	rs61596028	intronic	ERBB2IP	.	.	0.005882	0.0175	.	.
5	65290544	G	rs56195564	intronic	ERBB2IP	.	.	0.005882	0.0074	.	.
5	65290713	G	rs75060345	intronic	ERBB2IP	.	.	0.005882	0.007	.	.
5	65317181	T	rs706679	exonic	ERBB2IP	synonymous SNV	p.L189L	0.7941	0.7708	.	.
5	65319244	C	rs753067662	intronic	ERBB2IP	.	.	0.005882	1.66E-05	.	.
5	65320118	T	rs533894721	intronic	ERBB2IP	.	.	0.01765	0.0073	.	.
5	65321784	T	rs146856556	exonic	ERBB2IP	synonymous SNV	p.Y333Y	0.005882	0.0007	.	.
5	65344623	A	rs56113810	intronic	ERBB2IP	.	.	0.005882	0.0057	.	.

7	91667692	G	rs2285333	intronic	AKAP9	.	.	0.3706	0.3914	.	.
7	91669949	-	rs34590567	intronic	AKAP9	.	.	0.005882	0.004	.	.
7	91669978	-	.	intronic	AKAP9	.	.	0.005882	.	.	.
7	91674302	A	rs9785013	intronic	AKAP9	.	.	0.3706	0.3993	.	.
7	91682296	C	rs201843283	intronic	AKAP9	.	.	0.005882	0.0038	.	.
7	91691601	T	rs10236397	exonic	AKAP9	synonymous SNV	p.G1926G	0.3706	0.3795	.	.
7	91691759	-	rs757365188	exonic	AKAP9	frameshift deletion	p.Q1979fs	0.005882	1.66E-05	.	.
7	91699314	T	rs146462493	intronic	AKAP9	.	.	0.005882	0.0011	.	.
7	91700339	C	rs748480385	intronic	AKAP9	.	.	0.005882	1.66E-05	.	.
7	91700353	A	rs749413872	intronic	AKAP9	.	.	0.005882	2.51E-05	.	.
7	91707197	T	rs733957	intronic	AKAP9	.	.	0.3706	0.4076	.	.
7	91708722	A	rs61757672	exonic	AKAP9	synonymous SNV	p.Q2425Q	0.01176	0.0028	.	.
7	91711791	G	rs2079082	intronic	AKAP9	.	.	0.3706	0.3834	.	.
7	91713972	T	rs10228334	exonic	AKAP9	synonymous SNV	p.L2889L	0.3765	0.3832	.	.
7	91715662	T	rs28927678	exonic	AKAP9	synonymous SNV	p.L3049L	0.3647	0.3589	.	.
7	91718874	G	rs56295910	intronic	AKAP9	.	.	0.02353	0.0239	.	7.624
7	91719010	A	rs149829152	intronic	AKAP9	.	.	0.01176	0.0013	.	.
7	91719035	T	rs181209481	intronic	AKAP9	.	.	0.005882	0.003	.	.
7	91726927	C	rs1063243	exonic	AKAP9	synonymous SNV	p.R3476R	0.3765	0.3846	.	.
7	91727385	A	rs6946356	intronic	AKAP9	.	.	0.3588	0.3947	.	.
7	91735107	A	rs74753191	intronic	AKAP9	.	.	0.005882	0.0044	.	.
7	100279437	T	rs76501659	intronic	GIGYF1	.	.	0.01176	0.0045	.	.
7	100279633	A	rs779157646	intronic	GIGYF1	.	.	0.005882	4.23E-05	.	.
7	100279705	GCC	rs769704427	exonic	GIGYF1	nonframeshift insertion	p.Q972delinsRQ	0.005882	8.54E-06	.	.
7	100279928	A	.	intronic	GIGYF1	.	.	0.005882	.	.	.
7	100279946	A	rs117080933	exonic	GIGYF1	synonymous SNV	p.D920D	0.02353	0.0284	.	.
7	100280039	T	.	exonic	GIGYF1	synonymous SNV	p.L889L	0.005882	.	.	.
7	100280896	C	rs221793	intronic	GIGYF1	.	.	0.9471	0.9311	.	.
7	100281264	A	rs199778365	intronic	GIGYF1	.	.	0.005882	0.0006	.	.
7	100282538	A	rs11974395	intronic	GIGYF1	.	.	0.1941	0.2219	.	.
7	100282707	A	rs117339142	exonic	GIGYF1	synonymous SNV	p.L419L	0.01765	0.0359	.	.
7	100283914	C	rs143725633	exonic	GIGYF1	synonymous SNV	p.R279R	0.01765	0.0101	.	.
7	100284246	G	rs11975502	intronic	GIGYF1	.	.	0.01765	0.013	.	.
7	100284606	T	rs181902629	intronic	GIGYF1	.	.	0.01765	0.0057	.	.
7	100284913	G	rs221796	intronic	GIGYF1	.	.	0.9118	0.895	.	.
7	100285369	A	.	intronic	GIGYF1	.	.	0.005882	.	.	.
7	100285757	C	rs75939759	intronic	GIGYF1	.	.	0.05294	0.0465	.	.
7	100285801	T	rs371623925	intronic	GIGYF1	.	.	0.005882	0.0003	.	.
7	100285812	A	rs116789774	intronic	GIGYF1	.	.	0.005882	0.0054	.	.
7	100285888	T	rs2272572	UTR5	GIGYF1	.	.	0.1941	0.2158	.	.
7	100285906	A	rs192084647	UTR5	GIGYF1	.	.	0.005882	0.02	.	.
7	100285941	C	rs117459930	UTR5	GIGYF1	.	.	0.05294	.	.	.
7	100490077	A	rs7636	exonic	ACHE	synonymous SNV	p.P477P	0.06471	0.0642	.	.
7	100490765	T	rs17234989	intronic	ACHE	.	.	0.005882	0.002	.	.
7	100491047	T	rs17228581	exonic	ACHE	synonymous SNV	p.T269T	0.005882	0.0016	.	.
7	104703947	G	rs769210509	exonic	KMT2E	synonymous SNV	p.G112G	0.005882	8.25E-06	.	.
7	104714062	A	rs145147042	intronic	KMT2E	.	.	0.005882	0.0074	.	.
7	104716591	A	rs371638874	intronic	KMT2E	.	.	0.005882	.	.	.
7	104717517	T	rs2240455	exonic	KMT2E	synonymous SNV	p.Y292Y	0.1529	0.2282	.	.
7	104722143	C	rs151074690	exonic	KMT2E	synonymous SNV	p.H419H	0.005882	0.0007	.	.
7	104742054	T	rs11976329	intronic	KMT2E	.	.	0.1471	0.2332	.	.
7	104742679	C	rs77383660	intronic	KMT2E	.	.	0.01765	0.0173	.	.
7	104745867	G	rs139771344	intronic	KMT2E	.	.	0.005882	0.0008	.	.
7	104748391	C	rs150402862	intronic	KMT2E	.	.	0.03529	0.029	.	.
7	104749687	C	rs57492989	intronic	KMT2E	.	.	0.005882	0.0093	.	.
7	104752838	CCTCCACCT	rs749591342	exonic	KMT2E	nonframeshift insertion	p.P1545delinsPPPP	0.005882	4.95E-05	.	.
7	104753093	A	.	exonic	KMT2E	synonymous SNV	p.P1630P	0.005882	1.65E-05	.	.
7	104753233	ACCCCC	rs751442182	exonic	KMT2E	nonframeshift insertion	p.L1677delinsLPP	0.005882	7.61E-05	.	.
7	117359713	A	rs199539621	intronic	CTTNBP2	.	.	0.005882	0.0002	.	.
7	117365146	C	rs143800569	exonic	CTTNBP2	synonymous SNV	p.L1407L	0.005882	0.0002	.	.
7	117365348	A	rs149317537	intronic	CTTNBP2	.	.	0.01176	0.0012	.	.
7	117395664	G	rs34491454	intronic	CTTNBP2	.	.	0.01765	0.0253	.	.
7	117420450	C	rs77345073	intronic	CTTNBP2	.	.	0.005882	0.0015	.	.
7	117422868	C	rs78183633	intronic	CTTNBP2	.	.	0.01176	0.0278	.	.
7	117424277	A	rs7811545	intronic	CTTNBP2	.	.	0.005882	0.0036	.	.
7	117424291	G	rs28609642	intronic	CTTNBP2	.	.	0.005882	0.0048	.	.
7	117431142	T	rs147505672	intronic	CTTNBP2	.	.	0.005882	0.0066	.	.
7	117432212	A	.	exonic	CTTNBP2	synonymous SNV	p.C346C	0.005882	.	.	.
7	117450792	C	rs2111204	intronic	CTTNBP2	.	.	0.6529	0.6314	.	.
7	117450802	G	rs41281090	intronic	CTTNBP2	.	.	0.005882	0.0064	.	.
7	117511581	G	rs10277241	intronic	CTTNBP2	.	.	0.03529	.	.	.
7	151842397	T	rs2240819	intronic	KMT2C	.	.	0.02941	0.0713	.	.
7	151847946	A	rs3757422	intronic	KMT2C	.	.	0.02353	0.0511	.	.
7	151851392	T	rs56753294	exonic	KMT2C	synonymous SNV	p.P4033P	0.005882	0.014	.	.
7	151851544	C	rs79605387	intronic	KMT2C	.	.	0.005882	0.0023	.	.
7	151853463	T	rs536483249	intronic	KMT2C	.	.	0.005882	0.0006	.	.
7	151855922	C	rs111293424	intronic	KMT2C	.	.	0.005882	1.65E-05	.	.
7	151859751	A	rs138353962	exonic	KMT2C	synonymous SNV	p.I3637I	0.01176	0.0044	.	.
7	151860846	C	rs113138058	exonic	KMT2C	synonymous SNV	p.Q3272Q	0.005882	3.30E-05	.	.

7	151873853	T	rs6464211	exonic	KMT2C	synonymous SNV	p.Q2895Q	0.1941	0.2494	.	.
7	151874498	T	rs10252263	exonic	KMT2C	synonymous SNV	p.Q2680Q	0.02941	0.0705	.	.
7	151877128	C	rs138464665	exonic	KMT2C	synonymous SNV	p.S2411S	0.005882	0.0008	.	.
7	151877889	C	rs17173370	exonic	KMT2C	synonymous SNV	p.V2352V	0.005882	0.0094	.	.
7	151879593	TGCTGC	rs749417254	exonic	KMT2C	nonframeshift insertion	p.Q1784delinsQQQ	0.005882	6.64E-05	.	.
7	151884447	T	rs61730536	exonic	KMT2C	synonymous SNV	p.T1636T	0.005882	0.0096	.	.
7	151884583	C	rs766239018	intronic	KMT2C	.	.	0.005882	8.37E-06	.	.
7	151884607	T	rs3800834	intronic	KMT2C	.	.	0.02941	0.0698	.	.
7	151891051	T	rs200920682	intronic	KMT2C	.	.	0.005882	0.0002	.	.
7	151896350	C	rs10487890	intronic	KMT2C	.	.	0.005882	0.0137	.	.
7	151896573	GTTA	rs112572300	intronic	KMT2C	.	.	0.02941	0.0704	.	.
7	152055777	T	rs187902675	intronic	KMT2C	.	.	0.01176	0.0061	.	.
7	152132812	T	rs191834730	exonic	KMT2C	synonymous SNV	p.E20E	0.005882	0.0223	.	.
8	61655690	A	rs7836586	intronic	CHD7	.	.	0.7294	0.8119	.	.
8	61693942	AAAGCA	rs377139749	exonic	CHD7	nonframeshift insertion	p.K683delinsKKA	0.02353	0.0068	.	.
8	61694001	A	.	intronic	CHD7	.	.	0.005882	.	.	.
8	61707572	C	rs79302359	exonic	CHD7	synonymous SNV	p.S708S	0.01176	0.0135	.	.
8	61707725	A	rs4540437	intronic	CHD7	.	.	0.7588	0.8316	.	.
8	61713126	TGGACT	rs397687085	intronic	CHD7	.	.	0.7412	0.7954	.	.
8	61714190	T	rs41272438	intronic	CHD7	.	.	0.05294	0.0284	.	.
8	61732518	G	rs79276682	intronic	CHD7	.	.	0.005882	0.0036	.	.
8	61732521	G	rs6471902	intronic	CHD7	.	.	0.7471	0.8259	.	.
8	61741378	A	rs199581494	intronic	CHD7	.	.	0.01176	0.0008	.	.
8	61742846	G	rs41272442	intronic	CHD7	.	.	0.01176	0.0228	.	.
8	61748893	A	rs7005873	intronic	CHD7	.	.	0.7353	.	.	.
8	61750718	A	rs41265246	exonic	CHD7	synonymous SNV	p.G1479G	0.005882	0.0009	.	.
8	61750860	G	rs7844902	intronic	CHD7	.	.	0.7176	0.8137	.	.
8	61757805	T	rs71640288	intronic	CHD7	.	.	0.02941	0.0056	.	.
8	61758019	A	rs143263433	intronic	CHD7	.	.	0.01176	.	.	.
8	61761171	T	rs151322460	intronic	CHD7	.	.	0.005882	0.0008	.	.
8	61764838	G	rs41265252	intronic	CHD7	.	.	0.01176	0.0076	.	.
8	61765273	T	rs3763592	intronic	CHD7	.	.	0.04706	0.0718	.	.
8	61765395	T	rs41312170	exonic	CHD7	synonymous SNV	p.P2037P	0.01176	0.0068	.	.
8	61765419	A	rs6999971	exonic	CHD7	synonymous SNV	p.P2045P	0.04706	0.0285	.	.
8	61765500	G	rs199828744	exonic	CHD7	synonymous SNV	p.P2072P	0.005882	0.0003	.	.
8	61765560	A	rs2068096	exonic	CHD7	synonymous SNV	p.E2092E	0.04118	0.0657	.	.
8	61768716	T	rs201046385	exonic	CHD7	synonymous SNV	p.S2373S	0.005882	5.76E-05	.	.
8	61769195	G	rs2272727	exonic	CHD7	synonymous SNV	p.T2452T	0.06471	0.0472	.	.
8	61774901	T	rs769103057	intronic	CHD7	.	.	0.005882	5.28E-05	.	.
8	61774929	A	rs774062065	intronic	CHD7	.	.	0.005882	0.0001	.	.
9	96392371	A	rs10992813	intronic	PHF2	.	.	0.3706	0.3011	.	.
9	96392389	A	rs12553775	intronic	PHF2	.	.	0.1059	0.086	.	.
9	96398655	T	rs148494944	intronic	PHF2	.	.	0.01765	0.0296	.	.
9	96398831	G	rs10992818	intronic	PHF2	.	.	0.1235	0.1621	.	.
9	96407920	T	rs7038310	exonic	PHF2	synonymous SNV	p.D103D	0.09412	0.0739	.	.
9	96407953	A	rs35505758	exonic	PHF2	synonymous SNV	p.T114T	0.01765	0.0186	.	.
9	96407983	T	rs9695734	exonic	PHF2	synonymous SNV	p.T124T	0.1294	0.1316	.	.
9	96411414	A	rs16912641	exonic	PHF2	synonymous SNV	p.L174L	0.02353	0.026	.	.
9	96415482	T	rs56134753	exonic	PHF2	synonymous SNV	p.P208P	0.005882	0.0201	.	.
9	96415653	T	rs3750354	intronic	PHF2	.	.	0.3882	0.3529	.	.
9	96416899	T	rs3750355	intronic	PHF2	.	.	0.01765	0.0209	.	.
9	96418315	T	rs113129271	intronic	PHF2	.	.	0.005882	0.0062	.	.
9	96420390	G	rs3763605	intronic	PHF2	.	.	0.6294	0.5875	.	.
9	96420414	T	rs76256243	intronic	PHF2	.	.	0.005882	0.0065	.	.
9	96422788	T	rs73523907	exonic	PHF2	synonymous SNV	p.L548L	0.01765	0.0221	.	.
9	96425138	C	rs3750358	intronic	PHF2	.	.	0.6353	0.5677	.	.
9	96425174	C	rs78407785	intronic	PHF2	.	.	0.005882	0.0041	.	.
9	96425350	C	rs117578535	intronic	PHF2	.	.	0.005882	0.0094	.	.
9	96425777	T	rs7036592	intronic	PHF2	.	.	0.3882	0.3272	.	.
9	96425915	T	rs138464551	exonic	PHF2	synonymous SNV	p.L645L	0.005882	0.0013	.	.
9	96428243	T	rs41276198	intronic	PHF2	.	.	0.1353	0.1264	.	.
9	96429439	A	rs41297181	exonic	PHF2	synonymous SNV	p.K755K	0.09412	0.0406	.	.
9	96435876	A	rs140677384	exonic	PHF2	synonymous SNV	p.P786P	0.01176	0.0076	.	.
9	96437179	A	rs79020256	intronic	PHF2	.	.	0.005882	0.0042	.	.
9	96438083	T	rs765541178	intronic	PHF2	.	.	0.005882	8.51E-06	.	.
9	96439070	A	rs143554999	exonic	PHF2	synonymous SNV	p.S1009S	0.005882	0.0007	.	.
9	96439979	-	rs139272538	UTR3	PHF2	.	.	0.005882	0.0033	.	.
9	135772681	A	rs45468995	exonic	TSC1	synonymous SNV	p.T904T	0.01176	0.0017	.	.
9	135772717	A	rs4962081	exonic	TSC1	synonymous SNV	p.A892A	0.08824	0.0758	.	.
9	135772977	A	rs118203720	exonic	TSC1	synonymous SNV	p.A831A	0.005882	0.0025	.	.
9	135776925	C	rs75802666	intronic	TSC1	.	.	0.02353	0.0711	.	.
9	135777958	T	rs116518821	intronic	TSC1	.	.	0.005882	0.0043	.	.
9	135781563	A	rs10901220	intronic	TSC1	.	.	0.1647	0.1522	.	.
9	135782221	C	rs7862221	exonic	TSC1	synonymous SNV	p.E394E	0.1647	0.1371	.	.
9	135786112	C	rs6597586	intronic	TSC1	.	.	0.1647	0.129	.	.
9	135797254	G	rs118203414	exonic	TSC1	synonymous SNV	p.S154S	0.005882	0.0006	.	.
9	135802555	T	rs118203350	intronic	TSC1	.	.	0.005882	0.0106	.	.
9	135804266	A	rs62621221	UTR5	TSC1	.	.	0.005882	0.0013	.	.
10	28822853	A	rs182451824	intronic	WAC	.	.	0.01176	.	.	.

10	28822982	T	rs143699084	intronic	WAC	.	.	0.005882	0.0043	.	.
10	28884612	-	rs773739669	intronic	WAC	.	.	0.005882	8.85E-06	.	.
10	28900770	G	rs2232792	exonic	WAC	synonymous SNV	p.S349S	0.005882	0.0077	.	.
10	28905079	G	rs332176	intronic	WAC	.	.	0.2235	0.2075	.	.
10	28905330	G	rs747909234	intronic	WAC	.	.	0.005882	0.0001	.	.
10	89623027	-	rs587781340	UTR5	KLLN	.	.	0.005882	.	.	.
10	89623056	T	rs587779981	UTR5	KLLN	.	.	0.005882	.	.	.
10	114710953	T	rs745800886	intronic	TCF7L2	.	.	0.005882	5.57E-05	.	.
10	114901092	A	rs148377922	intronic	TCF7L2	.	.	0.005882	0.0011	.	.
10	114917865	G	rs199913706	intronic	TCF7L2	.	.	0.005882	0.0024	.	.
10	114920321	T	rs117423278	intronic	TCF7L2	.	.	0.01176	.	.	.
10	114925407	G	rs149031135	exonic	TCF7L2	synonymous SNV	p.P472P	0.01176	0.0045	.	.
10	114925758	C	rs1056877	UTR3	TCF7L2	.	.	0.01176	0.0511	.	.
11	4566513	A	rs11032738	exonic	OR52M1	synonymous SNV	p.I31I	0.01765	0.0046	.	.
11	4566711	T	rs2709182	exonic	OR52M1	synonymous SNV	p.D97D	0.5294	0.4594	.	.
11	4566870	C	rs61747520	exonic	OR52M1	synonymous SNV	p.S150S	0.02941	0.029	.	.
11	4566939	G	rs61747538	exonic	OR52M1	synonymous SNV	p.K173K	0.005882	0.0049	.	.
11	4567140	A	rs61734243	exonic	OR52M1	synonymous SNV	p.T240T	0.005882	0.0086	.	.
11	4567158	CA	rs145064459	exonic	OR52M1	frameshift insertion	p.S246fs	0.005882	0.0025	.	.
11	4567185	C	rs12295898	exonic	OR52M1	synonymous SNV	p.Y255Y	0.08824	0.0697	.	.
11	67926088	-	rs373030307	exonic	KMT5B	nonframeshift deletion	p.335_335del	0.005882	0.0012	.	.
11	67938474	C	rs368120903	intronic	KMT5B	.	.	0.005882	2.51E-05	.	.
11	67938848	C	rs7949511	intronic	KMT5B	.	.	0.1	.	.	.
11	67939134	C	rs140135686	exonic	KMT5B	synonymous SNV	p.A209A	0.005882	0.0001	.	.
11	67941365	T	rs114727354	exonic	KMT5B	synonymous SNV	p.R164R	0.01176	0.005	.	.
11	67947575	C	rs184694662	intronic	KMT5B	.	.	0.005882	0.0008	.	.
11	70331937	A	rs143595073	exonic	SHANK2	synonymous SNV	p.A899A	0.01176	0.0223	.	.
11	70333047	C	rs142550207	exonic	SHANK2	synonymous SNV	p.P529P	0.005882	0.0007	.	.
11	70333655	G	rs782117005	exonic	SHANK2	synonymous SNV	p.R327R	0.005882	6.09E-05	.	.
11	70348009	G	rs7928538	intronic	SHANK2	.	.	0.02353	0.0527	.	.
11	70348104	C	rs370998426	intronic	SHANK2	.	.	0.005882	.	.	.
11	70348261	T	rs146778438	intronic	SHANK2	.	.	0.005882	0.0024	.	.
11	70505882	A	rs188803860	intronic	SHANK2	.	.	0.005882	0.001	.	.
11	70544864	C	rs112497741	exonic	SHANK2	unknown	.	0.01176	0.0132	.	.
11	70544921	A	rs11237214	intronic	SHANK2	.	.	0.06471	0.1214	.	.
11	70666732	T	rs781906747	exonic	SHANK2	unknown	.	0.005882	0.0012	.	.
11	70798896	-	rs35132270	intronic	SHANK2	.	.	0.3	0.3301	.	.
11	70798902	T	rs34650500	intronic	SHANK2	.	.	0.3	0.3155	.	.
11	70798925	-	rs139112112	intronic	SHANK2	.	.	0.005882	0.0145	.	.
11	70805534	T	rs556851353	intronic	SHANK2	.	.	0.005882	0.0006	.	.
11	70824326	A	rs117706585	intronic	SHANK2	.	.	0.005882	0.0222	.	.
11	70829858	C	rs3924047	intronic	SHANK2	.	.	0.5	0.526	.	.
11	70829950	A	rs76014490	exonic	SHANK2	unknown	.	0.005882	0.0109	.	.
11	70858322	A	rs376267466	exonic	SHANK2	unknown	.	0.005882	0.0014	.	.
11	119212363	T	rs138370910	exonic	MFRP	synonymous SNV	p.A545A	0.01176	0.0053	.	.
11	119213303	T	rs11217241	intronic	C1QTNF5,MFRP	.	.	0.1294	0.1683	.	.
11	119213570	T	rs776266547	intronic	C1QTNF5,MFRP	.	.	0.005882	8.39E-06	.	.
11	119215007	G	rs185451482	intronic	C1QTNF5,MFRP	.	.	0.005882	0.0066	.	.
11	119215046	T	rs35885438	exonic	MFRP	synonymous SNV	p.L318L	0.05882	0.0551	.	.
11	119215307	T	rs758656745	intronic	C1QTNF5,MFRP	.	.	0.005882	4.15E-05	.	.
11	119216231	G	rs2510143	exonic	MFRP	synonymous SNV	p.H180H	0.9471	0.9364	.	.
11	119216279	A	rs36015759	exonic	MFRP	synonymous SNV	p.Y164Y	0.2176	0.2136	.	.
11	119216673	A	rs111578461	intronic	C1QTNF5,MFRP	.	.	0.005882	0.0078	.	.
11	119216910	C	.	intronic	C1QTNF5,MFRP	.	.	0.005882	.	.	.
11	119217254	T	rs883247	UTR5	C1QTNF5,MFRP	.	.	0.5647	0.5951	.	.
12	13715954	A	rs1805246	exonic	GRIN2B	synonymous SNV	p.F1406F	0.05294	0.0387	.	.
12	13715975	G	rs1805247	exonic	GRIN2B	synonymous SNV	p.H1399H	0.1059	0.1376	.	.
12	13716638	A	rs1806191	exonic	GRIN2B	synonymous SNV	p.H1178H	0.4882	0.3915	.	.
12	13716674	A	rs45600931	exonic	GRIN2B	synonymous SNV	p.S1166S	0.01765	0.0065	.	.
12	13717508	A	rs1806201	exonic	GRIN2B	synonymous SNV	p.T888T	0.2824	0.3102	.	.
12	13717597	A	rs201881500	intronic	GRIN2B	.	.	0.005882	3.43E-05	.	.
12	13720043	A	rs3026160	exonic	GRIN2B	synonymous SNV	p.C838C	0.1471	0.0887	.	.
12	13722876	T	rs757990373	exonic	GRIN2B	synonymous SNV	p.V749V	0.005882	4.96E-05	.	.
12	13724942	-	rs370168771	intronic	GRIN2B	.	.	0.01176	0.0026	.	.
12	13761741	A	rs1805522	exonic	GRIN2B	synonymous SNV	p.I602I	0.04118	0.065	.	.
12	13764774	A	rs1805482	exonic	GRIN2B	synonymous SNV	p.S555S	0.3529	0.2647	.	.
12	13768586	A	rs35025065	exonic	GRIN2B	synonymous SNV	p.D447D	0.02353	0.0067	.	.
12	13769603	C	rs76777620	intronic	GRIN2B	.	.	0.01176	0.0143	.	.
12	13828659	C	rs11055581	intronic	GRIN2B	.	.	0.1882	0.1577	.	.
12	14018777	C	rs7301328	exonic	GRIN2B	synonymous SNV	p.P122P	0.4294	0.4172	.	.
12	14019128	T	rs34315573	exonic	GRIN2B	synonymous SNV	p.A5A	0.02941	0.0508	.	.
13	51929304	T	rs4942995	intronic	SERPINE3	.	.	0.02941	0.0552	.	.
13	51941943	C	rs61749884	exonic	INTS6	synonymous SNV	p.S856S	0.01176	0.0213	.	.
13	51943057	A	rs45535533	intronic	INTS6	.	.	0.01765	0.0186	.	.
13	51943462	T	rs73195996	intronic	INTS6	.	.	0.02941	0.0531	.	.
13	51950136	-	rs3831047	intronic	INTS6	.	.	0.01176	0.0183	.	.
13	51956122	C	rs9526753	intronic	INTS6	.	.	0.5882	0.697	.	.
13	51956225	G	rs61749885	exonic	INTS6	synonymous SNV	p.T405T	0.01176	0.0215	.	.
13	51963595	-	rs368336972	intronic	INTS6	.	.	0.01765	0.0336	.	.

13	51963594	AAG	rs5769933324	intronic	INTS6	.	.	0.005882	0.0034	.	.
14	21853853	G	rs61748933	exonic	CHD8	synonymous SNV	p.D2555D	0.02941	0.014	.	.
14	21859080	-	rs35057134	intronic	CHD8	.	.	0.3294	0.3349	.	.
14	21861631	C	rs779588873	intronic	CHD8	.	.	0.005882	0	.	.
14	21862388	A	rs199912058	intronic	CHD8	.	.	0.005882	0.0001	.	.
14	21862666	G	rs77562111	intronic	CHD8	.	.	0.005882	0.0027	.	.
14	21863396	A	rs184930403	intronic	CHD8	.	.	0.005882	.	.	.
14	21866026	C	rs61752837	exonic	CHD8	synonymous SNV	p.A1669A	0.005882	0.0238	.	.
14	21867754	C	rs375587003	intronic	CHD8	.	.	0.005882	0.0009	.	.
14	21868798	A	rs17792647	intronic	CHD8	.	.	0.01176	0.0145	.	.
14	21869708	C	rs199499304	intronic	CHD8	.	.	0.01176	0.0321	.	.
14	21871172	C	rs191258109	intronic	CHD8	.	.	0.005882	0.0002	.	.
14	21871653	T	rs8022395	exonic	CHD8	synonymous SNV	p.V1159V	0.9471	0.9093	.	.
14	21874068	C	rs7155123	intronic	CHD8	.	.	0.9294	0.8616	.	.
14	21876729	C	rs374458289	intronic	CHD8	.	.	0.005882	0.0011	.	.
14	21876762	A	rs145389674	intronic	CHD8	.	.	0.01765	0.0267	.	.
14	21876976	T	rs61752838	exonic	CHD8	synonymous SNV	p.P791P	0.005882	0.0024	.	.
14	21877015	A	rs202188781	intronic	CHD8	.	.	0.005882	0.0003	.	.
14	21877019	A	rs200436134	intronic	CHD8	.	.	0.005882	0.0003	.	.
14	21883866	C	.	intronic	CHD8	.	.	0.005882	.	.	.
14	21896192	A	rs61744173	exonic	CHD8	synonymous SNV	p.N479N	0.01765	0.0084	.	.
14	21899909	T	.	intronic	CHD8	.	.	0.005882	.	.	.
14	77491829	A	rs145752129	exonic	IRF2BPL	synonymous SNV	p.S769S	0.005882	0.0022	.	.
14	77492891	A	rs879027	exonic	IRF2BPL	synonymous SNV	p.Y415Y	0.5882	0.478	.	.
14	77494484	C	rs76980172	UTR5	IRF2BPL	.	.	0.02941	.	.	.
14	77494557	T	rs3742745	UTR5	IRF2BPL	.	.	0.005882	.	.	.
14	77494578	G	.	UTR5	IRF2BPL	.	.	0.005882	.	.	.
14	77494682	A	rs12897921	UTR5	IRF2BPL	.	.	0.5588	.	.	.
14	77494722	T	rs78010105	UTR5	IRF2BPL	.	.	0.08235	.	.	.
15	26806064	C	rs3751582	intronic	GABRB3	.	.	0.2647	0.3456	.	.
15	26812780	T	rs76812964	exonic	GABRB3	synonymous SNV	p.S176S	0.02941	0.0247	.	.
15	27017536	A	rs8179186	intronic	GABRB3	.	.	0.1471	0.2221	.	.
15	27018797	A	rs20318	exonic	GABRB3	synonymous SNV	p.P25P	0.1588	0.2983	.	.
15	27128252	G	rs74006529	intronic	GABRA5	.	.	0.005882	0.0018	.	.
15	27128254	G	rs73363996	intronic	GABRA5	.	.	0.01765	0.0206	.	.
15	27128461	T	rs140680	intronic	GABRA5	.	.	0.08824	0.1191	.	.
15	27128498	T	rs79274924	exonic	GABRA5	synonymous SNV	p.D97D	0.005882	0.0005	.	.
15	27182357	C	rs140682	exonic	GABRA5	synonymous SNV	p.V202V	0.4765	0.5927	.	.
15	27182491	C	.	intronic	GABRA5	.	.	0.005882	.	.	.
15	93485032	T	rs141271290	intronic	CHD2	.	.	0.005882	0.0133	.	.
15	93485184	T	rs372257500	exonic	CHD2	synonymous SNV	p.G275G	0.005882	.	.	.
15	93498721	C	rs144093014	exonic	CHD2	synonymous SNV	p.Y596Y	0.005882	0.006	.	.
15	93510540	G	rs201137739	intronic	CHD2	.	.	0.005882	0.0006	.	.
15	93510603	G	rs4777755	exonic	CHD2	synonymous SNV	p.E683E	0.8529	0.8419	.	.
15	93514952	T	rs189630679	intronic	CHD2	.	.	0.005882	0.0002	.	.
15	93521604	G	rs11074121	exonic	CHD2	synonymous SNV	p.Q906Q	0.8471	0.8353	.	.
15	93521651	G	rs11074122	intronic	CHD2	.	.	0.8471	0.8353	.	.
15	93523998	G	rs77401998	intronic	CHD2	.	.	0.005882	0.0105	.	.
15	93524149	T	.	intronic	CHD2	.	.	0.005882	.	.	.
15	93527619	T	rs150268140	exonic	CHD2	synonymous SNV	p.D1042D	0.01176	0.0044	.	.
15	93528716	T	.	intronic	CHD2	.	.	0.005882	.	.	.
15	93536185	G	rs144292068	exonic	CHD2	synonymous SNV	p.A1184A	0.005882	0.0006	.	.
15	93536197	T	rs2272457	exonic	CHD2	synonymous SNV	p.Y1188Y	0.2647	0.2266	.	.
15	93543907	G	rs147884853	intronic	CHD2	.	.	0.01176	0.006	.	.
15	93552330	T	rs12915582	intronic	CHD2	.	.	0.8529	0.8369	.	.
15	93552349	A	rs72647789	intronic	CHD2	.	.	0.02941	0.0259	.	.
15	93552488	T	rs34315566	exonic	CHD2	synonymous SNV	p.I1509I	0.05294	0.046	.	.
15	93555541	G	rs117430127	intronic	CHD2	.	.	0.01176	0.0082	.	.
15	93555626	A	rs138836603	exonic	CHD2	synonymous SNV	p.L1548L	0.005882	8.28E-06	.	.
15	93555717	G	rs2119010	intronic	CHD2	.	.	0.2235	0.1812	.	.
15	93563483	C	rs773497328	exonic	CHD2	synonymous SNV	p.Y1716Y	0.005882	8.25E-06	.	.
15	93567864	C	rs12906163	exonic	CHD2	synonymous SNV	p.R1806R	0.3588	0.2946	.	.
16	2103417	A	rs45517100	exonic	TSC2	synonymous SNV	p.A100A	0.005882	0.0003	.	.
16	2104475	T	rs376332051	intronic	TSC2	.	.	0.005882	2.48E-05	.	.
16	2105400	T	rs1800720	intronic	TSC2	.	.	0.09412	0.0871	.	.
16	2110805	A	rs1800742	exonic	TSC2	synonymous SNV	p.Q370Q	0.01176	0.011	.	.
16	2112651	-	rs137854304	intronic	TSC2	.	.	0.005882	0.0013	.	.
16	2112941	G	rs45517166	intronic	TSC2	.	.	0.01176	0.0026	.	.
16	2114407	T	rs34012042	exonic	TSC2	synonymous SNV	p.S526S	0.06471	0.0573	.	.
16	2115481	T	rs45477195	intronic	TSC2	.	.	0.07059	0.0712	.	.
16	2115506	T	rs45517185	intronic	TSC2	.	.	0.06471	0.0566	.	.
16	2121869	T	rs45517208	exonic	TSC2	synonymous SNV	p.P677P	0.005882	0.0033	.	.
16	2121978	G	rs186681035	intronic	TSC2	.	.	0.01176	0.0104	.	.
16	2122822	G	rs7196184	intronic	TSC2	.	.	0.02941	0.0324	.	.
16	2124416	A	rs45517242	intronic	TSC2	.	.	0.01765	0.0028	.	.
16	2125769	A	rs145738496	intronic	TSC2	.	.	0.01765	0.0114	.	.
16	2125788	T	rs13331451	intronic	TSC2	.	.	0.08824	0.1049	.	.
16	2125834	C	rs13337626	exonic	TSC2	synonymous SNV	p.F860F	0.08235	0.0725	.	.
16	2125935	T	rs201973730	intronic	TSC2	.	.	0.005882	0.0002	.	.

16	2125937	G	rs1800715	intronic	TSC2	.	.	0.1	0.1163	.	.
16	2126452	G	rs45517265	intronic	TSC2	.	.	0.01176	0.0072	.	.
16	2129045	A	rs45517277	exonic	TSC2	synonymous SNV	p.T949T	0.005882	0.0009	.	.
16	2131729	T	rs45517307	exonic	TSC2	synonymous SNV	p.A1204A	0.005882	0.0003	.	.
16	2133727	A	rs11551373	exonic	TSC2	synonymous SNV	p.P1238P	0.01176	0.0103	.	.
16	2134221	T	rs45517325	intronic	TSC2	.	.	0.005882	0.0049	.	.
16	2134982	-	rs137854239	exonic	TSC2	nonframeshift deletion	p.1441_1442del	0.01176	0.0053	.	.
16	2134994	T	rs35986575	exonic	TSC2	synonymous SNV	p.D1445D	0.01176	0.0028	.	.
16	2135073	T	rs45482793	intronic	TSC2	.	.	0.01765	0.0103	.	.
16	2136842	T	rs45517384	exonic	TSC2	synonymous SNV	p.S1586S	0.005882	0.0133	.	.
16	2137925	-	rs137854209	exonic	TSC2	nonframeshift deletion	p.1617_1623del	0.005882	0.0023	.	.
16	2138218	C	rs1800718	intronic	TSC2	.	.	0.2235	0.2067	.	.
16	2138219	T	rs45515893	intronic	TSC2	.	.	0.005882	0.0035	.	.
16	2138269	C	rs1748	exonic	TSC2	synonymous SNV	p.D1667D	0.1765	0.1883	.	.
16	2138398	T	rs13332221	intronic	TSC2	.	.	0.1176	0.1102	.	.
16	2138422	G	rs13332222	intronic	TSC2	.	.	0.1235	0.121	.	.
16	2138584	C	rs1051771	exonic	TSC2	synonymous SNV	p.S1732S	0.07647	0.0767	.	.
17	3577146	T	rs1063497	ncRNA_intronic	P2RX5-TAX1BP3	.	.	0.1412	0.125	.	.
17	3585135	G	rs368454045	ncRNA_intronic	P2RX5-TAX1BP3	.	.	0.005882	.	.	.
17	3585283	A	rs117935905	ncRNA_intronic	P2RX5-TAX1BP3	.	.	0.02353	0.0087	.	.
17	3591453	C	rs78411513	ncRNA_intronic	P2RX5-TAX1BP3	.	.	0.01176	.	.	.
17	3592740	C	rs220487	ncRNA_intronic	P2RX5-TAX1BP3	.	.	0.1588	0.1759	.	.
17	3594277	-	rs3215407	exonic	P2RX5	frameshift deletion	p.P111fs	0.6235	0.6676	.	.
17	3594965	T	rs754977980	exonic	P2RX5	frameshift insertion	p.D87fs	0.005882	7.43E-05	.	.
17	3595137	A	rs761882993	ncRNA_intronic	P2RX5-TAX1BP3	.	.	0.005882	4.14E-05	.	.
17	3599115	G	rs558854970	ncRNA_intronic	P2RX5-TAX1BP3	.	.	0.01176	0.0038	.	.
17	7749044	G	rs140362144	intronic	KDM6B	.	.	0.005882	0.0005	.	.
17	7749621	G	rs80152199	intronic	KDM6B	.	.	0.06471	0.117	.	.
17	7750083	A	rs117985215	intronic	KDM6B	.	.	0.01176	0.0225	.	.
17	7750262	G	rs146637535	exonic	KDM6B	synonymous SNV	p.P279P	0.005882	0.0013	.	.
17	7750357	C	rs56880362	intronic	KDM6B	.	.	0.02353	0.0482	.	.
17	7750847	C	rs2270517	intronic	KDM6B	.	.	0.2294	0.3594	.	.
17	7751388	A	rs3744247	exonic	KDM6B	synonymous SNV	p.P594P	0.07647	0.0952	.	.
17	7751751	T	rs3744248	exonic	KDM6B	synonymous SNV	p.H715H	0.05882	0.0941	.	.
17	7752258	A	rs779781827	exonic	KDM6B	synonymous SNV	p.A884A	0.005882	0.0005	.	.
17	7752900	C	rs748469270	exonic	KDM6B	synonymous SNV	p.G1098G	0.005882	2.62E-05	.	.
17	7754296	-	rs146364592	intronic	KDM6B	.	.	0.07647	0.1134	.	.
17	7754460	A	rs3736306	exonic	KDM6B	synonymous SNV	p.I1265I	0.01176	0.023	.	.
17	7755980	T	rs140381590	intronic	KDM6B	.	.	0.01176	0.0273	.	.
17	29483108	T	rs17881168	exonic	NF1	synonymous SNV	p.S56S	0.03529	0.0119	.	.
17	29486152	A	rs2952976	intronic	NF1	.	.	0.6471	0.623	.	.
17	29490200	G	.	intronic	NF1	.	.	0.005882	.	.	.
17	29508699	C	rs182325576	intronic	NF1	.	.	0.005882	0.0021	.	.
17	29508775	A	rs1801052	exonic	NF1	synonymous SNV	p.L234L	0.6471	0.6228	.	.
17	29509641	A	rs138840528	exonic	NF1	synonymous SNV	p.Q282Q	0.01765	0.0061	.	.
17	29527406	-	.	intronic	NF1	.	.	0.005882	.	.	.
17	29545987	T	rs112806382	intronic	NF1	.	.	0.6471	0.6585	.	.
17	29546175	C	rs2905880	intronic	NF1	.	.	0.6471	0.634	.	.
17	29552064	C	.	intronic	NF1	.	.	0.005882	.	.	.
17	29553485	A	rs2285892	exonic	NF1	synonymous SNV	p.P678P	0.3588	0.3827	.	.
17	29554205	T	rs141082540	intronic	NF1	.	.	0.005882	0.0042	.	.
17	29556837	C	rs17880825	intronic	NF1	.	.	0.02353	0.0176	.	.
17	29559871	T	rs147955381	exonic	NF1	synonymous SNV	p.N1156N	0.005882	0.0007	.	.
17	29559932	A	rs2066736	intronic	NF1	.	.	0.3588	0.3676	.	.
17	29560159	A	rs145126193	exonic	NF1	synonymous SNV	p.V1212V	0.005882	0.0001	.	.
17	29562893	C	rs370179525	intronic	NF1	.	.	0.01176	0.0004	.	.
17	29563079	G	rs760137326	intronic	NF1	.	.	0.005882	0.0305	.	.
17	29587341	C	rs17881285	intronic	NF1	.	.	0.05294	0.053	.	.
17	29632664	C	rs16972176	intronic	EVI2B,NF1	.	.	0.005882	0.0033	.	.
17	29653237	A	rs17887014	exonic	NF1	synonymous SNV	p.K1724K	0.01176	0.0034	.	.
17	29654876	A	rs2285894	intronic	NF1	.	.	0.5706	0.5364	.	.
17	29664354	A	rs370220255	intronic	NF1	.	.	0.005882	0.0003	.	.
17	29664645	A	rs17883614	intronic	NF1	.	.	0.02353	0.0534	.	.
17	29670190	G	rs7405740	intronic	NF1	.	.	0.9235	0.9069	.	.
17	29677163	G	rs768496641	intronic	NF1	.	.	0.005882	8.24E-06	.	.
17	29677357	G	rs17881641	intronic	NF1	.	.	0.005882	0.003	.	.
17	29679246	A	rs964288	intronic	NF1	.	.	0.5824	0.5334	.	.
17	29683632	G	.	intronic	NF1	.	.	0.005882	.	.	.
17	29683646	A	rs199641099	intronic	NF1	.	.	0.005882	0.0071	.	.
17	29685660	A	rs55747230	intronic	NF1	.	.	0.02941	0.0126	.	.
17	29685689	T	rs114915525	intronic	NF1	.	.	0.005882	0.002	.	.
17	29686024	A	rs2285895	exonic	NF1	synonymous SNV	p.P2696P	0.005882	0.0017	.	.
17	29709035	G	.	intergenic	NF1,RAB11FIP4	.	.	0.005882	.	.	.
18	19345704	C	rs3017047	intronic	MIB1	.	.	0.01176	0.0109	.	.
18	19345932	T	rs3017048	intronic	MIB1	.	.	0.01176	0.0109	.	.
18	19348555	C	rs72886651	intronic	MIB1	.	.	0.005882	0.0036	.	.
18	19383993	G	rs12605999	intronic	MIB1	.	.	0.01176	0.0194	.	.
18	19408950	T	rs9989532	ncRNA_exonic	MIR133A1HG	.	.	0.9647	0.9617	.	.
18	19427096	G	rs11877131	intronic	MIB1	.	.	0.09412	0.0857	.	.

18	19444661	C	rs76056004	UTR3	MIB1	.	.	0.09412	0.075	.	.
18	44559578	A	rs149717505	exonic	TCEB3B	synonymous SNV	p.S686S	0.005882	0.0007	.	.
18	44561795	T	rs2571026	UTR5	TCEB3B	.	.	0.5353	.	.	.
18	44562030	A	rs2576049	intronic	KATNAL2	.	.	0.01765	.	.	.
18	44580762	A	rs144819676	intronic	KATNAL2	.	.	0.005882	0.0034	.	.
18	44584675	A	rs144860304	exonic	KATNAL2	synonymous SNV	p.A62A	0.005882	0.0014	.	.
18	44584734	T	rs112463898	intronic	KATNAL2	.	.	0.005882	0.0082	.	.
18	44589342	A	.	splicing	KATNAL2	.	.	0.005882	.	.	23.6
18	44589450	T	rs80196369	intronic	KATNAL2	.	.	0.005882	0.0142	.	.
18	44589744	C	.	intronic	KATNAL2	.	.	0.005882	.	.	.
18	44595647	T	rs56297904	exonic	KATNAL2	synonymous SNV	p.G222G	0.01176	0.0108	.	.
18	44595809	T	rs2289036	intronic	KATNAL2	.	.	0.5412	0.553	.	.
18	44601611	G	rs3816125	intronic	KATNAL2	.	.	0.3941	0.417	.	.
18	44626630	G	rs2289130	exonic	KATNAL2	synonymous SNV	p.T388T	0.07059	0.0818	.	.
18	44627200	T	rs117014566	intronic	KATNAL2	.	.	0.005882	0.0015	.	.
19	9449149	G	rs8110197	intronic	ZNF559,ZNF559-ZNF177	.	.	0.5235	0.5874	.	.
19	9449337	A	rs8110584	intronic	ZNF559,ZNF559-ZNF177	.	.	0.5294	.	.	.
19	9450029	C	rs16979618	intronic	ZNF559,ZNF559-ZNF177	.	.	0.02353	0.0218	.	.
19	9451722	T	rs78740264	intronic	ZNF559,ZNF559-ZNF177	.	.	0.005882	.	.	.
19	9452868	A	rs16979666	exonic	ZNF559	synonymous SNV	p.G311G	0.02353	0.021	.	.
19	39221716	-	rs141208726	intronic	CAPN12	.	.	0.1941	0.3155	.	.
19	39224413	G	rs4801861	exonic	CAPN12	synonymous SNV	p.F629F	0.8059	0.7654	.	.
19	39224858	T	rs377079795	intronic	CAPN12	.	.	0.005882	0.0002	.	.
19	39224902	A	rs35807146	intronic	CAPN12	.	.	0.06471	.	.	.
19	39224934	T	rs62120072	intronic	CAPN12	.	.	0.06471	0.0632	.	.
19	39225046	A	rs45599933	intronic	CAPN12	.	.	0.05882	0.0629	.	.
19	39225079	T	rs117513237	intronic	CAPN12	.	.	0.02941	0.0138	.	.
19	39226223	A	rs114233772	intronic	CAPN12	.	.	0.005882	0.0074	.	.
19	39226249	A	rs117242020	intronic	CAPN12	.	.	0.01176	.	.	.
19	39228868	CA	rs34488364	intronic	CAPN12	.	.	0.7824	0.7243	.	.
19	39229316	G	rs12983550	intronic	CAPN12	.	.	0.06471	0.064	.	.
19	39230748	G	rs34911882	exonic	CAPN12	synonymous SNV	p.L224L	0.06471	0.0659	.	.
19	39233146	G	rs936524	exonic	CAPN12	synonymous SNV	p.A110A	0.7882	0.7278	.	.
19	39234538	CCCAGAG	rs772210026	intronic	CAPN12	.	.	0.005882	0.0002	.	.
19	51850304	A	rs144640661	exonic	ETFB	synonymous SNV	p.F240F	0.005882	0.0076	.	.
19	51857614	G	rs141529162	exonic	ETFB	frameshift insertion	p.P93fs	0.05294	0.0337	.	.
19	51869456	A	rs370777932	intronic	ETFB	.	.	0.005882	.	.	.
19	51869638	A	rs111454736	UTR5	ETFB	.	.	0.02353	.	.	.
20	25434059	T	rs16987767	UTR3	NINL	.	.	0.04706	0.0504	.	.
20	25436283	G	rs2235607	intronic	NINL	.	.	0.1647	0.1682	.	.
20	25436301	C	rs73333584	intronic	NINL	.	.	0.005882	0.0046	.	.
20	25436466	T	rs12481409	intronic	NINL	.	.	0.1294	0.1736	.	.
20	25442989	GGGAGCC	rs3036810	intronic	NINL	.	.	0.2471	0.3249	.	.
20	25443076	G	.	exonic	NINL	synonymous SNV	p.V1175V	0.005882	.	.	.
20	25448153	T	rs56795563	intronic	NINL	.	.	0.005882	.	.	.
20	25456888	G	rs437635	exonic	NINL	synonymous SNV	p.S1013S	0.4118	0.466	.	.
20	25459660	C	rs148734140	exonic	NINL	synonymous SNV	p.T700T	0.005882	0.0018	.	.
20	25459764	G	rs11905437	exonic	NINL	synonymous SNV	p.R666R	0.1118	0.0989	.	.
20	25469844	G	rs2072977	intronic	NINL	.	.	0.1647	0.1684	.	.
20	25469969	T	rs181531254	intronic	NINL	.	.	0.005882	0.0051	.	.
20	25472059	T	rs45529236	exonic	NINL	synonymous SNV	p.A471A	0.02353	0.0069	.	.
20	25477427	T	rs143652872	exonic	NINL	synonymous SNV	p.E394E	0.01176	0.0113	.	.
20	25481693	T	rs200872736	intronic	NINL	.	.	0.005882	0.0006	.	.
20	25484705	A	rs544525213	exonic	NINL	synonymous SNV	p.D248D	0.005882	4.30E-05	.	.
20	25493446	T	rs111851591	intronic	NINL	.	.	0.005882	0.0069	.	.
20	25507012	-	rs10536286	intronic	NINL	.	.	0.05294	0.0485	.	.
20	25507019	A	rs11087523	intronic	NINL	.	.	0.05294	0.0466	.	.
20	25507268	C	rs141651341	intronic	NINL	.	.	0.01176	0.0051	.	.
20	49507924	T	rs142825371	UTR3	ADNP	.	.	0.005882	0.0131	.	.
20	49507972	A	rs142247083	exonic	ADNP	synonymous SNV	p.A1093A	0.005882	0.0009	.	.
20	49508320	C	rs144684998	exonic	ADNP	synonymous SNV	p.G977G	0.005882	0.0004	.	.
20	49508683	A	rs1062651	exonic	ADNP	synonymous SNV	p.V856V	0.01176	0.0423	.	.
20	49508776	A	rs148502910	exonic	ADNP	synonymous SNV	p.G825G	0.005882	0.005	.	.
20	49509175	T	rs6096168	exonic	ADNP	synonymous SNV	p.K692K	0.01176	0.0226	.	.
20	49509184	A	rs17790938	exonic	ADNP	synonymous SNV	p.G689G	0.08235	0.0823	.	.
20	49518583	A	rs755893677	exonic	ADNP	synonymous SNV	p.L58L	0.005882	1.65E-05	.	.
20	49520385	G	.	intronic	ADNP	.	.	0.005882	.	.	.
21	38850470	G	rs140466090	intronic	DYRK1A	.	.	0.005882	0.0034	.	.
21	38850640	A	rs928763	intronic	DYRK1A	.	.	0.8412	0.8649	.	.
21	38858938	A	rs55650427	intronic	DYRK1A	.	.	0.2471	0.2337	.	.
21	38862643	A	rs138086853	exonic	DYRK1A	synonymous SNV	p.A277A	0.005882	0.0005	.	.
21	38862794	G	.	intronic	DYRK1A	.	.	0.005882	.	.	.
21	38865507	G	rs2835772	intronic	DYRK1A	.	.	0.2824	0.31	.	.
21	38877544	G	rs188224885	intronic	DYRK1A	.	.	0.005882	0.0024	.	.
21	41414267	T	rs532903018	intronic	DSCAM	.	.	0.005882	2.69E-05	.	.
21	41414279	C	rs148999611	intronic	DSCAM	.	.	0.005882	0.0005	.	.
21	41414590	C	rs200764944	exonic	DSCAM	synonymous SNV	p.R1798R	0.005882	0.0026	.	.
21	41415963	A	rs776908213	intronic	DSCAM	.	.	0.005882	8.35E-06	.	.
21	41416066	T	rs16999204	exonic	DSCAM	synonymous SNV	p.R1774R	0.04706	0.0477	.	.

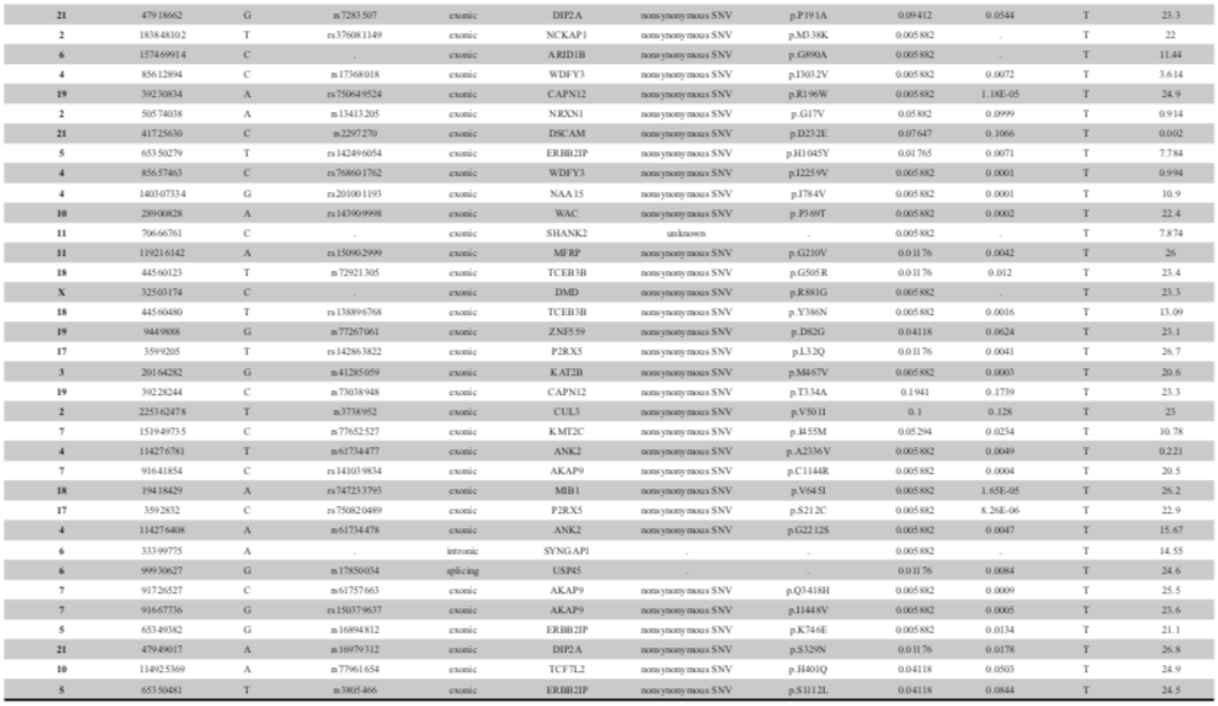
21	41416108	C	rs201433234	exonic	DSCAM	synonymous SNV	p.S1760S	0.01176	0.0068	.	.
21	41427615	G	rs76885144	intronic	DSCAM	.	.	0.005882	0.0068	.	.
21	41434841	C	rs2837409	intronic	DSCAM	.	.	0.2412	0.2931	.	.
21	41446936	A	rs7275294	intronic	DSCAM	.	.	0.04706	0.04	.	.
21	41447058	C	rs7275460	exonic	DSCAM	synonymous SNV	p.T1598T	0.04706	0.0379	.	.
21	41450573	T	rs2297259	intronic	DSCAM	.	.	0.08235	0.109	.	.
21	41450636	A	rs201376842	exonic	DSCAM	synonymous SNV	p.F1563F	0.005882	4.16E-05	.	.
21	41452029	T	rs73221371	intronic	DSCAM	.	.	0.005882	0.0027	.	.
21	41452034	T	rs2837424	intronic	DSCAM	.	.	0.1824	0.179	.	.
21	41455854	A	rs2297263	exonic	DSCAM	synonymous SNV	p.N1404N	0.05294	0.0471	.	.
21	41457502	G	rs41462546	intronic	DSCAM	.	.	0.01176	0.0156	.	.
21	41465664	A	rs62237594	exonic	DSCAM	synonymous SNV	p.V1278V	0.02353	0.0184	.	.
21	41465748	A	rs41445251	exonic	DSCAM	synonymous SNV	p.N1250N	0.005882	0.0054	.	.
21	41539259	A	rs74862130	intronic	DSCAM	.	.	0.01176	0.0381	.	.
21	41539263	G	rs73362176	intronic	DSCAM	.	.	0.005882	0.019	.	.
21	41559182	A	rs2297267	exonic	DSCAM	synonymous SNV	p.P885P	0.03529	0.0292	.	.
21	41559210	T	rs368137392	intronic	DSCAM	.	.	0.005882	3.43E-05	.	.
21	41559804	C	rs190516818	intronic	DSCAM	.	.	0.005882	0.0004	.	.
21	41648220	C	rs74381926	intronic	DSCAM	.	.	0.01765	0.0057	.	.
21	41683969	G	rs16999660	intronic	DSCAM	.	.	0.005882	0.0063	.	.
21	41684090	T	rs34336407	exonic	DSCAM	synonymous SNV	p.S660S	0.05882	0.0759	.	.
21	41710065	C	rs76195942	exonic	DSCAM	synonymous SNV	p.Q582Q	0.005882	0.0019	.	.
21	41710083	A	rs139863593	exonic	DSCAM	synonymous SNV	p.N576N	0.01176	0.0035	.	.
21	41711016	T	rs3215894	intronic	DSCAM	.	.	0.01176	0.0409	.	.
21	41725579	A	rs41367350	exonic	DSCAM	synonymous SNV	p.L249L	0.02353	0.0163	.	.
21	41741016	G	rs2837585	intronic	DSCAM	.	.	0.02353	0.0371	.	.
21	42064785	T	rs79669041	exonic	DSCAM	synonymous SNV	p.A153A	0.01176	0.0085	.	.
21	42064833	C	rs375720021	exonic	DSCAM	synonymous SNV	p.R137R	0.005882	0.0002	.	.
21	42218601	T	rs748752340	UTR5	DSCAM	.	.	0.005882	5.31E-05	.	.
21	47904792	-	rs534623367	intronic	DIP2A	.	.	0.01176	0.0005	.	.
21	47910523	G	rs7279002	exonic	DIP2A	synonymous SNV	p.P58P	0.3118	0.3766	.	.
21	47910659	G	rs9976530	intronic	DIP2A	.	.	0.4412	0.3944	.	.
21	47910668	T	rs80349069	intronic	DIP2A	.	.	0.005882	0.0006	.	.
21	47917048	-	rs11315869	intronic	DIP2A	.	.	0.8765	0.8051	.	.
21	47918652	A	rs763189134	exonic	DIP2A	synonymous SNV	p.P187P	0.005882	4.98E-05	.	.
21	47918794	C	rs376953348	intronic	DIP2A	.	.	0.005882	0.0001	.	.
21	47954545	T	rs374691753	exonic	DIP2A	synonymous SNV	p.H486H	0.005882	4.20E-05	.	.
21	47957332	A	rs2070432	intronic	DIP2A	.	.	0.3647	0.2989	.	.
21	47961634	C	rs2070434	intronic	DIP2A	.	.	0.2882	0.2975	.	.
21	47961711	A	rs2070435	exonic	DIP2A	synonymous SNV	p.L650L	0.3882	0.34	.	.
21	47965205	T	rs150180949	intronic	DIP2A	.	.	0.01176	0.0095	.	.
21	47965766	G	rs144135274	intronic	DIP2A	.	.	0.005882	0.001	.	.
21	47965942	T	rs193216059	UTR3	DIP2A	.	.	0.005882	0.0015	.	.
21	47966791	A	rs73152864	intronic	DIP2A	.	.	0.02353	0.0112	.	.
21	47966810	C	rs2839318	intronic	DIP2A	.	.	0.2059	0.189	.	.
21	47966977	T	rs16979358	UTR3	DIP2A	.	.	0.1059	0.1359	.	.
21	47969653	T	rs57388012	intronic	DIP2A	.	.	0.005882	0.0116	.	.
21	47969793	T	rs17302700	exonic	DIP2A	synonymous SNV	p.L874L	0.06471	0.0629	.	.
21	47970581	A	rs2255397	exonic	DIP2A	synonymous SNV	p.T917T	0.2176	0.2238	.	.
21	47970696	G	rs2839319	intronic	DIP2A	.	.	0.01176	0.0594	.	.
21	47971539	A	rs2255526	intronic	DIP2A	.	.	0.7353	0.7247	.	.
21	47971839	A	rs200257829	intronic	DIP2A	.	.	0.005882	0.0008	.	.
21	47974055	G	rs16979371	intronic	DIP2A	.	.	0.005882	0.0167	.	.
21	47974456	A	rs540894130	intronic	DIP2A	.	.	0.005882	5.82E-05	.	.
21	47974582	G	rs1107065	exonic	DIP2A	synonymous SNV	p.T1079T	0.3941	0.4442	.	.
21	47974855	T	rs780892394	intronic	DIP2A	.	.	0.005882	5.18E-05	.	.
21	47980570	G	rs2070429	intronic	DIP2A	.	.	0.7059	.	.	.
21	47980760	G	rs2839324	intronic	DIP2A	.	.	0.1529	.	.	.
21	47983806	G	rs201790767	exonic	DIP2A	synonymous SNV	p.T1371T	0.005882	0.0105	.	.
21	47985655	T	rs2248636	exonic	DIP2A	synonymous SNV	p.T1394T	0.07647	0.0931	.	.
21	47985694	T	rs16979409	exonic	DIP2A	synonymous SNV	p.A1407A	0.005882	0.0126	.	.
21	47986633	A	rs3819044	intronic	DIP2A	.	.	0.9	0.9056	.	.
21	47987412	A	rs146953731	exonic	DIP2A	synonymous SNV	p.V1527V	0.02353	0.0097	.	.
21	47987547	T	rs8127941	UTR3	DIP2A	.	.	0.1529	0.156	.	.
22	40574170	A	rs6001827	intronic	TNRC6B	.	.	0.6	0.6363	.	.
22	40641963	A	rs145191448	intronic	TNRC6B	.	.	0.01765	.	.	.
22	40641980	A	rs3752513	intronic	TNRC6B	.	.	0.4	0.3396	.	.
22	40662751	A	rs373711092	exonic	TNRC6B	synonymous SNV	p.S839S	0.005882	3.34E-05	.	.
22	40666241	T	rs200825621	exonic	TNRC6B	synonymous SNV	p.G974G	0.005882	0.0015	.	.
22	40673999	T	rs2413621	intronic	TNRC6B	.	.	0.7118	0.7184	.	.
22	40681788	T	rs6001862	intronic	TNRC6B	.	.	0.3176	0.3727	.	.
22	40696881	A	rs111897485	intronic	TNRC6B	.	.	0.005882	0.0089	.	.
22	40696884	T	rs376354581	intronic	TNRC6B	.	.	0.005882	0.0011	.	.
22	40697377	G	rs5995843	intronic	TNRC6B	.	.	0.3176	0.3916	.	.
22	40708679	C	rs2072858	intronic	TNRC6B	.	.	0.3235	0.3807	.	.
22	51113502	A	rs368658976	exonic	SHANK3	synonymous SNV	p.A30A	0.005882	0.0003	.	.
22	51113661	T	rs530240325	exonic	SHANK3	synonymous SNV	p.T83T	0.01176	0.005	.	.
22	51117137	A	rs9616914	intronic	SHANK3	.	.	0.5294	0.3574	.	.
22	51117364	C	rs373824758	intronic	SHANK3	.	.	0.005882	0.0001	.	.

22	51121760	T	rs9628236	intronic	SHANK3	.	.	0.005882	0.0032	.	.
22	51121773	T	rs201282170	exonic	SHANK3	synonymous SNV	p.S297S	0.01176	0.0034	.	.
22	51122946	T	rs116756427	intronic	SHANK3	.	.	0.005882	.	.	.
22	51123025	A	rs371238756	exonic	SHANK3	nonsynonymous SNV	p.A326T	0.005882	3.36E-05	.	.
22	51133518	A	rs13055562	intronic	SHANK3	.	.	0.6235	0.5885	.	.
22	51133524	T	rs76224556	intronic	SHANK3	.	.	0.005882	0.0205	.	.
22	51137094	A	rs9616942	intronic	SHANK3	.	.	0.08824	0.1397	.	.
22	51137249	C	rs1557620	intronic	SHANK3	.	.	0.8471	0.8039	.	.
22	51137253	G	rs11704325	intronic	SHANK3	.	.	0.005882	0.0034	.	.
22	51142381	A	rs12483981	intronic	SHANK3	.	.	0.2471	0.2503	.	.
22	51142692	A	rs74975830	intronic	SHANK3	.	.	0.005882	0.0108	.	.
22	51143285	C	.	exonic	SHANK3	nonsynonymous SNV	p.I617L	0.005882	.	.	.
22	51143309	C	rs8141844	intronic	SHANK3	.	.	0.03529	0.0514	.	.
22	51143351	T	rs147115189	intronic	SHANK3	.	.	0.005882	0.0032	.	.
22	51144513	G	rs61731160	exonic	SHANK3	synonymous SNV	p.P667P	0.01765	0.0096	.	.
22	51149992	T	rs116503692	intronic	SHANK3	.	.	0.005882	.	.	.
22	51153371	A	rs61729471	exonic	SHANK3	nonsynonymous SNV	p.A707T	0.03529	0.0645	.	.
22	51153509	A	rs1483155568	intronic	SHANK3	.	.	0.005882	0.0024	.	.
22	51154049	C	rs763213714	intronic	SHANK3	.	.	0.005882	5.10E-05	.	.
22	51154141	A	rs117066889	exonic	SHANK3	synonymous SNV	p.P756P	0.005882	0.0062	.	.
22	51159624	T	rs200077311	exonic	SHANK3	synonymous SNV	p.S1107S	0.01176	0.0046	.	.
22	51159798	A	rs145196448	exonic	SHANK3	synonymous SNV	p.K1165K	0.005882	0.0073	.	.
22	51160140	T	rs201793890	exonic	SHANK3	synonymous SNV	p.S1279S	0.005882	0.0026	.	.
22	51169491	T	rs557669600	exonic	SHANK3	synonymous SNV	p.P1635P	0.01176	0.0107	.	.
X	31152354	T	rs745473007	intronic	DMD	.	.	0.01176	1.81E-05	.	.
X	31165350	C	rs72466537	intronic	DMD	.	.	0.005882	0.0073	.	.
X	31165400	A	rs1800281	exonic	DMD	synonymous SNV	p.L516L	0.01765	0.0047	.	.
X	31191589	T	rs2404496	intronic	DMD	.	.	0.8941	.	.	.
X	31196942	A	rs41303187	intronic	DMD	.	.	0.005882	0.0021	.	.
X	31224684	G	rs2293668	intronic	DMD	.	.	0.8765	0.8836	.	.
X	31241107	C	.	intronic	DMD	.	.	0.01176	.	.	.
X	31497197	G	rs72466570	exonic	DMD	synonymous SNV	p.T128T	0.005882	0.0015	.	.
X	31676096	A	rs2270672	intronic	DMD	.	.	0.3706	0.3398	.	.
X	31697636	G	rs1801188	exonic	DMD	synonymous SNV	p.N116N	0.1529	0.1713	.	.
X	31792260	C	rs780283825	exonic	DMD	synonymous SNV	p.E1112E	0.005882	3.44E-05	.	.
X	31792345	A	rs72466586	intronic	DMD	.	.	0.01765	0.012	.	.
X	31986430	A	rs3761604	intronic	DMD	.	.	0.2471	0.3293	.	.
X	31986669	T	rs67729860	intronic	DMD	.	.	0.005882	0.0011	.	.
X	32305619	T	rs3788896	intronic	DMD	.	.	0.03529	0.1184	.	.
X	32364030	T	rs72468623	intronic	DMD	.	.	0.01176	0.0174	.	.
X	32366476	C	rs372608114	intronic	DMD	.	.	0.01176	6.93E-05	.	.
X	32383095	A	rs372345571	intronic	DMD	.	.	0.01176	0.0003	.	.
X	32383284	A	rs61733574	exonic	DMD	synonymous SNV	p.V285V	0.005882	0.0076	.	.
X	32404616	T	rs72468639	intronic	DMD	.	.	0.005882	0.0227	.	.
X	32408149	T	rs72468644	intronic	DMD	.	.	0.005882	0.002	.	.
X	32408311	C	rs41303181	intronic	DMD	.	.	0.1176	0.0623	.	.
X	32486756	T	rs1800268	exonic	DMD	synonymous SNV	p.S999S	0.01765	0.0127	.	.
X	32503227	C	rs1028360	intronic	DMD	.	.	0.005882	0.0091	.	.
X	32563263	G	rs228373	intronic	DMD	.	.	0.2412	0.3203	.	.
X	32563488	C	rs115571	intronic	DMD	.	.	0.6118	0.7207	.	.
X	32583942	A	rs1800267	exonic	DMD	synonymous SNV	p.L615L	0.005882	0.0114	.	.
X	32591811	G	rs5927082	intronic	DMD	.	.	0.1647	0.1004	.	.
X	32591931	C	rs5927083	exonic	DMD	synonymous SNV	p.R537R	0.1647	0.1165	.	.
X	32659676	-	rs772637416	intronic	DMD	.	.	0.005882	3.11E-05	.	.
X	32662223	C	rs41303189	intronic	DMD	.	.	0.005882	0.0086	.	.
X	32715937	-	rs72470512	intronic	DMD	.	.	0.005882	0.0212	.	.
X	32716110	T	rs1800265	exonic	DMD	synonymous SNV	p.T271T	0.01765	0.0866	.	.
X	32716132	T	rs72470514	intronic	DMD	.	.	0.04118	0.0214	.	.
X	32716133	C	rs72470515	intronic	DMD	.	.	0.04118	0.0212	.	.
X	32841370	G	rs72470529	intronic	DMD	.	.	0.005882	0.003	.	.
X	32867810	A	rs72470531	intronic	DMD	.	.	0.01176	0.0069	.	.
X	32867945	A	rs3834997	intronic	DMD	.	.	0.1588	0.0963	.	.
X	70368097	T	rs368578220	intronic	NLGN3	.	.	0.01176	0.0008	.	.
X	70368772	A	rs763607264	intronic	NLGN3	.	.	0.005882	0.0002	.	.
X	70375023	A	rs185402974	intronic	NLGN3	.	.	0.03529	0.011	.	.
X	70375080	C	rs144247281	exonic	NLGN3	synonymous SNV	p.G158G	0.005882	0.0009	.	.
X	70386825	G	rs2233442	intronic	NLGN3	.	.	0.005882	0.013	.	.
X	70389650	T	rs5981083	exonic	NLGN3	synonymous SNV	p.A710A	0.005882	0.0049	.	.
X	147003404	G	rs80358323	intronic	FMR1	.	.	0.02353	0.0063	.	.
X	147003545	G	rs141796490	intronic	FMR1	.	.	0.005882	0.0037	.	.
X	147010320	A	rs25707	exonic	FMR1	synonymous SNV	p.R138R	0.1059	0.0868	.	.
X	147014145	T	rs150724379	intronic	FMR1	.	.	0.04706	0.0301	.	.
X	147018146	T	rs25714	intronic	FMR1	.	.	0.1529	0.1639	.	.
X	153295874	T	.	exonic	MECP2	nonsynonymous SNV	p.P481T	0.005882	.	D	22.8
14	21863188	T	rs755975357	exonic	CHD8	nonsynonymous SNV	p.R1758H	0.005882	.	D	34
4	114279628	C	rs36210417	exonic	ANK2	nonsynonymous SNV	p.I3285T	0.005882	0.0082	D	25
10	114925406	A	rs77673441	exonic	TCF7L2	nonsynonymous SNV	p.P472H	0.005882	0.0025	D	33
3	9517375	T	rs201582360	exonic	SETD5	nonsynonymous SNV	p.S1310L	0.005882	0.0008	D	32
7	104753553	T	rs145540034	exonic	KMT2E	nonsynonymous SNV	p.P1784S	0.005882	0.0019	D	24.8

14	21897467	A	rs192989929	exonic	CHD8	nonsynonymous SNV	p.L291F	0.005882	0.0009	D	26.3
1	155307458	T	rs765245236	exonic	ASH1L	nonsynonymous SNV	p.R2963Q	0.005882	1.65E-05	D	35
7	100491451	C	rs17885778	exonic	ACHE	nonsynonymous SNV	p.P135A	0.005882	0.0011	D	24.1
3	9517561	T	rs62246321	exonic	SETD5	nonsynonymous SNV	p.T1372I	0.005882	0.0011	D	23.4
4	114284542	A	rs374991526	exonic	ANK2	nonsynonymous SNV	p.R1517Q	0.005882	2.48E-05	D	34
6	43112281	T	rs754421182	exonic	PTK7	nonsynonymous SNV	p.R652C	0.005882	1.65E-05	D	34
X	32717364	C	rs145668843	exonic	DMD	nonsynonymous SNV	p.I224M	0.005882	0.0002	D	17.67
X	153296697	A	rs61749711	exonic	MECP2	synonymous SNV	p.S206S	0.005882	0.0022	D	4.854
2	25468174	C	rs149738328	exonic	DNMT3A	nonsynonymous SNV	p.N312S	0.005882	0.0003	D	17.25
7	151868413	A	rs149118569	exonic	KMT2C	nonsynonymous SNV	p.G3130V	0.005882	0.0002	D	25.8
7	151902197	G	rs138119145	exonic	KMT2C	nonsynonymous SNV	p.D1319H	0.005882	0.0085	D	29.7
4	114279674	A	rs34270799	exonic	ANK2	nonsynonymous SNV	p.S3300R	0.01765	0.0188	D	18.63
7	104746115	G	rs145944822	exonic	KMT2E	nonsynonymous SNV	p.E809G	0.005882	2.50E-05	D	23.9
5	170336717	A	rs144195437	exonic	RANBP17	nonsynonymous SNV	p.R181H	0.005882	0.0009	D	33
7	104747899	T	rs117986340	exonic	KMT2E	nonsynonymous SNV	p.G999C	0.08235	0.0377	D	27
17	3591933	T	rs151100959	exonic	P2RX5	nonsynonymous SNV	p.G268R	0.005882	0.007	D	34
3	9512345	G	rs138685269	exonic	SETD5	nonsynonymous SNV	p.A976G	0.005882	0.0008	D	23
16	2120559	A	rs45517203	exonic	TSC2	nonsynonymous SNV	p.A607T	0.005882	0.0007	D	25.1
7	100282480	A	.	exonic	GIGYF1	nonsynonymous SNV	p.S443F	0.005882	.	D	20.1
1	155408644	A	rs138474502	exonic	ASH1L	nonsynonymous SNV	p.P1768S	0.005882	0.0007	D	18.14
4	85594116	A	rs149067356	exonic	WDFY3	nonsynonymous SNV	p.R3496C	0.005882	0.0004	D	34
4	114277327	G	.	exonic	ANK2	nonsynonymous SNV	p.E2518G	0.005882	.	D	29.6
10	114925675	C	rs147841431	exonic	TCF7L2	nonsynonymous SNV	p.S562P	0.005882	0.0037	T	2.941
7	151860230	C	rs142835638	exonic	KMT2C	nonsynonymous SNV	p.Q3478E	0.005882	0.0045	D	22.4
16	2129638	C	.	exonic	TSC2	nonsynonymous SNV	p.R1078P	0.005882	.	D	24.8
11	70336479	T	rs117843717	exonic	SHANK2	nonsynonymous SNV	p.R230H	0.005882	0.0048	T	33
4	85717696	G	rs151088392	exonic	WDFY3	nonsynonymous SNV	p.D1049H	0.005882	0.0003	T	28.7
4	85598475	T	rs369567418	exonic	WDFY3	nonsynonymous SNV	p.R3445H	0.005882	1.66E-05	T	34
1	155313143	T	.	exonic	ASH1L	nonsynonymous SNV	p.G2752E	0.005882	.	T	29.5
6	43111342	C	rs9472017	exonic	PTK7	nonsynonymous SNV	p.E615D	0.01765	0.0126	T	26.5
7	117351826	A	rs150547726	exonic	CTTNBP2	nonsynonymous SNV	p.P1586L	0.01176	0.0055	T	28.2
16	2103392	T	rs137853994	exonic	TSC2	nonsynonymous SNV	p.E92V	0.005882	0.0019	T	25.9
16	2110795	A	rs1800725	exonic	TSC2	nonsynonymous SNV	p.R367Q	0.01765	0.0136	T	23.8
16	2112989	A	rs137854154	exonic	TSC2	nonsynonymous SNV	p.A460T	0.005882	0.0024	D	18.97
19	39226155	C	.	exonic	CAPN12	nonsynonymous SNV	p.L538R	0.005882	.	D	25.7
15	93567716	C	rs201950393	exonic	CHD2	nonsynonymous SNV	p.Q1756H	0.005882	0.0002	D	22.8
7	100285473	T	rs139996819	exonic	GIGYF1	nonsynonymous SNV	p.A67T	0.005882	0.0012	T	23.7
X	32613890	G	.	exonic	DMD	nonsynonymous SNV	p.L521S	0.01176	.	T	28.1
4	85623617	A	rs749089254	exonic	WDFY3	nonsynonymous SNV	p.R2829C	0.005882	8.38E-06	T	35
6	43109528	T	.	exonic	PTK7	nonsynonymous SNV	p.R451C	0.005882	.	T	34
4	85675001	A	rs202056534	exonic	WDFY3	nonsynonymous SNV	p.W1863L	0.005882	0.0001	T	28.1
16	2110710	A	rs559727962	exonic	TSC2	nonsynonymous SNV	p.V339I	0.005882	.	T	25
11	119216294	C	rs140629667	exonic	MFRP	nonsynonymous SNV	p.N159K	0.005882	0.0001	T	27.6
22	40662481	C	rs200598254	exonic	TNRC6B	nonsynonymous SNV	p.W749C	0.005882	0.0001	T	23.3
21	47929238	T	rs200877060	exonic	DIP2A	nonsynonymous SNV	p.R242C	0.005882	0.0004	T	34
4	114262881	A	rs34065266	exonic	ANK2	nonsynonymous SNV	p.V1311I	0.005882	4.12E-05	T	26.4
4	114177036	T	rs186264035	exonic	ANK2	nonsynonymous SNV	p.R379L	0.005882	8.24E-05	T	35
2	32289189	A	rs372005558	exonic	SPAST	nonsynonymous SNV	p.P97T	0.01176	0.0006	T	10.16
2	149241329	T	rs188449443	intronic	MBD5	.	.	0.01176	0.0028	T	12.13
11	4566916	T	rs201510674	exonic	OR52M1	nonsynonymous SNV	p.R166C	0.005882	0.0007	T	24.9
1	155448415	G	rs13373934	exonic	ASH1L	nonsynonymous SNV	p.S1416P	0.005882	0.003	T	15.18
2	166152389	A	rs17183814	exonic	SCN2A	nonsynonymous SNV	p.R19K	0.08235	0.082	T	10.94
2	50765412	T	rs56086732	exonic	NRXN1	nonsynonymous SNV	p.L708I	0.01765	0.0039	T	23.6
1	155490964	G	rs111605066	exonic	ASH1L	nonsynonymous SNV	p.I116T	0.005882	0.0034	T	13.14
7	151878670	A	rs140719911	exonic	KMT2C	nonsynonymous SNV	p.D2092V	0.005882	0.0007	T	24.9
7	151860470	A	rs746034306	exonic	KMT2C	nonsynonymous SNV	p.R3398W	0.005882	2.47E-05	T	34
19	39234589	T	rs145436226	exonic	CAPN12	nonsynonymous SNV	p.V73M	0.005882	0.0001	T	25.7
21	47954567	T	rs201190474	exonic	DIP2A	nonsynonymous SNV	p.R494W	0.005882	0.0006	T	32
X	147030322	T	rs45540244	exonic	FMR1	nonsynonymous SNV	p.T508I	0.005882	0.0009	T	11.16
7	151860866	A	.	exonic	KMT2C	nonsynonymous SNV	p.R3266W	0.005882	.	T	33
6	43100425	T	rs34021075	exonic	PTK7	nonsynonymous SNV	p.T418S	0.02941	0.0144	T	19.85
8	61654727	C	rs375325395	exonic	CHD7	nonsynonymous SNV	p.A246P	0.005882	8.29E-06	T	23
X	153296689	A	rs61749714	exonic	MECP2	nonsynonymous SNV	p.T209M	0.005882	0.0005	T	10.51
7	117400548	C	rs144920028	exonic	CTTNBP2	nonsynonymous SNV	p.N1038S	0.005882	0.0006	T	0.42
15	93557954	C	rs56227200	exonic	CHD2	nonsynonymous SNV	p.G1574A	0.01765	0.0225	T	6.335
X	32509543	C	rs769988372	exonic	DMD	nonsynonymous SNV	p.W817G	0.005882	2.34E-05	T	23.7
8	61778448	T	rs184814820	exonic	CHD7	nonsynonymous SNV	p.L935F	0.005882	0.0064	T	23.1
3	9517369	T	rs11542009	exonic	SETD5	nonsynonymous SNV	p.T1308I	0.08824	0.076	T	23.3
4	85678252	C	rs150181993	exonic	WDFY3	nonsynonymous SNV	p.C1751G	0.005882	0.0012	T	22.5
2	50765589	C	rs200074974	exonic	NRXN1	nonsynonymous SNV	p.I649V	0.005882	0.0006	T	15.69
9	96437286	A	rs139530562	exonic	PHF2	nonsynonymous SNV	p.A902T	0.005882	8.29E-05	T	22.4
21	47952128	C	rs192323584	exonic	DIP2A	nonsynonymous SNV	p.I385T	0.005882	0.0023	T	27.2
15	93486081	A	.	exonic	CHD2	nonsynonymous SNV	p.A279T	0.005882	.	T	23.5
7	104752798	A	rs74959149	exonic	KMT2E	nonsynonymous SNV	p.S1532N	0.005882	0.018	T	6.931
6	43109789	A	rs373575913	exonic	PTK7	nonsynonymous SNV	p.R500H	0.005882	7.48E-05	T	26.1
8	61655556	T	rs142962579	exonic	CHD7	nonsynonymous SNV	p.G522V	0.01176	0.0023	T	23.3
11	70331641	A	rs141276059	exonic	SHANK2	nonsynonymous SNV	p.P998L	0.005882	0.0009	T	27.9
7	117375374	C	rs62617115	exonic	CTTNBP2	nonsynonymous SNV	p.L1213V	0.02941	0.0277	T	22.7
6	43112267	T	rs34764696	exonic	PTK7	nonsynonymous SNV	p.A647V	0.05294	0.0478	T	15.12

6	99956560	C	rs7744845	exonic	USP45	nonsynonymous SNV	p.K67E	0.2647	0.3277	T	26.2
X	32472822	C	rs760481477	exonic	DMD	nonsynonymous SNV	p.Y1179C	0.005882	1.15E-05	T	26
9	135801087	T	rs118203357	exonic	TSC1	nonsynonymous SNV	p.A84T	0.005882	0.0007	T	14.12
5	170336704	A	rs80184931	exonic	RANBP17	nonsynonymous SNV	p.A177T	0.005882	0.0054	T	25.1
7	104749426	T	.	exonic	KMT2E	nonsynonymous SNV	p.A1169V	0.005882	.	T	26.9
5	170668092	A	.	exonic	RANBP17	nonsynonymous SNV	p.N861K	0.005882	.	T	16.04
15	93545488	A	rs61756301	exonic	CHD2	nonsynonymous SNV	p.S1407T	0.005882	0.0153	T	16.17
6	43128519	A	rs34865794	exonic	PTK7	nonsynonymous SNV	p.R908Q	0.01765	0.0136	T	23.3
2	149221327	A	rs34995577	exonic	MBD5	nonsynonymous SNV	p.G79E	0.005882	0.0008	T	25.4
1	202710733	G	rs112284833	exonic	KDM5B	nonsynonymous SNV	p.E903Q	0.02353	0.0235	T	25.3
20	49510344	C	.	exonic	ADNP	nonsynonymous SNV	p.M303V	0.005882	.	T	0.454
5	14508347	T	rs55900671	exonic	TRIO	nonsynonymous SNV	p.K3037M	0.005882	0.0041	T	21.8
4	114274556	G	rs775386505	exonic	ANK2	nonsynonymous SNV	p.I1594M	0.005882	0.0002	T	12.62
X	32361370	T	rs142441725	exonic	DMD	nonsynonymous SNV	p.E533K	0.005882	0.0003	T	23
21	38865433	G	rs145857775	exonic	DYRK1A	nonsynonymous SNV	p.T356A	0.005882	0.0015	T	24
11	70338541	G	rs55968949	exonic	SHANK2	nonsynonymous SNV	p.K192Q	0.005882	0.001	T	16.7
19	51857658	T	rs140608276	exonic	ETFB	nonsynonymous SNV	p.V79I	0.005882	0.0051	T	17.65
1	202700068	T	rs34216958	exonic	KDM5B	nonsynonymous SNV	p.R1382Q	0.01176	0.0037	T	21
11	119213319	A	rs145881139	exonic	MFRP	nonsynonymous SNV	p.L458F	0.01765	0.0085	T	23.5
17	7752884	A	rs202119281	exonic	KDM6B	nonsynonymous SNV	p.R1093H	0.005882	4.37E-05	T	24.5
11	70333498	C	rs62622853	exonic	SHANK2	nonsynonymous SNV	p.Y379C	0.01176	0.0178	T	22.2
21	47924334	A	rs201002582	exonic	DIP2A	nonsynonymous SNV	p.R239H	0.005882	8.49E-05	T	25.7
19	51857774	G	rs143144671	exonic	ETFB	nonsynonymous SNV	p.E40A	0.01765	0.0023	T	17.57
19	51850290	A	rs1130426	exonic	ETFB	nonsynonymous SNV	p.T245M	0.5706	0.498	T	32
17	7749954	G	rs60738318	exonic	KDM6B	nonsynonymous SNV	p.P203A	0.02353	0.0476	T	22.1
17	7751050	T	rs62059713	exonic	KDM6B	nonsynonymous SNV	p.P482S	0.1353	0.1868	T	17.08
6	157099799	A	rs375160616	exonic	ARID1B	nonsynonymous SNV	p.G246S	0.005882	0.0042	T	23.4
7	91729127	G	rs34327395	exonic	AKAP9	nonsynonymous SNV	p.M3614V	0.02353	0.0072	T	12.02
X	32380996	T	rs1801187	exonic	DMD	nonsynonymous SNV	p.R404H	0.5118	0.5142	T	32
9	135781205	C	rs118203576	exonic	TSC1	nonsynonymous SNV	p.K536R	0.01176	0.0188	T	14.22
6	157527482	C	rs149518409	exonic	ARID1B	nonsynonymous SNV	p.E1723A	0.005882	0.0022	T	6.985
5	14488142	A	rs750105964	exonic	TRIO	nonsynonymous SNV	p.G2469S	0.005882	0.0005	T	19.92
17	7749972	T	rs148641957	exonic	KDM6B	nonsynonymous SNV	p.V209L	0.005882	0.0041	T	11.91
18	44560739	G	.	exonic	TCEB3B	nonsynonymous SNV	p.E299D	0.005882	.	T	0.01
6	99930639	A	rs769168377	exonic	USP45	nonsynonymous SNV	p.L279F	0.005882	8.29E-06	T	31
7	91712698	G	rs6960867	exonic	AKAP9	nonsynonymous SNV	p.N2792S	0.3647	0.3593	T	1.912
11	70644619	T	rs146580493	exonic	SHANK2	unknown	.	0.01765	0.0049	T	34
18	44559933	A	rs139477376	exonic	TCEB3B	nonsynonymous SNV	p.Y568F	0.005882	0.0006	T	21.7
6	157507504	T	rs113430057	intronic	ARID1B	.	.	0.04118	0.039	T	6.201
6	33411673	C	rs191549504	exonic	SYNGAP1	nonsynonymous SNV	p.I1115T	0.01176	0.0101	T	5.59
4	114278128	T	rs145895389	exonic	ANK2	nonsynonymous SNV	p.S2785L	0.005882	0.0019	T	3.781
7	91708898	G	rs35759833	exonic	AKAP9	nonsynonymous SNV	p.K2484R	0.1059	0.109	T	0.002
7	100285476	T	rs77794375	exonic	GIGYF1	nonsynonymous SNV	p.A66T	0.07647	0.1005	T	21.6
3	20161096	G	rs17006625	exonic	KAT2B	nonsynonymous SNV	p.N386S	0.02353	0.032	T	0.001
17	7752244	G	rs373157695	exonic	KDM6B	nonsynonymous SNV	p.R880G	0.005882	0.0005	T	21.7
7	100488658	C	rs1799806	exonic	ACHE	nonsynonymous SNV	p.P592R	0.4882	0.473	T	18.14
18	44560875	A	rs2010834	exonic	TCEB3B	nonsynonymous SNV	p.C254F	0.5353	0.5058	T	0.003
X	32486806	G	rs72468667	exonic	DMD	nonsynonymous SNV	p.E983Q	0.005882	0.0015	T	22.3
20	25456698	T	rs35666277	exonic	NINL	nonsynonymous SNV	p.D1077N	0.04118	0.0447	T	11.54
2	149241063	A	rs72861124	intronic	MBD5	.	.	0.01176	0.0042	T	15.05
22	40662984	C	rs143708410	exonic	TNRC6B	nonsynonymous SNV	p.G917A	0.005882	0.0041	T	13.74
X	32509625	C	rs72468681	exonic	DMD	nonsynonymous SNV	p.N789K	0.01765	0.0077	T	21.8
4	114278277	T	rs3733617	exonic	ANK2	nonsynonymous SNV	p.P2835S	0.04706	0.09	T	6.374
17	7750936	G	rs73233606	exonic	KDM6B	nonsynonymous SNV	p.S444G	0.02353	0.0419	T	1.341
17	7750903	C	rs138395797	exonic	KDM6B	nonsynonymous SNV	p.S433P	0.005882	0.0162	T	13.09
X	31496350	T	rs1800280	exonic	DMD	nonsynonymous SNV	p.R208Q	0.9471	0.9045	T	17.1
6	157150496	G	rs17318151	exonic	ARID1B	nonsynonymous SNV	p.I560V	0.01765	0.0129	T	0.001
7	151877127	T	rs13231116	exonic	KMT2C	nonsynonymous SNV	p.P2412T	0.04118	0.0191	T	3.632
19	9452698	A	rs74575837	exonic	ZNF559	nonsynonymous SNV	p.E255K	0.01176	0.0188	T	18.6
11	70805674	C	.	exonic	SHANK2	unknown	.	0.005882	.	T	0.506
20	25456793	T	rs202203038	exonic	NINL	nonsynonymous SNV	p.G1045E	0.005882	0.0002	T	0.001
7	91603056	T	rs142401936	exonic	AKAP9	nonsynonymous SNV	p.S27L	0.005882	0.0017	T	17.66
18	44560429	A	rs892586	exonic	TCEB3B	nonsynonymous SNV	p.A403S	0.5941	0.571	T	0.002
X	31986607	A	rs1800273	exonic	DMD	nonsynonymous SNV	p.R814W	0.04118	0.026	T	25.3
17	3594291	T	rs372259448	exonic	P2RX5	nonsynonymous SNV	p.L107M	0.005882	2.64E-05	T	22.7
19	39221780	A	rs563664489	exonic	CAPN12	nonsynonymous SNV	p.V681L	0.005882	9.14E-06	T	23.5
4	85715702	C	rs148407700	exonic	WDFY3	nonsynonymous SNV	p.S1153A	0.005882	0.0009	T	26.8
16	2138508	C	rs9209	exonic	TSC2	nonsynonymous SNV	p.S1707T	0.005882	0.003	T	0.307
6	157507698	C	rs113232635	intronic	ARID1B	.	.	0.03529	0.0266	T	1.267
7	91630620	T	rs6964587	exonic	AKAP9	nonsynonymous SNV	p.M463I	0.3765	0.3829	T	0.443
18	44561100	G	rs2571028	exonic	TCEB3B	nonsynonymous SNV	p.R179P	0.5647	0.5614	T	0.249
X	32632470	A	rs375337020	exonic	DMD	nonsynonymous SNV	p.P470S	0.01176	1.16E-05	T	20.8
X	32509441	A	rs187926894	exonic	DMD	nonsynonymous SNV	p.T851S	0.005882	0.0001	T	0.01
20	49509113	A	rs375998490	exonic	ADNP	nonsynonymous SNV	p.A713V	0.005882	1.65E-05	T	12.75
17	29645473	T	rs140933050	exonic	EVI2A	nonsynonymous SNV	p.G187S	0.005882	0.0032	T	24.9
18	44560300	T	rs3744863	exonic	TCEB3B	nonsynonymous SNV	p.A446T	0.5353	0.506	T	19.87
7	91732083	A	rs141856443	exonic	AKAP9	nonsynonymous SNV	p.R3758H	0.005882	0.0016	T	21.6
X	32503194	C	rs228406	exonic	DMD	nonsynonymous SNV	p.D874G	0.6294	0.7211	T	8.366
5	14358417	G	rs773220678	exonic	TRIO	nonsynonymous SNV	p.N726S	0.005882	8.25E-06	T	16.28

7	151860113	A	.	exonic	KMT2C	nonsynonymous SNV	p.P3517S	0.005882	.	T	11.92
11	119213626	C	.	exonic	MFRP	nonsynonymous SNV	p.S404R	0.005882	.	T	16.61
9	96422544	A	rs376063250	exonic	PHF2	nonsynonymous SNV	p.R467Q	0.005882	2.19E-05	T	6.452
5	14507299	T	rs200337620	exonic	TRIO	nonsynonymous SNV	p.A2894V	0.005882	2.49E-05	T	23
11	70829901	T	rs73521173	exonic	SHANK2	unknown	.	0.01176	0.0208	T	18.56
6	157431662	A	rs147784000	exonic	ARID1B	nonsynonymous SNV	p.A767T	0.01765	0.0007	T	26.9
15	27018841	A	rs25409	exonic	GABRB3	nonsynonymous SNV	p.P11S	0.005882	0.0058	T	16.19
18	44560038	A	rs72921303	exonic	TCEB3B	nonsynonymous SNV	p.C533F	0.01176	0.0119	T	23.8
20	25484623	A	rs13044759	exonic	NINL	nonsynonymous SNV	p.R276W	0.04706	0.036	T	22.4
X	31496398	C	rs1800279	exonic	DMD	nonsynonymous SNV	p.H192R	0.02353	0.0263	T	1.633
11	119216504	T	rs3814762	exonic	MFRP	nonsynonymous SNV	p.V136M	0.3059	0.2664	T	0.239
21	41450656	T	rs200532632	exonic	DSCAM	nonsynonymous SNV	p.A1557T	0.005882	0.0002	T	15.92
7	151859683	A	rs74483926	exonic	KMT2C	nonsynonymous SNV	p.S3660L	0.005882	0.0477	T	8.146
5	65307924	G	rs61758158	exonic	ERBB2IP	nonsynonymous SNV	p.I119V	0.01176	0.0098	T	5.207
20	25472105	A	rs141376094	exonic	NINL	nonsynonymous SNV	p.A456V	0.005882	0.0007	T	1.037
7	151949068	G	rs3735156	exonic	KMT2C	nonsynonymous SNV	p.R526P	0.03529	0.0768	T	15.75
7	117358107	C	rs142089340	exonic	CTTNBP2	nonsynonymous SNV	p.T1571A	0.005882	0.0014	T	9.618
X	32429940	C	rs28715870	exonic	DMD	nonsynonymous SNV	p.F47V	0.005882	0.0098	T	9.041
1	151377922	C	rs199639268	exonic	POGZ	nonsynonymous SNV	p.S1102G	0.005882	2.47E-05	T	6.906
18	44560678	C	.	exonic	TCEB3B	nonsynonymous SNV	p.L320V	0.005882	.	T	9.5
3	11070958	A	rs112095333	exonic	SLC6A1	nonsynonymous SNV	p.L415I	0.005882	0.0034	T	20.1
17	29653037	G	rs147327414	exonic	NF1	nonsynonymous SNV	p.I1658V	0.005882	0.0037	T	7.41
20	25485659	C	rs34585177	exonic	NINL	nonsynonymous SNV	p.S191R	0.005882	0.0014	T	14.74
11	67926306	T	rs550206217	exonic	KMT5B	nonsynonymous SNV	p.A263T	0.005882	4.96E-05	T	1.894
7	91714951	T	rs149341527	exonic	AKAP9	nonsynonymous SNV	p.S2992L	0.005882	0.0002	T	25.7
14	21861835	C	rs148494847	exonic	CHD8	nonsynonymous SNV	p.D2040G	0.005882	0.0018	T	22.3
20	25442190	C	rs139278158	exonic	NINL	nonsynonymous SNV	p.L1222V	0.005882	0.0032	T	15.3
21	47975907	G	rs573139009	exonic	DIP2A	nonsynonymous SNV	p.K1130R	0.005882	0.0011	T	23.1
19	39229089	G	rs12983010	exonic	CAPN12	nonsynonymous SNV	p.C287R	0.07059	0.081	T	13.86
19	9453511	C	.	exonic	ZNF559	nonsynonymous SNV	p.E526Q	0.005882	.	T	7.53
X	31893307	G	rs1800275	splicing	DMD	.	.	0.2765	0.2781	T	15.62
7	151848538	C	rs139111507	exonic	KMT2C	nonsynonymous SNV	p.L4219V	0.005882	0.0031	T	0.002
17	7752523	C	rs61764072	exonic	KDM6B	nonsynonymous SNV	p.K973Q	0.02353	0.0147	T	24.7
9	96436037	A	rs41276200	exonic	PHF2	nonsynonymous SNV	p.S840N	0.01176	0.011	T	8.693
2	149228026	C	.	exonic	MBD5	nonsynonymous SNV	p.E838D	0.005882	.	T	10.8
17	3582954	T	rs61748727	exonic	P2RX5	nonsynonymous SNV	p.E373K	0.005882	0.0104	T	11.93
11	4567335	G	rs2657167	exonic	OR52M1	nonsynonymous SNV	p.S305R	0.4235	0.3922	T	0.002
5	65321311	T	rs3213837	exonic	ERBB2IP	nonsynonymous SNV	p.S274L	0.1412	0.1546	T	16.72
17	7751140	T	.	exonic	KDM6B	nonsynonymous SNV	p.P512S	0.005882	.	T	0.946
11	70666499	A	rs115457448	exonic	SHANK2	unknown	.	0.005882	0.0105	T	22.4
17	29645538	T	rs147909684	exonic	EVI2A	nonsynonymous SNV	p.S165Y	0.005882	0.0006	T	7.049
11	67925546	C	rs144458991	exonic	KMT5B	nonsynonymous SNV	p.N516S	0.005882	0.0028	T	0.002
7	91603115	T	rs35669569	exonic	AKAP9	nonsynonymous SNV	p.H47Y	0.005882	0.0081	T	0.009
6	99894086	G	rs41288947	exonic	USP45	nonsynonymous SNV	p.R521T	0.3824	0.276	T	8.621
7	91714911	T	rs1063242	exonic	AKAP9	nonsynonymous SNV	p.P2979S	0.9941	0.9962	T	1.518
19	39221513	T	.	exonic	CAPN12	nonsynonymous SNV	p.M713I	0.005882	.	T	21.1
4	114276906	A	rs141191319	exonic	ANK2	nonsynonymous SNV	p.E2378K	0.005882	0.0026	T	7.91
14	21896112	T	rs768690204	exonic	CHD8	nonsynonymous SNV	p.S506N	0.005882	8.28E-06	T	12.25
20	49508584	T	rs750568080	exonic	ADNP	nonsynonymous SNV	p.S889R	0.005882	9.89E-05	T	17.44
8	61654298	A	rs41272435	exonic	CHD7	nonsynonymous SNV	p.S103T	0.01176	0.0116	T	19.78
21	41725625	T	rs41395652	exonic	DSCAM	nonsynonymous SNV	p.R234H	0.01765	0.0303	T	22.9
18	44559844	A	rs61743415	exonic	TCEB3B	nonsynonymous SNV	p.P598S	0.01176	0.0119	T	23.3
X	32383146	T	rs16990264	exonic	DMD	nonsynonymous SNV	p.N331K	0.005882	0.0076	T	20.5
2	1926507	C	.	exonic	MYT1L	nonsynonymous SNV	p.N345S	0.005882	.	T	0.02
18	44559730	T	rs78014467	exonic	TCEB3B	nonsynonymous SNV	p.G636S	0.04706	0.0227	T	9.889
19	51857738	A	rs79338777	exonic	ETFB	nonsynonymous SNV	p.P52L	0.07647	0.0739	T	19.41
5	170725810	T	rs755426271	exonic	RANBP17	nonsynonymous SNV	p.A1072V	0.005882	8.25E-06	T	11.46
X	70389349	T	rs370863612	exonic	NLGN3	nonsynonymous SNV	p.P610L	0.005882	2.42E-05	T	17.23
7	91732041	C	rs148318643	exonic	AKAP9	nonsynonymous SNV	p.G3744A	0.005882	9.89E-05	T	24.6
7	91712609	C	rs144875383	exonic	AKAP9	nonsynonymous SNV	p.K2762N	0.005882	0.0013	T	0.211
11	70507842	T	rs149581714	exonic	SHANK2	nonsynonymous SNV	p.A11T	0.005882	9.06E-05	T	17.05
20	25434139	T	rs17857107	exonic	NINL	nonsynonymous SNV	p.R1366H	0.1059	0.0912	T	15.46
4	114276880	C	rs28377576	exonic	ANK2	nonsynonymous SNV	p.V2369A	0.1176	0.098	T	0.001
20	25472061	A	rs567610080	exonic	NINL	nonsynonymous SNV	p.A471S	0.005882	8.30E-06	T	6.103
18	44585955	A	rs7233515	exonic	KATNAL2	nonsynonymous SNV	p.S88N	0.4588	0.4236	T	0.014
22	40552119	A	rs9611280	exonic	TNRC6B	nonsynonymous SNV	p.V16M	0.1	0.111	T	23.3
17	7751531	T	rs201403136	exonic	KDM6B	nonsynonymous SNV	p.P642L	0.005882	0.004	T	23.1
11	70830059	A	rs199717803	exonic	SHANK2	unknown	.	0.005882	9.38E-05	T	19.54
20	25439036	A	rs41310175	exonic	NINL	nonsynonymous SNV	p.R1276C	0.02353	0.0323	T	12.99
11	67925354	C	rs758639921	exonic	KMT5B	nonsynonymous SNV	p.Q580R	0.005882	1.65E-05	T	11.15
17	7750010	C	rs79548905	exonic	KDM6B	nonsynonymous SNV	p.E221D	0.01176	0.0191	T	13.27
X	32662355	A	rs34155804	exonic	DMD	nonsynonymous SNV	p.T401S	0.005882	0.0039	T	8.415
6	33399778	T	rs9394145	intronic	SYNGAP1	.	.	0.2588	0.3066	T	12.41
7	100490797	T	rs1799805	exonic	ACHE	nonsynonymous SNV	p.H353N	0.05882	0.0423	T	12.3
11	119216555	C	rs4639950	exonic	MFRP	nonsynonymous SNV	p.I119V	0.005882	0.0049	T	0.003
19	9452879	A	rs16979670	exonic	ZNF559	nonsynonymous SNV	p.T315N	0.02353	0.0211	T	0.023
17	29623288	T	rs11080149	exonic	OMG	nonsynonymous SNV	p.G21D	0.1353	0.0927	T	19.64
2	183792911	G	.	exonic	NCKAP1	nonsynonymous SNV	p.Q1038H	0.005882	.	T	22.4
9	135786904	G	rs1073123	exonic	TSC1	nonsynonymous SNV	p.M271T	0.1647	0.129	T	6.707



21	47918662	G	rs7283507	exonic	DIP2A	nonsynonymous SNV	p.P191A	0.09412	0.0544	T	23.3
2	183848102	T	rs376081149	exonic	NCKAP1	nonsynonymous SNV	p.M338K	0.005882	.	T	22
6	157469914	C	.	exonic	ARID1B	nonsynonymous SNV	p.G890A	0.005882	.	T	11.44
4	85612894	C	rs17368018	exonic	WDFY3	nonsynonymous SNV	p.I3032V	0.005882	0.0072	T	3.614
19	39230834	A	rs750649524	exonic	CAPN12	nonsynonymous SNV	p.R196W	0.005882	1.18E-05	T	24.9
2	50574038	A	rs13413205	exonic	NRXN1	nonsynonymous SNV	p.G17V	0.05882	0.0999	T	0.914
21	41725630	C	rs2297270	exonic	DSCAM	nonsynonymous SNV	p.D232E	0.07647	0.1066	T	0.002
5	65350279	T	rs142496054	exonic	ERBB2IP	nonsynonymous SNV	p.H1045Y	0.01765	0.0071	T	7.784
4	85657463	C	rs768601762	exonic	WDFY3	nonsynonymous SNV	p.I2259V	0.005882	0.0001	T	0.994
4	140307334	G	rs201001193	exonic	NAA15	nonsynonymous SNV	p.I784V	0.005882	0.0001	T	10.9
10	28900828	A	rs143909998	exonic	WAC	nonsynonymous SNV	p.P369T	0.005882	0.0002	T	22.4
11	70666761	C	.	exonic	SHANK2	unknown	.	0.005882	.	T	7.874
11	119216142	A	rs150902999	exonic	MFRP	nonsynonymous SNV	p.G210V	0.01176	0.0042	T	26
18	44560123	T	rs72921305	exonic	TCEB3B	nonsynonymous SNV	p.G505R	0.01176	0.012	T	23.4
X	32503174	C	.	exonic	DMD	nonsynonymous SNV	p.R881G	0.005882	.	T	23.3
18	44560480	T	rs138896768	exonic	TCEB3B	nonsynonymous SNV	p.Y386N	0.005882	0.0016	T	13.09
19	9449888	G	rs77267061	exonic	ZNF559	nonsynonymous SNV	p.D82G	0.04118	0.0624	T	23.1
17	3599205	T	rs142863822	exonic	P2RX5	nonsynonymous SNV	p.L32Q	0.01176	0.0041	T	26.7
3	20164282	G	rs41285059	exonic	KAT2B	nonsynonymous SNV	p.M467V	0.005882	0.0003	T	20.6
19	39228244	C	rs73038948	exonic	CAPN12	nonsynonymous SNV	p.T334A	0.1941	0.1739	T	23.3
2	225362478	T	rs3738952	exonic	CUL3	nonsynonymous SNV	p.V501I	0.1	0.128	T	23
7	151949735	C	rs77652527	exonic	KMT2C	nonsynonymous SNV	p.I455M	0.05294	0.0234	T	10.78
4	114276781	T	rs61734477	exonic	ANK2	nonsynonymous SNV	p.A2336V	0.005882	0.0049	T	0.221
7	91641854	C	rs141039834	exonic	AKAP9	nonsynonymous SNV	p.C1144R	0.005882	0.0004	T	20.5
18	19418429	A	rs747233793	exonic	MIB1	nonsynonymous SNV	p.V645I	0.005882	1.65E-05	T	26.2
17	3592832	C	rs750820489	exonic	P2RX5	nonsynonymous SNV	p.S212C	0.005882	8.26E-06	T	22.9
4	114276408	A	rs61734478	exonic	ANK2	nonsynonymous SNV	p.G2212S	0.005882	0.0047	T	15.67
6	33399775	A	.	intronic	SYNGAP1	.	.	0.005882	.	T	14.55
6	99930627	G	rs17850034	splicing	USP45	.	.	0.01176	0.0084	T	24.6
7	91726527	C	rs61757663	exonic	AKAP9	nonsynonymous SNV	p.Q3418H	0.005882	0.0009	T	25.5
7	91667736	G	rs150379637	exonic	AKAP9	nonsynonymous SNV	p.I1448V	0.005882	0.0005	T	23.6
5	65349382	G	rs16894812	exonic	ERBB2IP	nonsynonymous SNV	p.K746E	0.005882	0.0134	T	21.1
21	47949017	A	rs16979312	exonic	DIP2A	nonsynonymous SNV	p.S329N	0.01176	0.0178	T	26.8
10	114925369	A	rs77961654	exonic	TCF7L2	nonsynonymous SNV	p.H401Q	0.04118	0.0503	T	24.9
5	65350481	T	rs3805466	exonic	ERBB2IP	nonsynonymous SNV	p.S1112L	0.04118	0.0844	T	24.5

Supplemental Table S3: Top 5 SNPs for each SRS subscore for variants in the Williams syndrome Critical Region

SNP	Alt allele	MAF	Transcript[a]	Gene	Consequence	Beta	95% Confidence interval	Raw p-value	FDR	SRS sub category
rs3812316	G	0.1	NM_032951	*MLXIPL*	p.Q241H	4.817	1.146-8.487	0.01206	0.4101	AWR
rs13235543	T	0.12	NM_032954	*MLXIPL*	p.P342P	3.399	0.05103-6.746	0.05016	0.5083	AWR
rs2074754	T	0.4	NM_032408	*BAZ1B*	p.S679S	2.24	-0.5935 – 4.54	0.06004	0.5083	AWR
rs61438591	C	0.2	.	*GTF2IRD1*	intronic	2.426	-0.3677 – 5.22	0.09284	0.5083	AWR
rs2071307	A	0.47	NM_001081752	*ELN*	p.G412S	1.963	-0.2993 – 4.225	0.09305	0.5083	AWR
rs2074754	T	0.4	NM_032408	*BAZ1B*	p.S679S	4.145	1.496-6.794	0.003006	0.1022	COG
rs61438591	C	0.2	.	*GTF2IRD1*	intronic	3.578	0.2897-6.867	0.03618	0.6151	COG
rs17851629	G	0.21	NM_016328	*GTF2IRD1*	E171E	3.129	-0.1129 – 6.37	0.06229	0.706	COG
rs61010704	G	0.23	.	*MLXIPL*	intronic	2.377	-0.872 – 5.626	0.1559	0.8179	COG
rs7795181	C	0.22	.	*VPS37D*	intronic	-2.082	-5.346 – 1.183	0.2153	0.8179	COG
rs2074754	T	0.4	NM_032408	*BAZ1B*	p.S679S	3.172	0.7265 – 5.617	0.01307	0.2675	COM
rs61438591	C	0.2	.	*GTF2IRD1*	intronic	3.732	0.7711 – 6.692	0.01573	0.2675	COM
rs17851629	G	0.21	NM_016328	*GTF2IRD1*	E171E	3.241	0.2964 – 6.186	0.0341	0.3865	COM
rs3812316	G	0.1	NM_032951	*MLXIPL*	p.Q241H	3.122	-0.995 – 7.238	0.1414	0.7693	COM
rs3135698	C	0.06	.	*RFC2*	intronic	-4.712	-11.52 – 2.093	0.1787	0.7693	COM
rs2074754	T	0.4	NM_032408	*BAZ1B*	p.S679S	2.411	0.301 – 4.521	0.02808	0.8893	MOT
rs61438591	C	0.2	.	*GTF2IRD1*	intronic	2.441	-.1236 – 5.007	0.06597	0.8893	MOT
rs17851629	G	0.21	NM_016328	*GTF2IRD1*	E171E	1.862	-0.6716 – 4.396	0.1538	0.8893	MOT
rs76029572	G	0.07	NM_012453	*TBL2*	p.E8Q	-2.966	-7.008 – 1.075	0.1543	0.8893	MOT
rs2240357	C	0.23	NM_016328	*GTF2IRD1*	p.Y404Y	1.745	-0.7424 – 4.232	0.1731	0.8893	MOT
rs2074754	T	0.4	NM_032408	*BAZ1B*	p.S679S	2.436	-0.2729 – 5.146	0.08205	0.7104	RRB
rs61438591	C	0.2	.	*GTF2IRD1*	intronic	2.762	-0.4886 – 6.012	0.09996	0.7104	RRB
rs2071307	A	0.47	NM_001081752	*ELN*	p.G412S	2.222	-0.4162 – 4.86	0.1029	0.7104	RRB
rs61010704	G	0.25	.	*MLXIPL*	intronic	2.476	-0.6671 – 5.619	0.1269	0.7104	RRB
rs3812316	G	0.1	NM_032951	*MLXIPL*	p.Q241H	3.245	-1.159 – 7.649	0.1528	0.7104	RRB

[a] "." Refers to information that is not applicable

Supplemental Table S4: Top 5 SNPs for each SRS subscore for variants in 71 genes associated with Autism spectrum disorder

SNP	Alt allele	MAF	Transcript[a]	Gene	Consequence	Beta	95% Confidence interval	Raw p-value	FDR	SRS sub category
rs45599933	A	0.06	.	*CAPN12*	intronic	11.72	4.596 – 18.85	0.001854	0.1735	AWR
rs12983010	G	0.07	NM_144691	*CAPN12*	p.C287R	10.47	4.028 – 16.91	0.002085	0.1735	AWR
rs3733615	G	0.16	NM_001148	*ANK2*	p.Q2370Q	6.806	2.486 – 11.13	0.0028	0.1735	AWR
rs33966911	T	0.11	NM_001148	*ANK2*	p.P1823P	7.609	2.764 – 12.45	0.002885	0.1735	AWR
rs28377576	C	0.11	NM_001148	*ANK2*	p.V2369A	7.578	2.734 – 12.45	0.002991	0.1735	AWR
rs3750354	T	0.39	.	*PHF2*	intronic	-6.727	-10.58 – -2.875	0.0009956	0.1897	COG
rs7036592	T	0.39	.	*PHF2*	intronic	-6.727	-10.58 – -2.875	0.0009956	0.1897	COG
rs10992813	A	0.37	.	*PHF2*	intronic	-6.476	-10.43 – -2.527	0.001922	0.2441	COG
rs3763605	G	0.63	.	*PHF2*	intronic	5.41	1.399 – 9.421	0.009964	0.5607	COG
rs3750358	C	0.63	.	*PHF2*	intronic	5.311	1.276 – 9.345	0.01178	0.5607	COG
rs112318565	G	0.06	.	*ARID1B*	intronic	11.63	3.582 – 19.67	0.005892	0.8653	COM
rs12553775	A	0.11	.	*PHF2*	intronic	7.386	1.482 – 13.29	0.01647	0.8653	COM
rs140682	C	0.48	NM_000810	*GABRA5*	p.V202V	-4.218	-7.659 – -0.7778	0.01867	0.8653	COM
rs4351684	G	0.51	.	*ILF2*	intronic	-4.662	-8.546 – -0.7782	0.02119	0.8653	COM
rs1805482	A	0.35	NM_000834	*GRIN2B*	p.S555S	4.615	0.6839 – 8.546	0.0241	0.8653	COM
rs30612	C	0.84	NM_007118	*TRIO*	p.T1700T	6.392	2.773 – 10.01	0.0008803	0.3354	MOT
rs12983010	G	0.07	NM_144691	*CAPN12*	p.C287R	8.842	2.839 – 14.84	0.005044	0.6218	MOT
rs45599933	A	0.06	.	*CAPN12*	intronic	9.766	3.115 – 16.42	0.005181	0.6218	MOT
rs7005873	A	0.74	.	*CHD7*	intronic	-4.68	-7.972 – -1.388	0.006735	0.6218	MOT
rs27100	T	0.43	.	*TRIO*	intronic	-4.171	-7.182– -1.161	0.008161	0.6218	MOT
rs7005873	A	0.74	.	*CHD7*	intronic	-6.41	-10.54 – -2.279	0.00323	0.5602	RRB
rs1805482	A	0.35	NM_000834	*GRIN2B*	p.S555S	6.074	1.944 – 10.2	0.005112	0.5602	RRB
rs112318565	G	0.06	.	*ARID1B*	intronic	12.3	3.682 – 20.91	0.006502	0.5602	RRB
rs7844902	G	0.72	.	*CHD7*	intronic	-5.388	-9.483 – -1.294	0.01186	0.5602	RRB
rs5891777	TGGACT	0.74	.	*CHD7*	intronic	-5.144	-9.263 – -1.025	0.01665	0.5602	RRB

[a] “.” Refers to information that is not applicable

Supplemental Table S5: Top 5 SNPs for each SRS subscore for variants discovered across the whole exome

SNP	Alt allele	MAF	Transcript[a]	Gene	Consequence	Beta	95% Confidence interval	Raw p-value	FDR	SRS sub category
rs35430620	T	0.79	.	*PCTP*	intronic	9.39	5.767 – 13.01	2.57E-06	0.1711	AWR
rs3803300	C	0.84	NM_001137601	*ZBTB42*	UTR3	-11.08	-15.71 – -6.448	1.20E-05	0.3714	AWR
Var-6-31322340	A	0.07	.	*HLA-B*	intronic	-12.76	-18.21 – -7.318	1.67E-05	0.3714	AWR
rs1804020	A	0.27	NM_001014972	*ZFN638*	p.V1726M	-8.316	-12.03 – -4.604	3.55E-05	0.4874	AWR
rs2960061	C	0.85	.	*PCTP*	intronic	10.29	5.69 – 14.9	3.66E-05	0.4874	AWR
rs527221	C	0.11	NM_001288765	*DMPK*	p.L334V	16.4	10.69 – 22.12	2.94E-07	0.01959	COG
rs572634	C	0.11	.	*DMPK*	intronic	14.76	9.043 – 20.48	2.80E-06	0.09309	COG
rs2292288	G	0.43	unknown	*SYNM*	unknown	-8.746	-12.5 – -4.996	1.81E-05	0.4029	COG
rs2305914	T	0.08	.	*WBP2*	intronic	-15.78	-22.86 – -8.704	3.85E-05	0.6403	COG
rs1064512	C	0.08	NM_003038	*SLC1A4*	p.G37R	13.12	7.061 – 19.18	6.14E-05	0.8187	COG
rs2076404	A	0.69	.	*TGM6*	intronic	-8.695	-12.35 – -5.038	1.30E-05	0.4552	COM
rs2546028	C	0.55	NM_175872	*ZNF792*	UTR5	-6.561	-9.394 – -3.728	2.05E-05	0.4552	COM
rs2546029	G	0.55	NM_175872	*ZNF792*	UTR5	-6.561	-9.394 – -3.728	2.05E-05	0.4552	COM
rs491873	T	0.59	.	*TUBA3C*	intronic	-7.645	-11.03 – -4.258	3.16E-05	0.5256	COM
rs1811	G	0.46	NM_001099437	*ZNF30*	p.Q124R	6.464	3.404 – 9.524	8.81E-05	0.734	COM
rs2651080	C	0.31	NM_175872	*ZNF792*	p.T333T	8.116	5.113 – 11.12	1.09E-06	0.02169	MOT
rs1345658	A	0.46	NM_001099437	*ZNF30*	p.R380K	6.535	4.07 – 9.001	1.63E-06	0.02169	MOT
rs1811	G	0.46	NM_001099437	*ZNF30*	p.Q124R	6.535	4.07 – 9.001	1.63E-06	0.02169	MOT
rs2651079	T	0.46	NM_175872	*ZNF792*	p.R177Q	6.535	4.07 – 9.001	1.63E-06	0.02169	MOT
rs2651109	C	0.46	NM_001099437	*ZNF30*	p.S215S	6.535	4.07 – 9.001	1.63E-06	0.02169	MOT
rs2546028	C	0.55	NM_175872	*ZNF792*	UTR5	-6.773	-9.831 – -3.715	4.26E-05	0.5321	RRB
rs2546029	G	0.55	NM_175872	*ZNF792*	UTR5	-6.773	-9.831 – -3.715	4.26E-05	0.5321	RRB
rs2059404	A	0.58	.	*ARID2*	intronic	-8.358	-12.16 – -4.558	4.75E-05	0.5321	RRB
rs7315731	T	0.42	NM_004719	*SCAF11*	p.V627I	-8.358	-12.16 – -4.558	4.75E-05	0.5321	RRB
rs13044892	A	0.06	.	*ATP9A*	intronic	-16.79	-24.58 – -9.009	6.57E-05	0.5321	RRB

[a] “.” Refers to information that is not applicable

Chapter 3: The effects of *Gtf2ird1* and *Gtf2i* DNA binding on transcription and behavior supports the important function of the N-terminal end of *Gtf2ird1*.

Nathan Kopp, Katherine McCullough, Susan E. Maloney, Joseph Dougherty

3.1 Abstract

The two transcription factors Gtf2i and Gtf2ird1 have been thought to play a role in the craniofacial, cognitive, and behavioral phenotypic domains of WS. There exist many mouse models of each of these transcription factors that show behavioral phenotypes. Further, some phenotypes such as balance, anxiety, and social behavior, mouse models of both transcription factors show deficits in the same direction, however the affect of these genes on behavior have not been studied in combination. To examine how these genes could mediate behavioral consequences we described the genomic binding sites of these transcription factors in the developing brain. We then characterized two new mouse models generated using the CRISPR/Cas9 system to test how mutating both *Gtf2i* and *Gtf2ird1* can modify the transcriptional and behavioral phenotype observed in a single *Gtf2ird1* mutant. The *Gtf2ird1* mutant was shown to make a N-truncated protein that has decreased capacity to bind the promoter of *Gtf2ird1* but still can bind genome-wide. Despite little differences in DNA-binding and transcriptome-wide expression, the mutation still caused balance, marble burying, and activity phenotypes, supporting a functional role for the N-terminus of Gtf2ird1. Mutating both *Gtf2i* and *Gtf2ird1* did not modify the transcriptomic or behavioral phenotypes, suggesting that *Gtf2ird1* mutation largely drives the behavioral phenotypes observed.

3.2 Introduction

The Williams syndrome critical region (WSCR) contains 26 genes that are typically deleted in Williams syndrome (WS) (OMIM#194050). The genes in this region are of interest for their potential to contribute to the unique physical, cognitive, and behavioral phenotypes of WS, which include craniofacial dysmorphology, mild to severe intellectual disability, poor visual spatial cognition, balance and coordination problems, and a characteristic hypersocial personality

(2, 13, 15). Single gene knock out mouse models exist for many of the genes in the region, with differing degrees of face validity to the phenotypes of WS (92–96, 101). Two genes have been highlighted in the human and mouse literature as playing a large role in the social and cognitive tasks, *Gtf2i* and *Gtf2ird1*. Mouse models of each gene have shown social phenotypes as well as balance and anxiety phenotypes (92, 96, 97, 101, 152, 153). Since there is evidence that each gene affects similar behaviors, we set out to test the hypothesis that that knocking down both genes simultaneously would lead to more severe phenotypes, suggesting that multiple genes in the WSCR locus affect similar behaviors. Investigating both genes together, rather than individually could provide a more complete understanding of how the genes in the WSCR contribute to the phenotypes of WS.

Gtf2i and *Gtf2ird1* are part of the General transcription factor 2i family of genes. A third member *Gtf2ird2* is located in the WSCR that is variably deleted in patients with WS that have larger deletions(41). This gene family has arisen from gene duplication events, which resulted in high sequence homology between the genes (68). The defining feature of this gene family is the presence of the helix-loop-helix I repeats, which are involved in DNA and protein binding (154). *Gtf2i* has roles that include regulating transcriptional activity in the nucleus, but this multifunctional transcription factor also resides in the cytoplasm where it conveys messages from extracellular stimuli and regulates calcium entry into the cell (74, 76). So far, *Gtf2ird1* has only been described in the nucleus of cells and is thought to regulate transcription and associate with chromatin modifiers (79). The DNA binding of these two transcription factors has been studied in ES cells and embryonic craniofacial tissue. They recognize similar and disparate genomic loci, suggesting that both genes interact to regulate specific regions of the genome (84, 155). However, the DNA binding of these genes has not been studied in the developing brain, which could provide insight on how the general transcription factor 2i family contributes to cognitive and behavioral phenotypes.

We performed ChIP-seq on Gtf2i and Gtf2ird1 in the developing mouse brain to define where these genes bind and also to test the downstream consequences of disrupting the binding. We used the CRISPR/Cas9 system to make a mouse model with a mutation in just *Gtf2ird1* and a mouse model with mutations in both *Gtf2i* and *Gtf2ird1* to test how adding a *Gtf2i* mutation modifies the affects of *Gtf2ird1* mutation. We showed that the mutation in *Gtf2ird1* resulted in the production of an N-truncated protein that disrupts the binding of Gtf2ird1 at the Gtf2ird1 promoter and deregulates the transcription of *Gtf2ird1*. While there are mild consequences of the mutation on transcription genome-wide the mutant mouse exhibited clear balance and marble burying deficits, as well as increased activity. Comparing the single gene mutant to the double mutant did not reveal more severe transcriptional changes or behavioral phenotypes. This suggests that Gtf2ird1 drives the majority of the phenotypes observed in the current studies, and the N-terminal end of this protein has functional consequences on DNA-binding and behavior.

3.3 Results

3.3.1 *Gtf2i* and *Gtf2ird1* bind at active promoters and conserved sites

The paralogous transcription factors, *Gtf2i* and *Gtf2ird1*, have been implicated in the craniofacial and behavioral phenotypes seen in humans with WS as well as mouse models (38, 96, 97, 100, 101, 153). However, the underlying mechanisms by which the general transcription factor 2i family acts are not well understood. One approach to begin to identify how these transcription factors can regulate phenotypes is by identifying where they bind in the genome. This has been done in ES cells and embryonic facial tissue and revealed that both of these transcription factors bind to genes involved in craniofacial development (84). However, these are not relevant tissues that could explain their affects on brain development and subsequent behavior. To overcome this we performed ChIP-seq for Gtf2ird1 and Gtf2i in the developing embryonic day 13.5 (E13.5) brain, a time point when both of these proteins are highly expressed.

We identified 1,410 peaks that were enriched in the Gtf2ird1 IP samples compared to the input. The Gtf2ird1 bound regions were strikingly enriched in the promoter of genes and along the gene body, more so than would be expected by randomly sampling the genome (**Figure 1A**) ($\chi^2 = 1537.8$, d.f. =7, $p < 2.2\text{x}10^{-16}$). The bound peaks were found mostly in H3K4me3 bound regions (Fisher's exact test, $p<2.2\text{x}10^{-16}$), suggesting that they are in active sites in the genome. While the Gtf2ird1 bound regions were also enriched in repressed regions of the genome as defined by H3K27me3 marks (Fisher's exact test, $p< 2.2\text{x}10^{-16}$), 94% of the peaks were in H3K4me3 regions opposed to the 11% of Gtf2ird1 peaks found in H3K27me3 regions (**Figure 1B**), suggesting the Gtf2ird1 may have more of a role in activation than repression.

To understand the common function of the genes that have Gtf2ird1 bound at the promoter we performed GO analysis. The top ten results were consistent with the functions previously described for Gtf2ird1, specifically regulation of transcription and chromatin organization, and we highlighted new categories, such as protein ubiquination (**Figure 1C**). To further test if these regions have functional consequences we compared the conservation of the Gtf2ird1 peaks to a random sample of the genome and found that the Gtf2ird1 peaks are more conserved (t=18.131,d.f.=2403, $p < 2\text{x}10^{-16}$) (**Figure 1D**). We conducted motif enrichment analysis using HOMER to identify other factors that share binding sites with Gtf2ird1 (**Figure**

1E). The GSC motif, which is similar to the core RGATTR motif for Gtf2i and Gtf2ird1, was identified in 4.64% of the targets (65). Interestingly, the CTCF motif was found at 11% of the Gtf2ird1 targets, further supporting its role in chromatin organization.

Gtf2i Chip-seq showed similar results to that of Gtf2ird1. We identified 1,755 WT Gtf2i peaks that had significantly higher coverage in the WT IP compared to the KO IP (**Supplemental Figure 1A**). These peaks were significantly enriched for promoter regions as well as the gene body when compared to random genomic targets (**Figure 2A**)($\chi^2 = 911.63$, d.f.=7, $p < 2.2x10^{-16}$). Similar to Gtf2ird1, the majority of the Gtf2i peaks (78.7%) overlapped H3K4me3 peaks (Fisher's exact test, $p< 2.2x10^{-16}$), with a smaller subset of peaks (20.7%) overlapping with the H3K27me3 mark (Fisher's exact test $p<2.2x10^{-16}$). This suggests that these peaks are located mainly in active regions of the genome (**Figure 2B**). Summarizing the common functions of these target genes by GO analysis, showed enrichment for biological processes such as intracellular signal transduction and phosphorylation (**Figure 2C**). For example, Gtf2i binds within the gene body of the Src gene (**Figure 2D**), which has been shown to phosphorylate Gtf2i itself to activate its transcriptional activity as well as regulate calcium entry into the cell (74, 76). Along with binding to gene promoters, the Gtf2i binding sites are significantly more conserved than random sampling the genome, further suggesting important functional roles of these regions (**Figure 2E**). Motif enrichment of the Gtf2i peaks revealed GC rich binding motifs such as for the KLF/SP family of transcription factors. Interestingly, the Lhx family of transcription factors motif is enriched. Finally, we see an enrichment of the CTCF motif, which Gtf2i has been shown to help target CTCF to specific genomic regions (156) (**Figure 2F**).

3.3.2 *Gtf2i* and *Gtf2ird1* binding sites have distinct features yet overlap at a subset of promoters

One way in which Gtf2i and Gtf2ird1 can interact is by binding the same sites in the genome. We set out to determine how the binding regions of these two genes were similar or different, as well as directly scanning for shared targets. First, we compared the Gtf2i and Gtf2ird1 chip peaks and found that the proportion of annotations of the binding sites are significantly different ($\chi^2 = 282.84$, d.f.=7, $p < 2.2\text{x}10^{-16}$) (**Figure 3A**). While both transcription factors mainly bind in promoters and the gene body, *Gtf2ird1* has a higher proportion of peaks at the promoter compared to *Gtf2i*, whereas *Gtf2i* has more peaks that fall in intergenic regions when compared to *Gtf2ird1*. Interestingly, when we compared them directly to each other the Gtf2ird1 bound peaks were significantly more conserved than the Gtf2i bound peaks (t=7.81,d.f.=2736.5, $p=8.2\text{x}10^{-15}$) (**Figure 3B**). Next, to identify common targets, we looked at the overlap of the genes that had either of the transcription factors at their promoter, and we identified a significant overlap of 148 genes (Fisher's exact test $p < 1\text{x}10^{-38}$) (**Figure 3C**). The GO functions of the overlapped genes highlight specific roles in synaptic functioning and signal transduction (**Figure 3D**). The *Mapk14* gene is an example of a gene involved in signal transduction that has both *Gtf2i* and *Gtf2ird1* bound at its promoter (**Figure 3E**). Interestingly, *Mapk14* is known to phosphorylate *Baz1b* (157), another transcription factor in the WSCR. Shared targets such as this one suggest there are points of convergence where having both genes deleted, such as in WS, might result in synergistic downstream impacts, and further implicates another gene in the WSCR.

3.3.3 Frameshift mutation in *Gtf2ird1* results in truncated protein and affects DNA binding at the *Gtf2ird1* promoter

To investigate the functional role of *Gtf2ird1* and *Gtf2i* at these bound sites and understand how these two genes interact, we set out to make loss of function models of *Gtf2ird1* individually and a double mutant with mutations in both *Gtf2i* and *Gtf2ird1*. We designed two gRNAs, one for *Gtf2ird1* and one for *Gtf2i*, and injected them simultaneously into FVB mouse embryos, to obtain single gene mutations, as well as double gene mutations. We first characterized the consequences of a one base pair adenine insertion in exon three of *Gtf2ird1*. This is an early constitutively expressed exon, and the frameshift mutation introduced a premature stop codon in exon three, which we expected to trigger nonsense-mediated decay (**Figure 4A**). We crossed heterozygous mutant animals to analyze *Gtf2i* and *Gtf2ird1* transcript and protein abundance in heterozygous and homozygous mutants compared to WT littermates (**Figure 4B**). The western blots and qPCR were performed using the whole brain at embryonic day 13.5 (E13.5). As expected, the Gtf2ird1 mutation did not affect the protein or transcript levels of *Gtf2i* (**Figure 4C,D**). Contrary to our prediction that the frameshift mutation would cause nonsense-mediated decay, we observed an ~0.8 CT increase in *Gtf2ird1* transcript with each copy of the mutation and a 40% reduction of the protein in homozygous mutants compared to WT with no significant difference between the WT and heterozygous mutants (**Figure 4E, F**). This suggests that the mutation did have an effect on protein abundance and disrupted the normal transcriptional regulation of the gene.

Similar results were reported in a mouse model that deleted exon two of *Gtf2ird1*, which showed reduced levels of an N-terminally truncated protein caused by a translation re-initiation event at methionine-65 (66). We noticed a slight shift in the homozygous mutant band, which may correspond to the loss of the N-terminal end of the protein. The N-terminal end codes for a conserved leucine zipper, which participates in dimerization as well as DNA-binding (66, 158). Mutating the leucine zipper was shown to affect binding of the protein to the Gtf2ird1 upstream regulatory (GUR) element that is located at the promoter of *Gtf2ird1* (**Figure 4G).** Given the previous findings that Gtf2ird1 negatively autoregulates its transcription and mutating the leucine zipper affects binding to the GUR, we hypothesized that the frameshift mutation diminished the ability of Gtf2ird1 to bind to its promoter resulting in increased transcript abundance. We tested this by performing ChIP-qPCR in the E13.5 brain in WT and *Gtf2ird1*$^{-/-}$ mutants. In the WT brain, Gtf2ird1 IP enriched for the GUR 13-20 times over off-target sequences, which was significantly higher than the Gtf2ird1 IP in the G*tf2ird1*$^{-/-}$ brain (**Figure 1H,I**). Taken together, nonsense transcripts of *Gtf2ird1* with a stop codon in exon three can reinitiate at a lower level to produce an N-truncated protein with diminished binding capacity at the GUR element.

3.3.4 Truncated Gtf2ird1 does not affect binding genome wide

Given that the one base pair insertion did not result in a full knock out of the protein, but did affect its DNA binding capacity at the GUR of *Gtf2ird1*, we tested whether the mutant was a loss of function for all DNA binding. We performed ChIP-seq in the E13.5 *Gtf2ird1*$^{-/-}$ mutants and compared it to the WT Chip-seq data to test the consequences of the mutation on DNA binding genome-wide. The ChIP-seq data confirmed the decrease in binding at the TSS of *Gtf2ird*, however, a small peak is still present at the TSS in the mutant animal, suggesting that the mutation has greatly decreased the binding at this locus (**Figure 4J**). We compared the coverage of the genomic regions identified in the WT ChIP-seq data as bound by Gtf2ird1 in the mutant and WT samples. Surprisingly, the only peak that was identified at an FDR < 0.1 as having differential coverage between the two genotypes was the peak at the TSS of Gtf2ird1

(**Figure 4K**). This suggests that this frameshift mutation has a very specific consequence on how the protein binds to its own promoter that does not affect its binding elsewhere in the genome. The Gtf2ird1 promoter has two instances of the R4 core motif in the sense direction and one instance of the motif in the antisense orientation. We searched the sequences under the identified peaks for similar orientations of the binding motif and found only three other peaks, of which none showed any difference in coverage between genotypes. None of the three other peaks matched the exact spacing of the three motifs found in the Gtf2ird1 promoter. This suggests that the leucine zipper is important for a specific configuration of binding sites that is only present in this one instance in the mouse genome.

3.3.5 Gtf2ird1 frameshift mutation shows mild transcriptional differences

The N-truncated Gtf2ird1 clearly affected the expression levels of *Gtf2ird1* and affected its binding at the promoter of *Gtfi2rd1.* Although we didn't see binding genome-wide perturbed, , it is possible losing the N terminal altered the proteins ability to recruit other transcriptional co-regulators, and thus impact expression. Therefore, we tested the effects of this mutation on transcription genome-wide in the E13.5 brain. We compared the whole brain transcriptome of WT littermates to either heterozygous or homozygous mutants.

Strikingly similarly to the ChIP-seq data, the only transcript with an FDR < 0.1 is *Gtf2ird1*, which was in the same direction as we saw in the qPCR (**Figure 4L** and **Supplemental Figure 2A**). We leveraged the WT ChIP-seq data to see if the presence of Gtf2ird1 at a promoter correlates with gene expression. Binning the genes according to expression level showed that the distribution of Gtf2ird1 targets was different than expected by chance ($\chi^2 = 48.83$, d.f.=3 $p < 1.42\times10^{-10}$), suggesting that highly expressed genes are more likely to have Gtf2ird1 bound at their promoters (**Figure 4M**). To see if there was a more subtle general effect below our sensitivity for a single gene, we tested if the bound Gtf2ird1 targets that are expressed in the brain at E13.5 as a population had their expression shifted. However, we saw only a small trend towards significance between the bound genes and unbound genes, with a mean increase in expression of 0.014 log2 CPM fold change in Gtf2ird1 targets (Kolgmogorov-Smirnov test D=0.038, p=0.079). While this is perhaps unsurprising, because the frameshift mutation did not disturb binding genome wide (**Figure 4N**), the homozygous mutant do have an overall decrease of ~ 50% protein levels which should mimic a WSCR deletion. Thus, transcriptional consequences of haploinsufficiency of this gene might be similarly small.

3.3.6 Frameshift mutation in *Gtf2ird1* is sufficient to affect behavior

Although we observed small differences in DNA binding and overall brain transcription, another Gtf2ird1 model also reported no little effects of Gtf2ird1 on expression transcriptome wide in the brain, yet the model still showed behavioral phenotypes (88, 101). Therefore we tested if our mutation had downstream consequences on adult mouse behavior. There are many single gene knock out models of *Gtf2ird1* and they each show distinct behavioral differences and in some instances the results are contradictory (39, 92, 100, 101). One consistent phenotype across models is motor coordination deficits, which is also an area of difficulty in individuals with WS. Similarly, we observed a significant effect of genotype (H_2=7.88, p =0.01945), on how long the animals could balance on a ledge. Homozygous animals fell off the ledge sooner than WT littermates (p=0.021) (**Figure 5A**). Marble burying has not been reported in other *Gtf2ird1* models, but in larger WS models that either delete the entire syntenic WSCR or delete the proximal half of the region that contains *Gtf2ird1* have shown decreased marble burying phenotypes (90, 93). We observed a similar significant effect of genotype on the number of marbles buried ($F_{2,75}$=7.92, p =0.00076), with the *Gtf2ird1*$^{-/-}$ mutants burying fewer marbles than

WT (p=0.0176) and *Gtf2ird1*$^{+/-}$ littermates (p=0.00067) (**Figure 5B**). Reports of overall activity levels in *Gtf2ird1* mouse models have been discrepant (92, 100). Here we showed that there was only a trend towards a significant main effect of genotype ($F_{2,71}$=2.97, p=0.057) on total distance traveled in a one hour locomotor task, but there is a main effect of sex ($F_{1,71}$=18.77, p=4.76x10^{-5}) and a genotype by sex interaction ($F_{2,71}$=4.98, p=0.0095) (**Figure 5C**). Activity levels were increased in the female *Gtf2ird1*$^{-/-}$ mutants at later time points compared to WT females, and to an intermediate extent in the *Gtf2ird1*$^{+/-}$ mutants (**Supplemental Figure 3A**). There were no differences in total distance traveled between the male genotypes (**Supplemental Figure 3B**). The time spent in the center of an open field is used as a measure of anxiety-like behavior in mice. Anxiety-like behaviors in *Gtf2ird1* models have also been discrepant in the literature (101). Here we showed that there was only a trend for a main effect of genotype when we controlled for sex ($F_{2,71}$=3.070, p=0.0526) (**Figure 5D** and **Supplemental Figure3C, D**).

Finally, as individuals with WS also show high prevalence of phobias, as well as intellectual disability, we tested learning and memory using the conditioned fear task (2, 21). On day one the mice were trained to associated a tone with a footshock and we observed that the mice increased their freezing over time ($F_{2,122}$=26.77, p=2.28x10^{-10}), as expected, and there was a time by genotype interaction ($F_{4,122}$=3.99, p=0.004) where the WT mice froze more during the last five minutes of the task compared to both the *Gtf2ird1*$^{+/-}$ (p=0.007) and the *Gtf2ird1*$^{-/-}$ mutants (p=0.002) (**Figure 5E**). On the second day, contextual fear memory was tested. We placed the mice in the same chamber in which they were delivered the footshock and measured their freezing behavior in the absence of the footshock and the tone. All genotypes exhibited a fear memory response as indicated by the significant effect of the context compared to baseline of day one ($F_{1,61}$=31.83, p=4.63x10^{-7}) but no main effect of genotype ($F_{2,61}$=1.24, p=0.30). Each genotype group froze more during the first two minutes of day two than on day one (WT: $p=4.7x10^{-6}$, *Gtf2ird1*$^{+/-}$: p=0.034, *Gtf2ird1*$^{-/-}$: p=0.0061) (**Supplemental Figure 3E**). When we analyzed the entire time of the experiment of contextual fear we similarly saw no main effect of genotype ($F_{2,61}$=2.36, p=0.010), but a significant effect of time ($F_{7,427}$=4.43, $p=9.14x10^{-5}$) and a time by genotype interaction ($F_{14,427}$=2.19, p=0.0077), suggesting that the freezing behavior of the genotypes differ at certain time points during the task. Post hoc analysis showed that during minute two the WT animals are freezing significantly more than the *Gtf2ird1*$^{+/-}$ mutants (p=0.0008) suggesting a reduced contextual fear memory response (**Figure 5F**). On day three of the experiment, we tested cued fear by placing the animals in a different context but played the tone that was paired with the shock on day one. All genotypes had a similar response to the tone ($F_{2,61}$=1.12, p=0.334) (**Figure 5G**). These differences could not be explained by differences in shock sensitivity (flinch: H_2=3.34, p=0.19, escape: H_2=2.98, p=0.23, vocalization: $F_{2,56}$=4.24, p=0.12) (**Supplemental Figure 3F**).

Overall, these behavior analyses show that the N-terminal truncation and/or the decreased total protein levels of the *Gtf2ird1* mutant can still result in adult behavioral phenotypes, specifically in the domains such as balance, activity, and marble burying. The most severe phenotypes were observed in the homozygous mutants.

3.3.7 Generation of *Gtf2i* and *Gtf2ird1* double mutant

The evidence that this frameshift mutation in *Gtf2ird1* has functional consequences on some of its DNA binding capacity as well as leads to behavioral phentoypes led us to characterize a double mutant that was generated during the dual gRNA CRISPR/Cas9 injection. This mutant allowed us to test the effects of knocking out *Gtf2i* along with mutating *Gtf2ird1*, as well as test the consistency of the previous *Gtf2ird1* phenotypes across different mutations. The double mutant described here has a two base pair deletion in exon five of *Gtf2i* and a 590 base pair deletion that encompasses most of exon three of *Gtf2ird1* (**Figure 6A**). We carried out a heterozygous cross of the double mutants to similarly test the protein and transcript abundance of each gene in the heterozygous and homozygous state. The homozygous double mutant is embryonic lethal due to the lack of *Gtf2i*, which has been described in other *Gtf2i* mutants (**Figure 6B**) (87, 96). We were able to detect homozygous embryos up to E15.5. Thus we focused molecular analyses on E13.5 mice for the reasons mentioned above. The two base pair deletion in exon five of *Gtf2i* leads to a premature stop codon and is a full knock out of the protein, and decreases the transcript abundance consistent with the degradation of the mRNA due to nonsense-mediated decay (**Figure 6C,D**). The 590 base pair deletion in *Gtf2ird1* removes all of exon three except the first 14 base pairs. This mutation has a larger effect on protein levels compared to the one base pair insertion, but a small amount of a truncated protein is still made at about 10% of the level of WT protein. We observed the same increase in transcript abundance that was detected in the one base pair insertion mutation (**Figure 6E,F**).

3.3.8 Knocking down both *Gtf2i* and *Gtf2ird1* produces mild transcriptome changes

To test if having both *Gtf2i* and *Gtf2ird1* mutated had a larger effect on the transcriptome we performed whole brain RNA-seq analysis on WT E13.5 brains and compared them to *Gtf2i*$^{+/-}$/*Gtf2ird1*$^{+/-}$ littermates. There were only mild differences between the transcriptomes of the two genotypes similar to what was seen when we compared WT littermates to *Gtf2ird1*$^{-/-}$ mutants (**Figure 6G**). We also compared WT transcriptomes to the homozygous double mutants, which showed a greater difference between genotypes. However, this is probably due to the fact that the homozygous double mutants have a very severe phenotype, which includes neural tube closure defects. The GO terms suggested that overall nervous system development and glial cell differentiation is disrupted (**Supplemental Figure 4A,B**). We also coupled the Gtf2i ChIP-seq data with the RNA-seq data. Unlike what we saw with Gtf2ird1 bound genes, there was association between the expression levels of genes and the presence of Gtf2i ($\chi^2 = 6.58$, d.f.=3 p=0.086) (**Figure 6H**). This is consistent with a previous report of *Gtf2i* ChIP-seq data. There is a slight but significant shift to higher expression of genes of about 0.02 log2 CPM fold change that are bound to Gtf2i compared to genes that are not bound (Kolgmogorov-Smirnov test D=0.075, p=$9.50\text{x}10^{-5}$) (**Figure 6I**).

3.3.9 Double mutants show similar behavioral consequences similar to single *Gtf2ird1* mutants

To test the effects of mutating both *Gtf2i* and *Gtf2ird1* we crossed the heterozygous double mutant to the single *Gtf2ird1* heterozygous mouse (**Figure 7A**). This breeding strategy produced four littermate genotypes, WT, $Gtf2ird1^{+/-}$, $Gtf2i^{+/-}/Gtf2ird1^{+/-}$, and $Gtf2i^{+/-}/Gtf2ird1^{-/-}$ for direct and well-controlled comparisons. To test the effects of adding a *Gtf2i* mutation along with a *Gtf2ird1* mutation we compared the $Gtf2ird1^{+/-}$ to their $Gtf2i^{+/-}/Gtf2ird1^{+/-}$ littermates. The final genotype tested the effects of the heterozygous *Gtf2i* mutation in the presence of both the *Gtf2ird1* mutations. To be thorough we tested the protein and transcript abundance of each gene in the four genotypes. As expected all genotypes with the *Gtf2i* mutation showed decreased protein and transcript levels. The *Gtf2ird1* results reflected what was previously shown for each mutation, however, the $Gtf2i^{+/-}/Gtf2ird1^{-/-}$ did not show any further detectable decrease in protein abundance compared to the $Gtf2i^{+/-}/Gtf2ird1^{+/-}$ genotype (**Supplemental Figure 5A-D**).

We repeated the same behaviors that were performed on the one base pair *Gtf2ird1* mutants. We saw a similar significant effect of genotype on balance (H_3=10.68, p=0.014), with the $Gtf2i^{+/-}/Gtf2ird1^{-/-}$ falling off sooner compared to WT littermates (p=0.025) (**Figure 7B**).

There was no significant difference between the *Gtf2ird1*$^{+/-}$ and *Gtf2i*$^{+/-}$/*Gtf2ird1*$^{+/-}$ genotypes, suggesting that in the heterozygous state decreasing the dosage of *Gtf2i* does not strongly modify the *Gtf2ird1*$^{+/-}$ phenotype. These results were replicated in a subsequent cohort (**Supplemental Figure 5E**). There was a significant effect of genotype on the number of marbles buried (F$_{3,76}$=2.93, p=0.039). Post hoc analysis showed a significant difference between only the *Gtf2ird1*$^{+/-}$ and *Gtf2i*$^{+/-}$/*Gtf2ird1*$^{-/-}$ littermates (p=0.050) (**Figure 7C**), with a trend in the same direction as was previously seen in the *Gtf2ird1*$^{-/-}$ mutants. We saw a main effect of genotype on activity levels in the one hour locomotor task (F$_{3,69}$=3.22, p=0.028), but we did not see the same main effect of sex (F$_{1,69}$=2.29, p=0.14), or a sex by genotype interaction (F$_{3,69}$=1.82, p=0.15); however we did see a three way sex by time by genotype interaction (F$_{15,345}$=1.95, p=0.018). The combined sex data showed that the *Gtf2i*$^{+/-}$/*Gtf2ird1*$^{-/-}$ travel more distance in the later time points than the WT and *Gtf2ird1*$^{+/-}$ at time point 40 (**Figure 7D**). When we looked at the data by sex we saw a larger effect in the females with the *Gtf2ird1*$^{+/-}$ and *Gtf2i*$^{+/-}$/*Gtf2ird1*$^{+/-}$ intermediate to the *Gtf2i*$^{+/-}$/*Gtf2ird1*$^{-/-}$ (**Supplemental Figure 5F, G**). There was also a main effect of genotype on the time spent in the center of the apparatus (F$_{3,69}$=3.60, p=0.018). The *Gtf2i*$^{+/-}$/*Gtf2ird1*$^{-/-}$ spent less time in the center during the first ten minutes of the task compared to WT (p=0.0019) littermates with the *Gtf2ird1*$^{+/-}$ and *Gtf2i*$^{+/-}$/*Gtf2ird1*$^{+/-}$ showing intermediate values (**Figure 7E**).

Finally, we repeated the conditioned fear memory task using this breeding strategy. All genotypes increased their freezing after each foot shock on day one as expected. The WT animals exhibited higher freezing during minute one of baseline, but this difference diminished during minute two (**Figure 7F**). All animals showed a contextual fear memory response when they were re-introduced to the chamber on day two (F$_{1,68}$=81.21, p=3.21x10^{-13}) (**Supplemental**

Figure 5H) but there was no main effect of genotype ($F_{3,68}$=1.61, p=0.19) (**Figure 7G**). On day three, when cued fear was tested, there was a significant effect of genotype on the freezing behavior ($F_{3,68}$=3.17, p=0.030) and a time by genotype interaction ($F_{21,476}$=1.63, p=0.040). During minute five of the task the *Gtf2i*$^{+/-}$/*Gtf2ird1*$^{-/-}$ mutants froze significantly more than the WT (p=0.030) as did the *Gtf2ird1*$^{+/-}$ (p=0.024) (**Figure 7H).** The cued fear phenotype could not be explained by differences in sensitivity to the foot shock (**Supplemental Figure 5I**).

By crossing these two mutant lines we tested the hypothesis that the double heterozygous mutant would be more severe than a mutation only affecting *Gtf2ird1*. Comparing the *Gtf2ird1*$^{+/-}$ and *Gtf2i*$^{+/-}$/*Gtf2ird1*$^{+/-}$, showed mild deficits compared to WT littermates that in some cases were intermediate to phenotypes of the *Gtf2i*$^{+/-}$/*Gtf2ird1*$^{-/-}$. There were no instances when either the *Gtf2ird1*$^{+/-}$ or *Gtf2i*$^{+/-}$/*Gtf2ird1*$^{+/-}$ genotype was significantly different than the other, suggesting that in the behaviors that we have tested, *Gtf2i* mutation does not modify the effects of a *Gtf2ird1* mutation. This unique cross also allowed us to characterize a new mouse line *Gtf2i*$^{+/-}$/*Gtf2ird1*$^{-/-}$, which had the largest impact on behaviors. The phenotypes of *Gtf2i*$^{+/-}$/*Gtf2ird1*$^{-/-}$ were always in the same direction as the phenotypes in the *Gtf2ird1*$^{-/-}$ mouse model, but we also saw a significant cued fear deficit when the *Gtf2i* mutation was added. This further supports that the behaviors tested here, such as activity levels, balance, anxiety-like behaviors, marble burying, and learning and memory are largely affected by homozygous mutations in *Gtf2ird1*.

3.4 Discussion

We have described the *in vivo* DNA binding sites of Gtf2ird1 and Gtf2i in the developing mouse brain. This is the first description of these two transcription factors in a tissue that is relevant for the behavioral phenotypes that are seen in mouse models of WS. Gtf2ird1 showed a preference for active sites and promoter regions. The conservation of the Gtf2ird1 targets was higher on average than would be expected by chance, which provides evidence that these are functionally important regions of the genome. The functions of genes that are bound by Gtf2ird1 include transcriptional regulation as well as post translational regulation. A role for Gtf2ird1 in regulating genes involved in protein ubiqluniation has not been described before. Genes involved in chromatin organization were also found to be bound by Gtf2ird1. This supports the role of Gtf2ird1 in regulating chromatin by transcriptionally controlling other chromatin modifiers. Along with its localization pattern in the nucleus and its direct interaction with other chromatin modifiers such as ZMYM5 (79, 82), this data suggests that Gtf2ird1 can exert its regulation of chromatin at several different levels of biological organization. The motif enrichment of Gtf2ird1 peaks showed that CTCF may be present along with Gtf2ird1, further implicating the importance of Gtf2ird1 in chromatin biology. Interestingly, Gtf2i has been show to interact with and target CTCF to specific sites in the genome (156). It would be interesting to test if Gtf2ird1 has a similar relationship with CTCF and targets it to unique genomic loci.

Overall, Gtf2i showed a similar preference for promoters and active regions, although it had more intergenic targets than Gtf2ird1, and the conservation of Gtf2i peaks was significantly lower than the Gtf2ird1 peaks. The genes bound by Gtf2i were enriched for signal transduction and phosphorylation. Interestingly, Gtf2i was bound to the gene body of the *Src* gene. Src is known to phosphorylate Gtf2i to induce its transcriptional activity (74). Phosphorylation of Gtf2i by Src also antagonizes calcium entry into the cell (76). While, knocking out Gtf2i did not affect the expression of *Src*, it would be interesting to understand the functional consequence of Gtf2i bound to *Src*, especially since knockout mice of *Src* exhibit similar behaviors as *Gtf2i* knock out mice (75).

The overlap of targets of Gtf2i and Gtf2ird1 was significant, and the genes that did overlap were enriched for synaptic activity and signal transduction. This was evidence that these genes could interact via their binding targets to produce cognitive and behavioral phenotypes. To test how mutating both Gtf2i and Gtf2ird1 would modify the phenotypes of just Gtf2ird1 we characterized two new mouse models. We used the CRISPR/Cas9 system to generate multiple mutations in the two genes individually as well as together from one embryo injection. The ease and combinatorial possibilities of this technology will be amenable to testing many unique combinations of genes in copy number variant regions, which will be important to fully understand the complex relationships of genes in these disorders.

We saw that a frameshift mutation that we expected to trigger non-sense mediate decay in Gtf2ird1 did not and resulted in a mild reduction in protein levels in the homozygous mutant and an N-terminal truncation. Even making a larger 590bp deletion of exon three in Gtf2ird1 did not result in the degradation of the mRNA, but did have a larger effect on the protein, even though some protein product was still made. This phenomenon has been seen in at least two other mouse models of Gtf2ird1 (66, 101). These were made using classic homologous recombination removing either exon two or exon two through part of exon five. In both models Gtf2ird1 transcript was still made, but no *in vivo* protein analysis was done due to poor quality antibodies and the low expression of the protein. The presence of an aberrant protein that can still bind the genome, as the mutant described here can, could explain the lack of transcriptome differences in the brain shown here as well as in (88). It could also be that the mutant protein can still interact with other binding partners and be trafficked to the appropriate genomic loci. This mutation did disrupt the binding of Gtf2ird1 to its own promoter, which resulted in an increase in transcript levels. The property that specifies the Gtf2ird1 binding to its own promoter must be very unique, as DNA binding genome-wide was not perturbed in the mutant.

Nonetheless, the mutated Gtf2ird1 protein was still sufficient to cause adult behavioral abnormalities. This supports the hypothesis that the N-terminal end of the protein has other important functions beyond DNA binding. Similarly, the N-truncation of Gtf2i did not affect DNA-binding, but still resulted in behavioral deficits (67). The single Gtf2ird1 homozygous mutant showed balance deficits, which is consistent across many mouse models of WS. We also observed decreased marble burying. This task is thought to be mediated by hippocampal function, suggesting a possible disruption of the hippocampus caused by this mutation (159). We saw an increase in overall activity levels in female Gtf2ird1 mutants. This could relate to the high prevalence of Attention Deficit/Hyperactivity Disorder seen in WS (22).

Given the prior evidence that these two transcription factors are both involved in the cognitive and behavioral phenotypes of WS (34, 95), and the evidence that their shared binding targets regulate synaptic genes, we tested if having both Gtf2i and Gtf2ird1 mutated could modify the phenotype seen when just Gtf2ird1 was mutated. Contrary to our prediction, we did not see a large effect of adding a Gtf2i mutation to differences in transcriptome wide expression or behavioral phenotypes. This was also surprising given that we successfully reduced *Gtf2i* protein and it has been described in the literature as regulating transcription (58). It could be that by using the whole E13.5 brain we are diminishing the effects of transcriptional differences seen in a specific rare cell types. This potential confound could be overcome using single cell sequencing technologies in the future.

When Gtf2i was knocked down it the presence of two Gtf2ird1 mutations, we saw phenotypes in the same direction as the homozygous one base pair insertion *Gtf2ird1* mutant as well as significant results in the cued fear memory task. Thus, the behaviors tested in this study seem to be mainly driven by Gtf2ird1 homozygosity, which is consistent across the two different mutations. This does not exclude the possibility that Gtf2i can modify the phenotype of Gtf2ird1 knockdown in other behavioral domains. For example, it would be interesting to see the effect of adding Gtf2i on top of a Gtf2ird1 mutation on social behaviors.

Our study has provided the first description of the DNA-binding of both Gtf2i and Gtf2ird1 in the developing mouse brain and showed that they have unique and overlapping targets. These data will be used to inform downstream studies to understand how these two transcription factors interact with the genome. We generated two new mouse models that tested the importance of the N-terminal end of Gtf2ird1 and the affect of mutating both *Gtf2i* and *Gtf2ird1*. We provided evidence that despite either gene having little effect on transcription the *Gtf2ird1* mutation affects balance, marble burying, activity levels, and cued fear memory.

3.5 Materials and Methods

Generating genome edited mice

To generate unique combinations of gene knockouts we designed gRNAs targeting early constitutive exons of the mouse *Gtf2i* and *Gtf2ird1* genes. The gRNAs were tested for cutting efficiency in cell culture by transfecting N2a cells with the pX330 Cas9 expression plasmids (Addgene) that had each gRNA cloned into it. The DNA was harvested from the cells and cutting was detected using the T7 endonuclease assay. The gRNAs were *in vitro* transcribed using the MEGAShortScript kit (Ambion) and the Cas9 mRNA was *in vitro* transcribed using the mMessageMachine kit (Ambion). The two gRNAs and Cas9 mRNA were then injected into FVB mouse embryos and implanted into donor females. FVB mice were used for their large pronuclei and large litter sizes. The resulting offspring were genotyped for mutations by designing gene specific primers that had the illumina adapter sequences concatenated to their 3’ prime end to allow for deep sequencing of the amplicons surrounding the expected cut sites. The large 590 bp deletion was detected by amplifying 3.5kb that included exon two, exon three and part of intron three then using a Nextera library prep (Illumina) to deep sequence the amplicon. We described two founder mice obtained from these injections. Each founder line was bred to FVB/ANTJ mice to ensure the mutations detected were in the germline and on the same chromosomes in the case of founders with mutations in both genes. The mice were also crossed until the mutations were on a complete FVB/ANTJ background, which differs from the FVB background at two loci; Tyr$^{c\text{-}ch}$, which gives the chinchilla coat color of FVB/ANTJ and 129P2/OlaHSd *Pde6b* allele, which the FVB/ANTJ are WT for and prevents them from becoming blind in adult hood. The coat color was genotyped by eye, and the *Pde6b* gene was genotyped using the primers provided by the Jackson Laboratory website.

Western blotting

Embryos were harvested on embryonic day 13.5 (E13.5) and the whole brain was dissected in cold PBS and flash frozen in liquid nitrogen. The frozen brains were stored at -80° C until they were to be lysed. The frozen brain was homogenized in 500ul of 1xRIPA buffer (10mM Tris HCl pH 7.5, 140mM NaCl, 1mM EDTA, 1% Triton X-100, 0.1% DOC, 0.1% SDS, 10mM Na_3V0_4, 10mM NaF, 1x protease inhibitor (Roche)) along with 1:1000 dilution of RNAase inhibitors (RNasin (Promega) and SUPERase In (Thermo Fisher Scientific). The homogenate incubated on ice for 20 minutes and was then spun at 10,000g for 10 minutes at 4° C to clear the lysate. The lysate was stored as two aliquots of 100ul in the -80° C for protein analysis and 250ul of the lysate was added to 750ul of Trizol LS and stored at -80° C for later RNA extraction and qPCR. Total protein was quantified using the BCA assay and 25-50ug of protein in 1x Lamelli Buffer with B-mercaptoethanol was loaded onto 4-15% TGX protean gels from Bio-Rad. The protein was transferred to a .2um PVDF membrane by wet transfer. The membrane was blocked with 5% milk in TBST for one hour at room temperature. The membrane was cut at the 75KDa protein marker and the bottom was probed with a Gapdh antibody as an endogenous loading control, and the top was probed with an antibody for either Gtf2i or Gtf2ird1. The primary incubation was performed overnight at 4° C. The membrane was then washed three times in TBST for five minutes then incubated with a secondary antibody HRP conjugated antibody diluted in 5% milk in TBST for one hour at room temperature. The blot was washed three times with TBST for five minutes then incubated with Clarity Western ECL substrate (Bio-Rad) for five minutes. The blot was imaged in a MyECL Imager (Thermo Scientific). The relative protein abundance was quantified using Fiji (NIH) and normalized to Gapdh levels in a reference WT sample. The antibodies and dilutions used in this study were: Rabbit anti-GTF2IRD1 (1:500, Novus, NBP1-91973), Mouse anti-GTF2I (1:1000 BD Transduction Laboratories, BAP-135), and Mouse anti-Gapdh (1:10,000, Sigma Aldrich, G8795), HRP-conjugated Goat anti Rabbit IgG (1:2000, Sigma Aldrich, AP307P) and HRP-conjugated Goat anti Mouse IgG (1:2000, Bio Rad, 1706516).

Transcript abundance using RT-qPCR

RNA was extracted from Trizol LS using the Zymo Clean and Concentrator-5 kit with on column DNAase-I digestion following the manufacturer's instructions. The RNA was eluted in 30ul of RNAse free water and quantified using a Nanodrop 2000 (Thermo Scientific). One ug of

RNA was transcribed into cDNA using the qScript cDNA synthesis kit (Quanta Biosciences). The cDNA was used in a 10ul PCR reaction with 500nM of target specific primer and the PowerUP Sybr green master mix (Applied Biosystems). The primers were designed to amplify exons that were constitutively expressed in both *Gtf2i* (exons 25 and 27) and *Gtf2ird1* (exons 8 and 9) and span an intron. The RT-qPCR was carried out in a QuantStudio6Flex machine (Applied Biosystems) using the following cycling conditions: 95° C 20 seconds, 95° C 1 second, 60° C 20 seconds, repeat steps 2 through 3 40 times. Each target and sample was run in triplicate technical replicates, with three biological replicates for each genotype. The relative transcript abundance was determined using the delta CT method normalizing to *Gapdh*.

ChIP

Chromatin was prepared as described in (160). Briefly, frozen brains were homogenized in 10mL of cross-linking buffer (10mM HEPES pH7.5, 100mM NaCl, 1mM EDTA, 1mM EGTA, 1% Formaldehyde (Sigma)). The homogenate was spun down and resuspended in 5mL of 1x L1 buffer (50mM HEPES pH 7.5, 140 mM NaCl, 1mM EDTA, 1mM EGTA, 0.25% Triton X-100, 0.5% NP40, 10.0% glycerol, 1mM BGP (Sigma), 1x Na Butyrate (Millipore), 20mM NaF, 1x protease inhibitor (Roche)) to release the nuclei. The nuclei were spun down and resuspended in 5mL of L2 buffer (10mM Tris-HCl pH 8.0, 200mM NaCl, 1mM BGP, 1x Na Butyrate, 20mM NaF, 1x protease inhibitor) and rocked at room temperature for five minutes. The nuclei were spun down and resuspended in 950ul of buffer L3 (10mM Tris-HCl pH 8.0, 1mM EDTA, 1mM EGTA, 0.3% SDS, 1mM BGP, 1x Na Butyrate, 20mM NaF, 1x protease inhibitor) and sonicated to a fragment size of 100-500bp in a Covaris E220 focused-ultrasonicator with 5% duty factor, 140 PIP, and 200cbp. The sonicated chromatin was diluted in with 950ul of L3 buffer and 950ul of 3x covaris buffer (20mM Tris-HCl pH 8.0, 3.0% Triton X-

100, 450mM NaCl, 3mM EDTA). The diluted chromatin was pre-cleared using 15ul of protein G coated streptavidin magnetic beads (ThermoFisher) for two hours at 4° C. For IP, 15ul of protein G coated streptavidin beads were conjugated to either 10ul of Gtf2ird1 antibody (Rb anti GTF2IRD1 NBP1-91973 LOT:R40410) or 10ul of Gtf2i antibody (Rb anti GTF2I Bethyl Laboratories) for one hour at room temperature. 80ul of the pre-cleared lysate was saved to be the input sample. 400ul of the pre-cleared lysate was added to the beads and incubated overnight at 4° C. The IP was then washed two times with low salt wash buffer (10mM Tris-HCl pH 8.0, 2mM EDTA, 150mM NaCl, 1.0% Triton X-100, 0.1% SDS), two times with a high salt buffer (10mM Trish-HCl pH 8.0, 2mM EDTA, 500mM NaCl, 1.0% Triton X-100, 0.1% SDS), two times with LiCl wash buffer (10mM Tris-HCl pH 8.0, 1mM EDTA, 250mM LiCl (Sigma), 0.5% NaDeoxycholate, 1.0% NP40), and one time with TE (10mM Tris-HCl pH 8.0, 1mM EDTA) buffer. The DNA was eluted off of the beads with 200ul of 1x TE and 1% SDS by incubating at 65° C in an Eppendorf R thermomixer shaking at 1400rpm. The DNA was de-crosslinked by incubating at 65° C for 15 hours in a thermocycler. RNA was removed by incubating with 10ug of RNAse A (Invitrogen) at 37° C for 30 minutes and then treated with 140ug of Proteinase K (NEB) incubating at 55° C in a thermomixer mixing at 900rpm for two hours. The DNA was extracted with 200ul of phenol/chloroform/isoamyl alcohol (Ambion) and cleaned up using the Qiagen PCR purification kit and eluted in 60ul of elution buffer. Concentration was assessed using the highsensitivity DNA kit for qubit (Thermo Fisher Scientific).

ChIP-qPCR

Primers were designed to amplify the upstream regulatory element of *Gtf2ird1*. Two off target primers were designed that are 10kb upstream of the transcription start site of *Bdnf* and 7kb upstream of the *Pcbp3* transcription start site. The input sample was diluted 1:3, 1:30, and

1:300 to create a standard curve for each primer set and sample. Each standard, input, and IP sample for each primer set was performed in triplicate in 10ul reactions using the PowerUP Sybr green master mix (Applied Biosystems) and 250nM of forward and reverse primers. The reactions were performed in a QuantStudio6Flex machine (Applied Biosystems) with the following cycling conditions: 50° C for 2 minutes, 95° C for 10 minutes, 95° C 15 seconds, 60° C for 1 minute, repeat steps 3 through 4 40 times. The relative concentration of the input and IP samples were determined from the standard curve for each primer set. Enrichment of the IP samples was determined by dividing the on target upstream regulatory element relative concentration by the off target relative concentration.

ChIP-seq

ChIP-seq libraries were prepared using the Swift Accel-NGS 2S plus DNA library prep kits with dual indexing (Swift Biosciences). The final libraries were enriched by thirteen cycles of PCR. The libraries were sequenced by the Genome Technology Access Center at Washington University School of Medicine on a HiSeq3000 producing 1x50 reads.

Raw reads were trimmed of adapter sequences and bases with a quality score less than 25 using the Trimmomatic Software (161). The trimmed reads were aligned to the mm10 genome using the default settings of bowtie2 (162). Reads that had a mapping quality of less than 10 were removed. Picard tools was used to remove duplicates from the filtered reads (http://broadinstitute.github.io/picard). Macs2 was used to call peaks on the WT IP, *Gtf2ird1*$^{-/-}$ IP, and *Gtf2i*$^{-/-}$/*Gtf2ird1*$^{-/-}$ IPs with the corresponding sample's input as the control sample for each biological replicate (163). Macs2 used an FDR of 0.01 as the threshold to call a significant peak. High confidence peaks were those peaks that had some overlap within each biological replicate for each genotype using bedtools intersect (164). The read coverage for the high confidence peaks identified in the WT IPs was determined using bedtools coverage for all genotypes. To identify peaks with differential coverage, we used EdgeR to compare the WT peaks coverage files to the corresponding mutant peak coverage and differential peaks were defined as having an $FDR < 0.1$ (165). The peaks with $FDR < 0.1$ and $\log2FC > 0$ fine the Gtf2i high confidence peaks calls, since this mutation represents a full knockout of the protein. Annotations of peaks and motif analysis was performed using the HOMER software on the high confidence peaks (166). Peaks were annotated at the transcription start (TSS) of genes if the peak overlapped the +2.5kbp or -1kbp of the TSS using a custom R script. GO analysis on the ChIP target genes was performed using the goseq R package. We used E13.5 H3K4me3 and E13.5 H3K27me3 forebrain narrow bed peak files from the mouse ENCODE project to overlap with our peak datasets (167). Deeptools was used to generate bigwig files normalized to the library size for each sample by splitting the genome into 50bp overlapping bins (168). Deeptools was used to visualize the ChIP-seq coverage within the H3K4me3 and H3K27me3 peak regions. PhyloP scores for the WT ChIP-seq peaks and random genomic regions of the same length were retrieved using the UCSC table browser 60 Vertebrate Conservation PhyloP table. The Epigenome browser was used to visualize the ChIP-seq data as tracks.

RNA-seq

1ug of E13.5 whole brain total RNA extracted from Trizol LS was used as input for rRNA depletion using the NEBNExt rRNA Depletion Kit (Human/Mouse/Rat). The rRNA depleted RNA was used as input for library construction using the NEBNext Ultra II RNA library prep kit for Illumina. The final libraries were indexed and enriched by PCR using the following thermocycler conditions: 98° C for 30 seconds, 98° C 10 seconds, 65° C 1 minute and

15 seconds, 65° C 5 minutes, hold at 4° C, repeat steps 2 through 3 6 times. The libraries were sequenced by the Genome Technology Access Center at Washington University School of Medicine on a HiSeq3000 producing 1x50 reads.

RNA-seq analysis

The raw RNA-seq reads were trimmed of Illumina adapters and bases with quality scores less than 25 using Trimmomatic Software. The trimmed reads were aligned to the mm10 mouse genome using the default parameters of STARv2.6.1b (169) . We used HTSeq-count to determine the read counts for features using the Ensembl GRCm38 version 93 gtf file (170). Differential gene expression analysis was done using EdgeR. We compared the expression of genes that are targets of either Gtf2ird1 or Gtf2i to non-bound genes by generating a cumulative distribution plot of the average log CPM of the genes between genotypes. GO analysis was performed using the goseq R package.

Behavioral tasks

Animal statement

All animal testing was done in accordance with the Washington University in St. Louis animal care committee regulations. Mice were group housed in same-sex, mixed-genotype cages with two to five mice in a cage in standard mouse cages with dimensions 28.5 x 17.5 x 12 cm with corn cob bedding. The mice had ad libitum access to food and water and followed a 12 hour light-dark cycle with the lights on from 6:00am-6:00pm. The rooms the animals were housed in were kept at 20-22° C and a relative humidity of 50%. All mice were maintained on the FVB/AntJ ((171)) background from Jackson Labs. All behaviors were done in adulthood between ages P60-P130. A week prior to beginning behavior testing the mice were handled by the male experimenter. On days of testing the mice were moved to the testing room and allowed to habituate to the room and the male experimenter for 30 minutes before testing started. The number of mice and behaviors are listed in Table 1 and Table 2.

Ledge

To test balance, we timed how long a mouse could balance on a plexiglass ledge with a width of 0.5cm and a height of 38cm as described in (171). The mice were timed up to 60 seconds. If the mouse fell off within the first five seconds the time was restarted and the mouse was given another attempt. If after the third attempt the mouse fell off within the first five seconds that time was recorded. We tested all mice on the ledge and then allowed for a 20 minute rest time then repeated the testing on all the mice for a total of two trials for each mouse. The average of the two trials were used in the analysis.

One hour locomotor activity

We assessed activity levels in a one hour locomotor task, as previously described (171). Mice were placed in the center of a standard rat cage with dimensions 47.6 x 25.4 x 20.6cm. The rat cage was located inside of a sound-attenuating box with white light set to 24 lux. The mice could freely explore the cage for one hour. A plexiglass lid with air holes was placed on top of the rat cage to prevent the mice from jumping out of the cage. The position and horizontal movement of the mice was tracked using the ANY-maze software (Stoelting Co.: RRID: SCR_014289). The apparatus was divided into two zones, the edge zone was 5.5cm bordering the cage, and a 33 x 11cm center zone. The animal was considered in a particular zone if 80% of the mouse was detected in the zone. ANY-maze recorded the time, distance, and number of entries into each zone. After the task, the mouse was returned to its home cage and the apparatus was thoroughly cleaned with 70% ethanol.

Marble burying

Marble burying is a species-specific task that measures the compulsive digging behavior of mice. Normal hippocampal is thought to be required for normal marble burying phenotypes. We tested marble burying as previously described (171). A rat cage was filled with aspen bedding to a depth of 3cm and placed in a sound-attenuating box with white light set at 24 lux. A 5 x 4 grid of evenly spaces marbles was laid out on top of the bedding. The experimental mouse was placed in the center of the chamber and allowed to freely explore and dig in the chamber for 30 minutes. A plexiglass lid with air holes was placed on top of the rat cage to prevent the mice from escaping. After 30 minutes the animal was returned to their home cage. Two scorers counted the number of marbles not buried. A marble was considered buried if two-thirds of the marble was covered with bedding. The number of marbles buried was then determined, and the average of the two scorers was used in the analysis. After the marbles were counted the bedding was disposed of and the rat cage and marbles were cleaned with 70% ethanol.

Contextual and Cued Fear Conditioning

Learning and memory were tested using the contextual and cued fear condition paradigm as previously described (172). Contextual fear memory is thought to be driven by hippocampal functioning whereas cued fear is thought to be driven by amygdala functioning. On day one of the experiment, animals were placed in a Plexiglas chamber (26cm x 18cm x 18cm; Med Associates Inc.) with a metal grid floor that had an unobtainable peppermint odor. A chamber light was on for the duration of the five-minute task. During the first two minutes, the animal freely explored the apparatus, and this was considered the baseline. An 80dB white noise tone was played for 20 seconds at 100 seconds, 160 seconds, and 220 seconds during the five-minute task. During the last two seconds of the tone, the mice received a 1.0mA foot shock. The tone is the conditioned stimulus (CS) and the foot shock is the unconditioned stimulus (UCS). The animal's freezing behavior was monitored by the FreezeFrame (Actimetrics, Evanston, IL) software in 0.75s intervals. Freezing was defined as no movement besides respiration, and was used as a measure of the fear response of mice. After the five-minute task the mice were returned to their home cage. On day two, we tested contextual fear memory. The mice were placed in the same chamber as day with the unobtainable peppermint odor, and the freezing behavior was measured over the eight-minute task. The first two minutes of day two were compared to the first two minutes of day one to test for the acquisition of the fear memory. The mice were then returned to their home cage. On day three, to test cued fear, the mice were placed in a new black and white chamber that was partitioned into a triangle shape and had an unobtainable coconut scent. The mice were allowed to explore the chamber and the first two minutes were considered baseline. After minute two the 85 dB tone (CS) was played for the remaining eight minutes.

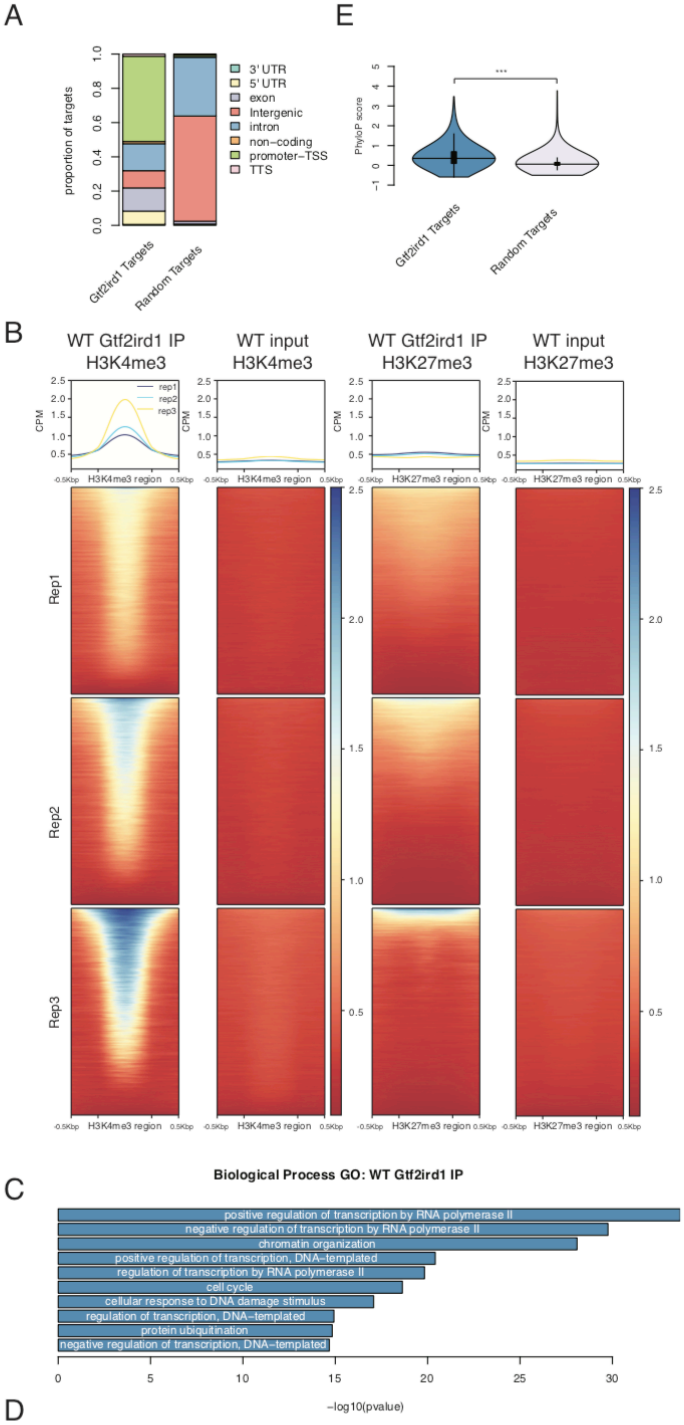
Statistical Analysis

All statistical analyses were performed in R v3.4.2. All statistical tests are reported in Supplemental Table 1. The ANOVA assumption of normality was assessed using the Shapiro-Wilkes test and manual inspection of qqPlots, and the assumption of equal variances was assessed with Levene's Test. When appropriate ANOVA was used to test for main effects and interaction terms. Post hoc analyses were done to compare between genotypes. If the data violated the assumptions of ANOVA non-parametric tests were performed. If the experiment was performed over time, linear mixed models were used to account for the repeated measures of an animal using the lme4 R package. Post hoc analyses were then conducted to compare between genotypes within time bins. Post hoc analyses were done using the multcomp R package (173). Animals were removed from analysis if they had a value that was 3.29 standard deviations above the mean or had poor video tracking and could not be analyzed.

3.6 Acknowledgements

This work was supported by 1R01MH107515 JDD, and the Autism Science Foundation, and the National Science Foundation Graduate Research Fellowship DGE-1745038 to NDK. We would like to thank the Genome Technology Access Center for technical support, as well as Dr. Beth Kozel and Dr. Harrison Gabel for critical advice on this project. We would also like to thank Dr. David Wozniak and the Animal Behavior Core at the Washington University School of Medicine for their time and resources.

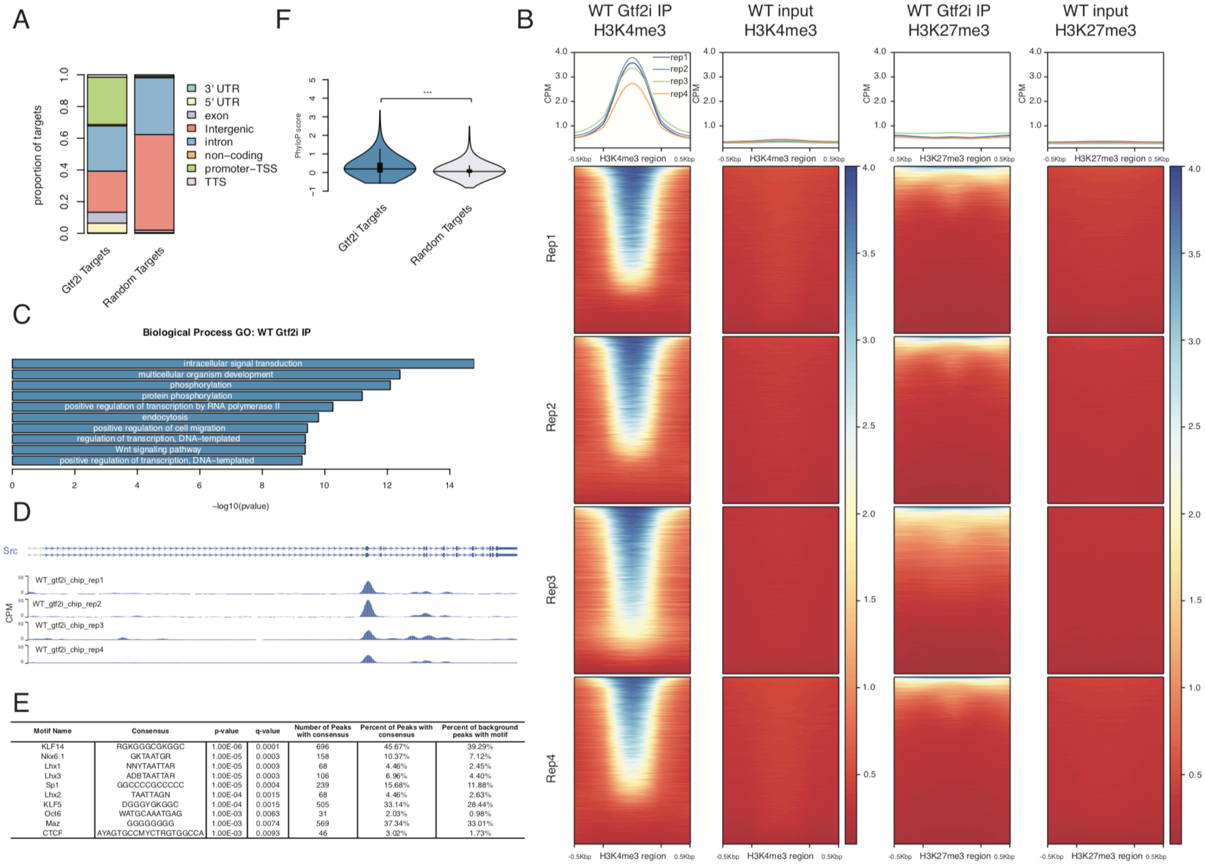
3.7 Figures



D

Motif Name	Consensus	p-value	q-value	Number of Peaks with consensus	Percent of Peaks with consensus	Percent of background peaks with motif
CTCF	AYAGTGCCMYCTRGTGGCCA	1.00E-79	0	161	11.68%	1.70%
GFX	ATTCTCGCGAGA	1.00E-67	0	101	7.32%	0.67%
BORIS	CNNBRGCGCCCCCTGSTGGC	1.00E-56	0	227	16.46%	4.87%
ZBTB33	GGVTCTCGCGAGAAC	1.00E-46	0	132	9.57%	2.04%
GSC	RGGATTAR	1.00E-05	0.0001	64	4.64%	2.45%
Otx2	NYTAATCCYB	1.00E-04	0.0017	47	3.41%	1.79%
ETS	AACCGGAAGT	1.00E-03	0.0185	133	9.64%	7.18%
E2F4	GGCGGGAAAH	1.00E-03	0.0186	274	19.87%	16.45%
GABPA	RACCGGAAGT	1.00E-03	0.0268	209	15.16%	12.24%
ETS1	ACAGGAAGTG	1.00E-02	0.0487	203	14.72%	12.02%

Figure 1: Gtf2ird1 binds preferentially to promoters in conserved, active sites in the genome. **A** Gtf2ird1 binding peaks are annotated primarily in promoters and gene bodies. The distribution of peak annotations is significantly different from random sampling the genome. **B** Gtf2ird1 peaks were enriched in H3K4me3 sites marking active regions of the genome and to a lesser extent in H3K27me3 marking repressed regions. **C** GO analysis of genes that have Gtf2ird1 bound to the promoter. **D** The conservation of sequence in Gtf2ird1 bound peaks is significantly higher than expected by chance. **E** Motifs of transcription factors enriched under Gtf2ird1 bound peaks.



Motif Name	Consensus	p-value	q-value	Number of Peaks with consensus	Percent of Peaks with consensus	Percent of background peaks with motif
KLF14	RGKGGGCGKGGC	1.00E-06	0.0001	696	45.67%	39.29%
Nkx6.1	GKTAATGR	1.00E-05	0.0003	158	10.37%	7.12%
Lhx1	NNYTAATTAR	1.00E-05	0.0003	68	4.46%	2.45%
Lhx3	ADBTAATTAR	1.00E-05	0.0003	106	6.96%	4.40%
Sp1	GGCCCCGCCCCC	1.00E-05	0.0004	239	15.68%	11.88%
Lhx2	TAATTAGN	1.00E-04	0.0015	68	4.46%	2.63%
KLF5	DGGGYGKGGC	1.00E-04	0.0015	505	33.14%	28.44%
Oct6	WATGCAAATGAG	1.00E-03	0.0063	31	2.03%	0.98%
Maz	GGGGGGGG	1.00E-03	0.0074	569	37.34%	33.01%
CTCF	AYAGTGCCMYCTRGTGGCCA	1.00E-03	0.0093	46	3.02%	1.73%

Figure 2: Gtf2i binds at promoters in conserved, active sites in the genome. **A** Gtf2i binding sites are annotated mostly in gene promoters and the gene body. The distribution of peaks is significantly different than would be expected by chance. **B** 78.7% of Gtf2i peaks overlap with H3K4me3 peaks marking active regions. 20.7% of the Gtf2i peaks fall within H3K27me3 peaks marking inactive regions. **C** GO analysis of genes that have Gtf2i bound at the promoter. **D** Epigenome browser shot of Gtf2i peak bound within the Src gene. **E** Genomic sequence under Gtf2i peaks are more conserved than we would expect by chance. **F** Motifs of transcription factors that are enriched in Gtf2i bound sequences.

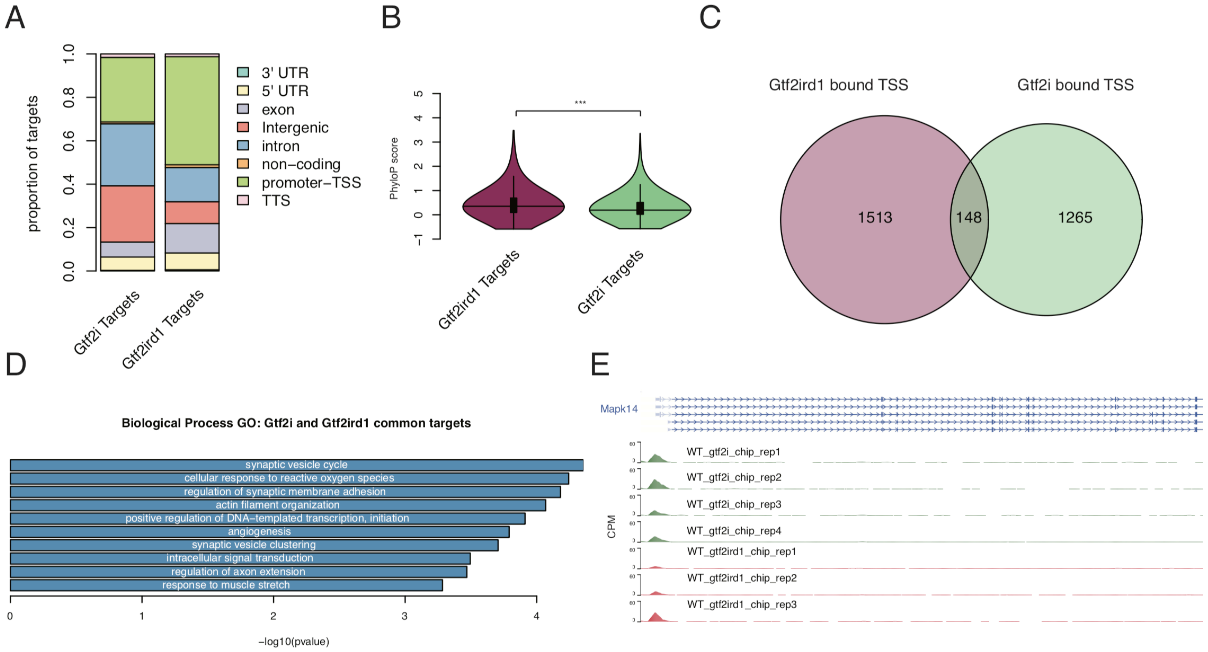


Figure 3: Comparison of Gtf2ird1 and Gtf2i binding sites. **A** Gtf2i and Gtf2ird1 have different distributions of annotated binding sites. **B** Gtf2ird1 bound sequences are more conserved than Gtf2i bound sequences. **C** The overlap of genes that have Gtf2i and Gtf2ird1 bound at their promoters. **D** GO analysis of genes that have both Gtf2i and Gtf2ird1 bound at their promoters. **E** Epigenome browser shot of Mapk14 showing peaks for both Gtf2i and Gtf2ird1.

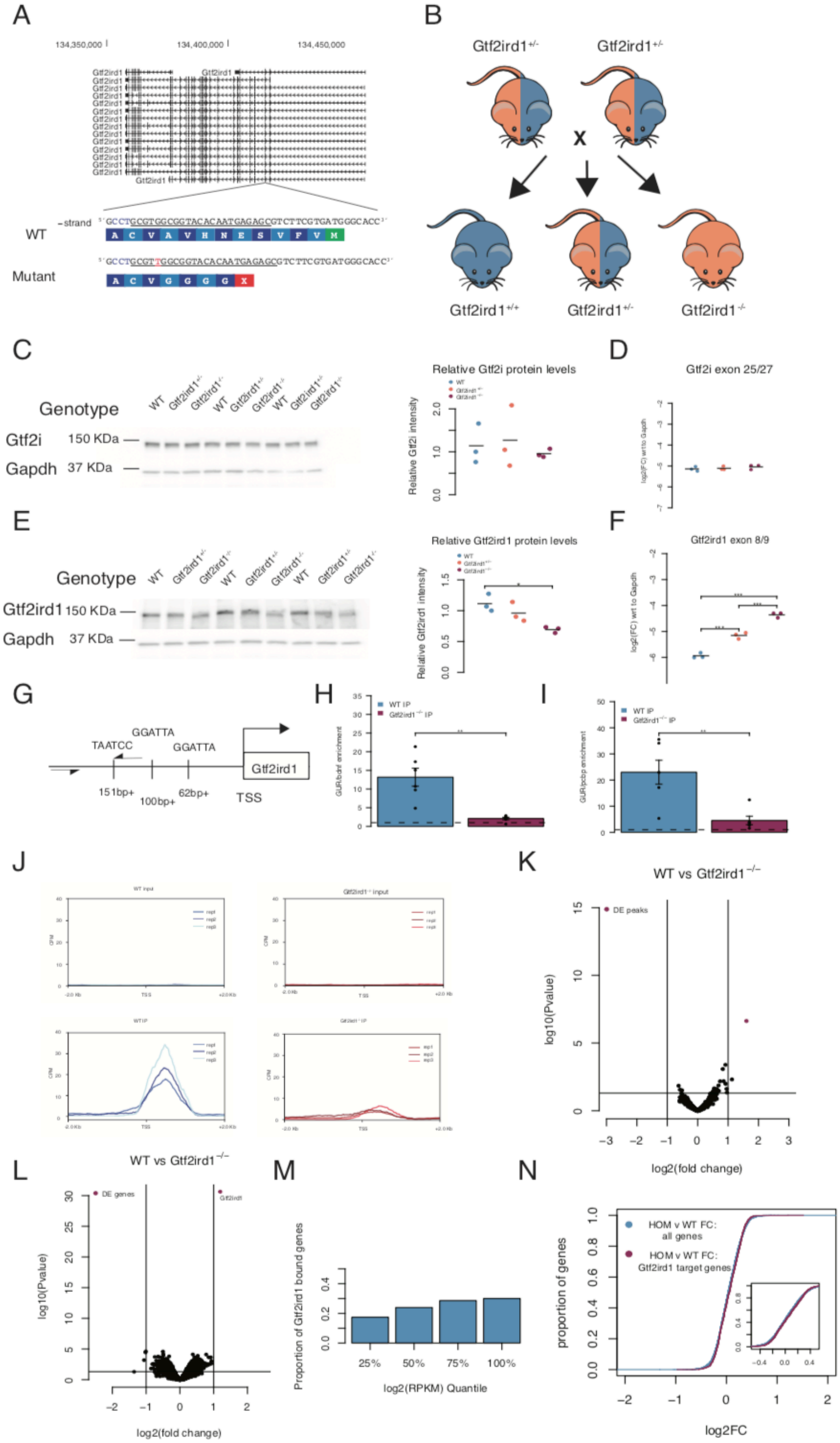

A
134,350,000
134,400,000
134,450,000
Gtf2ird1
-strand
WT
5'GCCTGCGTGGCGGTACACAATGAGAGCGTCTTCGTGATGGGCACC3'
A C V A V H N E S V F V M
Mutant
5'GCCTGCGTTGGCGGTACACAATGAGAGCGTCTTCGTGATGGGCACC3'
A C V G G G G X
B
Gtf2ird1+/-
Gtf2ird1+/-
X
Gtf2ird1+/+
Gtf2ird1+/-
Gtf2ird1-/-
C
Genotype
WT
Gtf2ird1+/-
Gtf2ird1-/-
Gtf2i 150 KDa
Gapdh 37 KDa
Relative Gtf2i protein levels
Relative Gtf2i intensity
D
Gtf2i exon 25/27
log2(FC) wrt to Gapdh
E
Genotype
Gtf2ird1 150 KDa
Gapdh 37 KDa
Relative Gtf2ird1 protein levels
Relative Gtf2ird1 intensity
F
Gtf2ird1 exon 8/9
log2(FC) wrt to Gapdh
G
GGATTA
TAATCC
GGATTA
Gtf2ird1
151bp+
100bp+
62bp+
TSS
H
WT IP
Gtf2ird1-/- IP
GtJR/bdnf enrichment
I
WT IP
Gtf2ird1-/- IP
GtJR/pcbp enrichment
J
WT input
Gtf2ird1-/- input
WT IP
Gtf2ird1-/- IP
TSS
K
WT vs Gtf2ird1-/-
DE peaks
log10(Pvalue)
log2(fold change)
L
WT vs Gtf2ird1-/-
DE genes
Gtf2ird1
log10(Pvalue)
log2(fold change)
M
Proportion of Gtf2ird1 bound genes
25%
50%
75%
100%
log2(RPKM) Quantile
N
HOM v WT FC: all genes
HOM v WT FC: Gtf2ird1 target genes
proportion of genes
log2FC

Figure 4: Frameshift mutation in Gtf2ird1 exon three results in a decreased amount of an N-truncated protein with diminished binding at Gtf2ird1 promoter and has little effect on transcription in the brain. A The sequence of exon three of Gtf2ird1 that was targeted by the gRNA underlined with the PAM sequence in blue. The mutant allele contains a one base pair insertion of an adenine nucleotide that results in a premature stop codon. **B** Breeding scheme of the intercross of *Gtf2ird1*$^{+/-}$ to produce genotypes used in the experiments. **C**, **D** Mutation in *Gtf2ird1* does not affect the protein or transcript levels of *Gtf2i*. **E** Frameshift mutation decreases the amount of protein in *Gtf2ird1*$^{-/-}$ and causes a slight shift to lower molecular weight. **F** The abundance of *Gtf2ird1* transcript increases with increasing dose of the mutation. **G** Schematic of Gtf2ird1 upstream regulatory element (GUR) that shows the three Gtf2ird1 binding motifs. The arrows indicate the location of the primers for amplifying the GUR in the ChIP-qPCR assay. **H,I** WT ChIP of Gtf2ird1 shows enrichment of the GUR over off target regions. There is more enrichment in the WT genotype compared to the *Gtf2ird1*$^{-/-}$ genotype. **J** Profile plots of Gtf2ird1 ChIP-seq data confirms diminished binding at the Gtf2ird1 promoter. **K** Differential peak analysis comparing WT and *Gtf2ird1*$^{-/-}$ ChIP-seq data showed only the peak at *Gtf2ird1* is changed between genotypes with an FDR <0.1. **L** Differential expression analysis in the E13.5 brain comparing WT and *Gtf2ird1*$^{-/-}$ showed only *Gtf2ird1* as changed with FDR < 0.1. **M** The presence of Gtf2ird1 at gene promoters is not evenly distributed across expression levels. **N** The expression of genes bound by Gtf2ird1 is not different compared to all other genes between WT and *Gtf2ird1*$^{-/-}$ mutants.

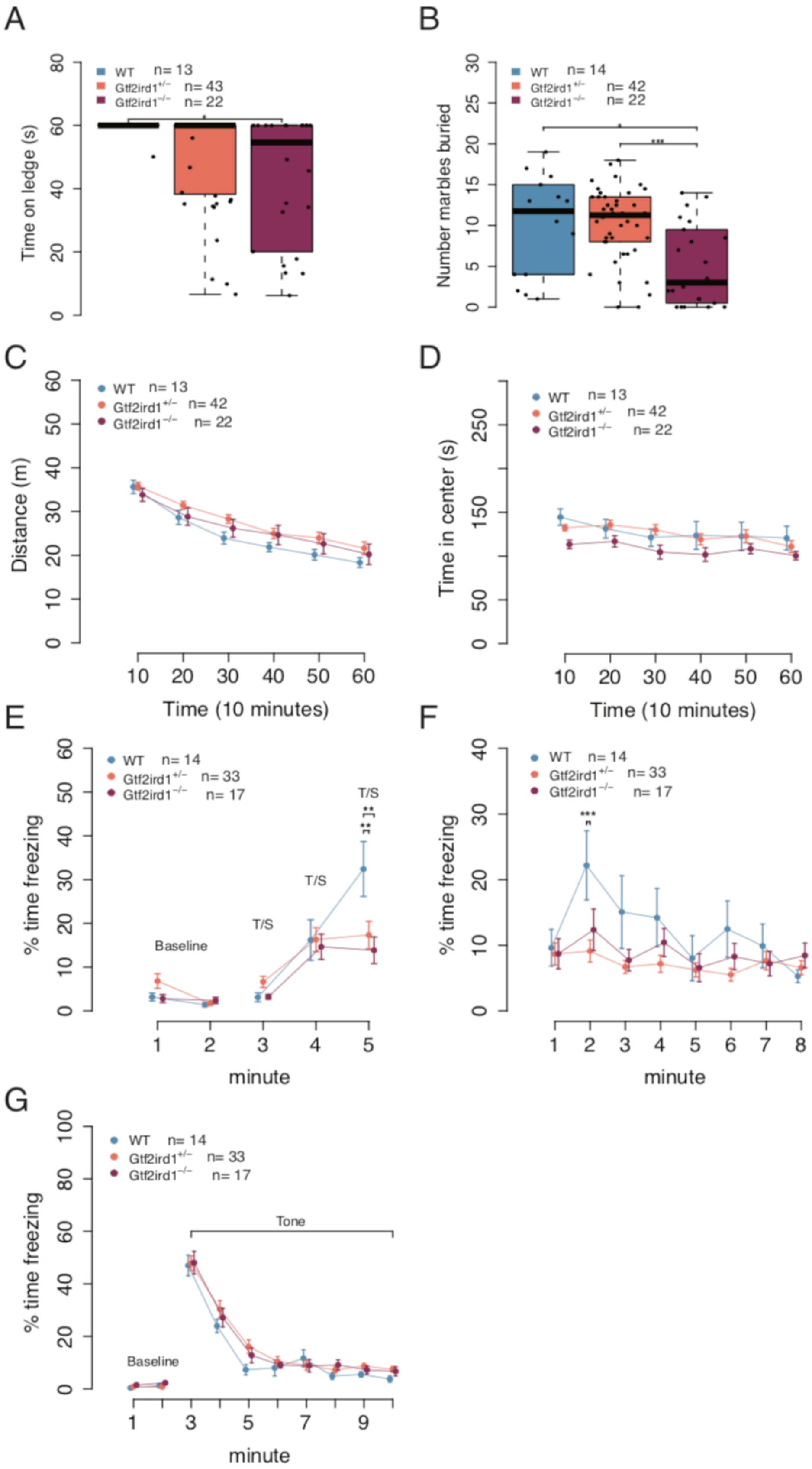

A
Time on ledge (s)
WT n= 13
Gtf2ird1+/- n= 43
Gtf2ird1-/- n= 22
B
Number marbles buried
WT n= 14
Gtf2ird1+/- n= 42
Gtf2ird1-/- n= 22
***
C
Distance (m)
WT n= 13
Gtf2ird1+/- n= 42
Gtf2ird1-/- n= 22
Time (10 minutes)
D
Time in center (s)
WT n= 13
Gtf2ird1+/- n= 42
Gtf2ird1-/- n= 22
Time (10 minutes)
E
% time freezing
WT n= 14
Gtf2ird1+/- n= 33
Gtf2ird1-/- n= 17
Baseline
T/S
T/S
T/S
**
**
minute
F
% time freezing
WT n= 14
Gtf2ird1+/- n= 33
Gtf2ird1-/- n= 17
***
minute
G
% time freezing
WT n= 14
Gtf2ird1+/- n= 33
Gtf2ird1-/- n= 17
Tone
Baseline
minute

Figure 5: Homozygous Frameshift mutation in *Gtf2ird1* is sufficient to cause behavioral phenotypes. A Homozgyous mutants have worse balance than WT littermates in ledge task. **B** Homozygous mutants bury fewer marbles than WT and heterozygous littermates. **C** Overall activity levels are not affected when both sexes are combined. **D** There is no difference in time spent in the center of the apparatus between genotypes. **E** Acquisition phase of fear condition paradigm. WT animals freeze more during the last five minutes of the task. **F** WT animals showed greater freezing in contextual fear memory task than *Gtf2ird1*$^{+/-}$. **G** There were no differences between genotypes in cued fear.

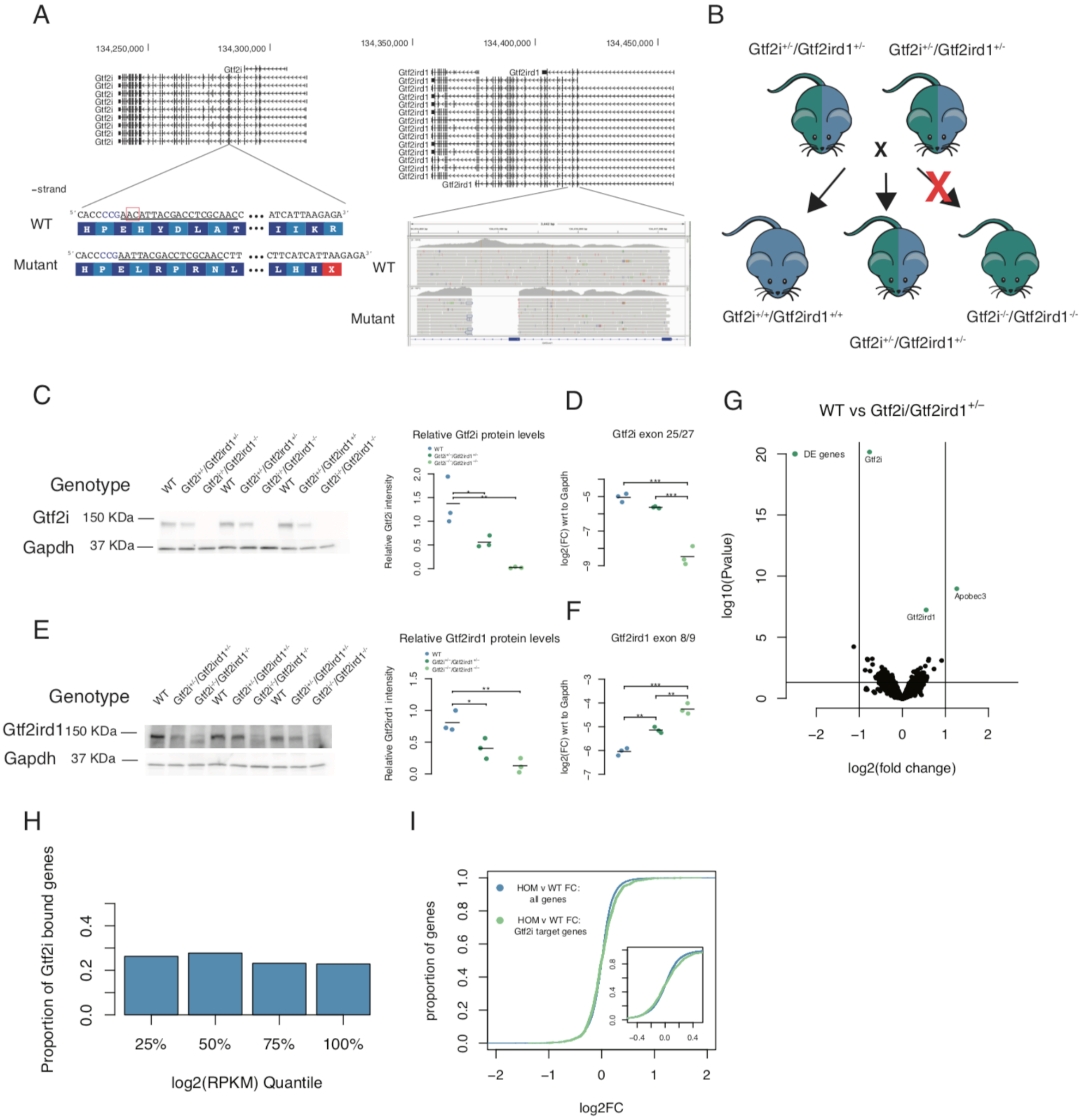


Figure 6: Mutating both *Gtf2i* and *Gtf2ird1* does not result in larger differences in brain transcriptomes. A Generation of double mutant. gRNA target is underlined in exon five of *Gtf2i* with the PAM sequence in blue. The two base pair deletion results in a premature stop codon within exon five. The *Gtf2ird1* mutation is a large 590 base pair deletion covering most of exon three as shown in the IGV browser shot. **B** Heterozygous intercross to generate

genotypes for ChIP and RNA-seq experiments. The homozygous double mutants are embryonic lethal but are present up to E15.5. **C** The two base pair deletion in *Gtf2i* decreases the protein by 50% in heterozygous mutant and no protein is detected in the homozygous E13.5 brain. **D** The mutation decreases the abundance of *Gtf2i* transcript consistent with nonsense-mediated decay. **E** The 590 base pair deletion in Gtf2ird1 leads to decrease protein levels in heterozygous and homozygous mutants. There is still a small amount of protein made in the homozygous mutant. **F** The 590 base pair deletion increases the amount of *Gtf2ird1* transcript. **G** Volcano plot comparing the expression in the E13.5 brain of WT and heterozygous double mutants. The highlighted genes represent an FDR < 0.1. **H** The presence of *Gtf2i* at the promoters does not correlate with the expression of a gene. **I** The fold change of genes between WT and homozygous double mutants that have Gtf2i bound at their promoters were slightly upregulated when compared to the fold change of genes that did not have Gtf2i bound.

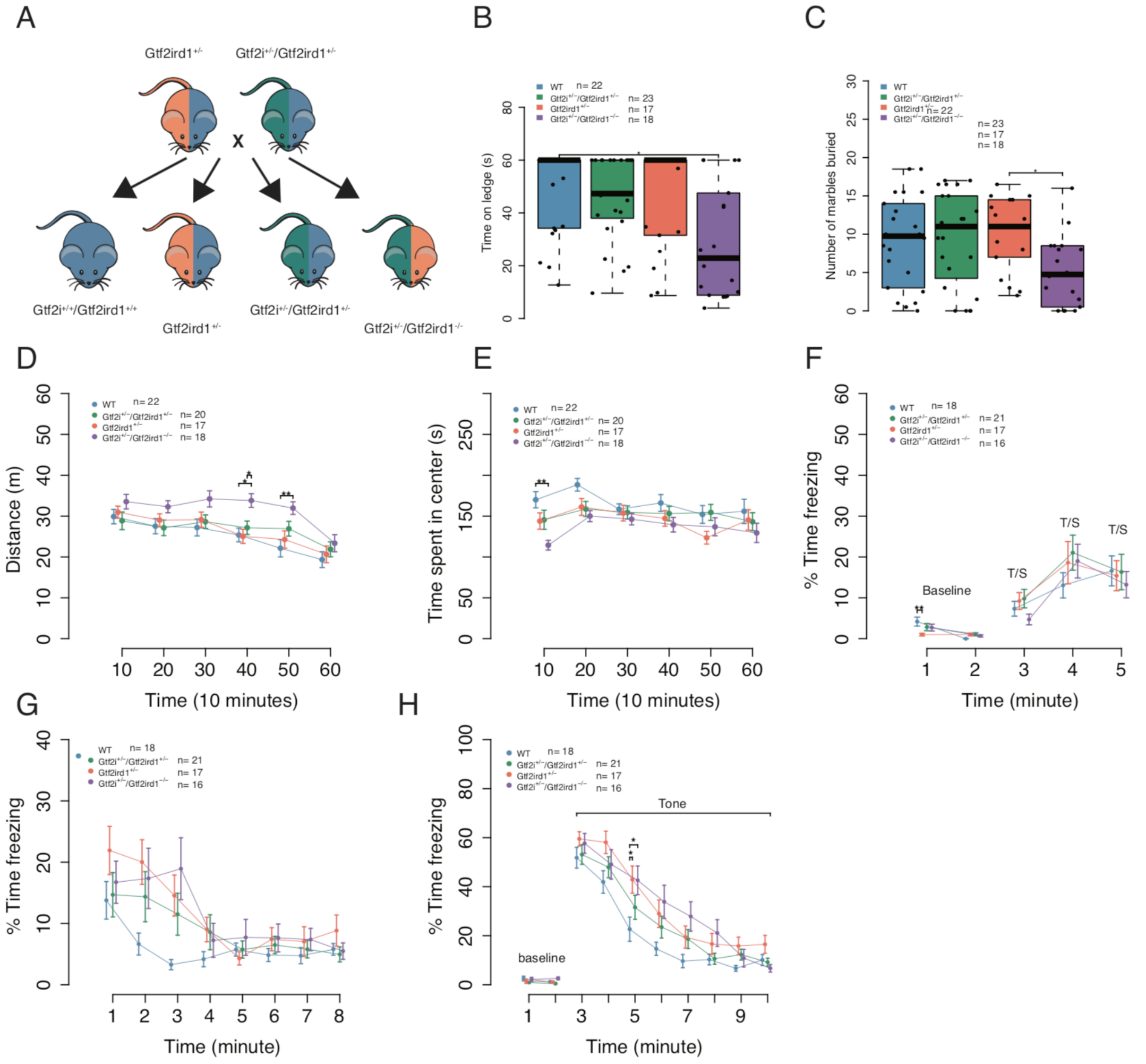
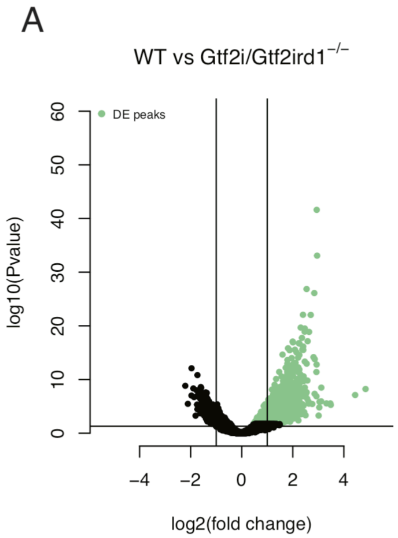
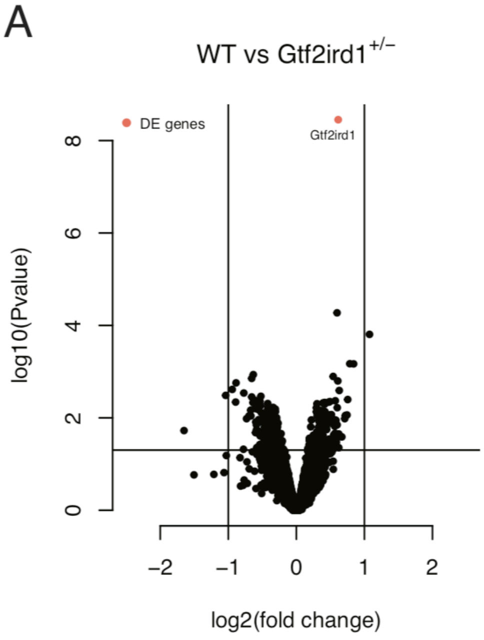


Figure 7: Gtf2i does not modify the phenotype of Gtf2ird1 mutation. A Breeding scheme for behavior experiments. **B** The *Gtf2i*$^{+/-}$/*Gtf2ird1*$^{-/-}$ animals fell off ledge sooner than WT littermates. **C** There was main effect of genotype on marbles buried. Post hoc analysis showed that the *Gtf2i*$^{+/-}$/*Gtf2ird1*$^{-/-}$ buried fewer marbles than the *Gtf2ird1*$^{-/-}$ genotype. **D** The *Gtf2i*$^{+/-}$/*Gtf2ird1*$^{-/-}$ had increased overall activity levels in a one hour activity task. **E** The *Gtf2i*$^{+/-}$/*Gtf2ird1*$^{-/-}$ showed decreased time in the center of the apparatus compared to WT, with the *Gtf2ird1*$^{+/-}$

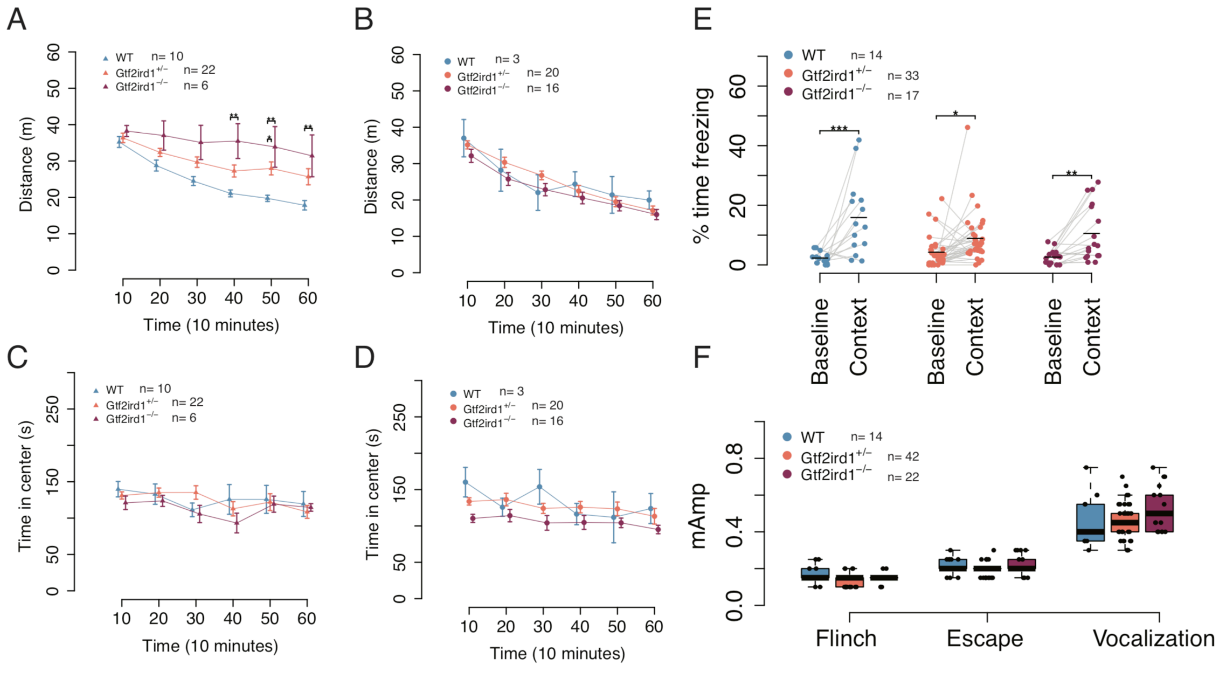
and *Gtf2i*$^{+/-}$/*Gtf2ird1*$^{+/-}$ having intermediate values. **F** All genotypes showed increased freezing with increased number of footshocks. **G** All genotypes showed a similar contextual fear response. **H** There was a main effect of genotype on cued fear with the *Gtf2ird1*$^{+/-}$ and *Gtf2i*$^{+/-}$/*Gtf2ird1*$^{-/-}$ showing an increased fear response compared to WT.



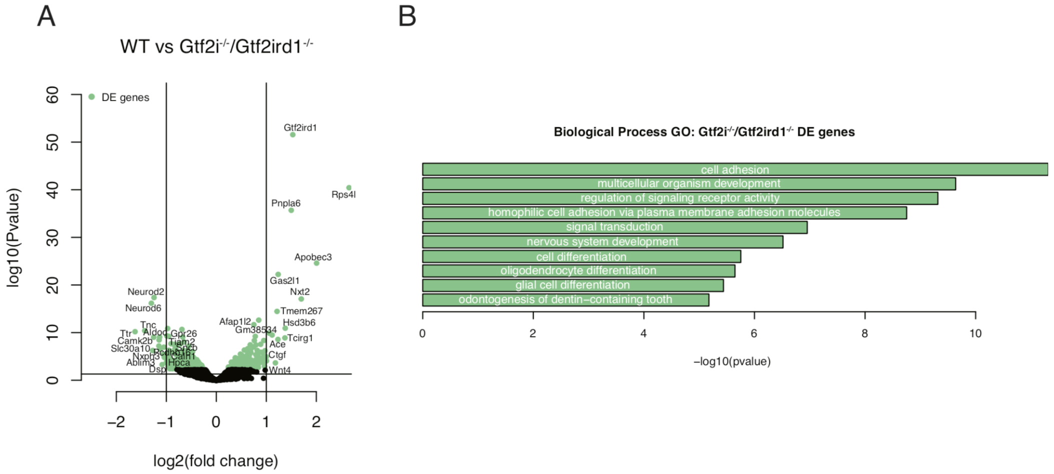
Supplemental Figure 1: Differential peak binding comparing the WT and homozygous Gtf2i IP. A The highlighted peaks have an FDR < 0.1 and a log2FC > 0. These were used as the high confidence Gtf2i peaks.



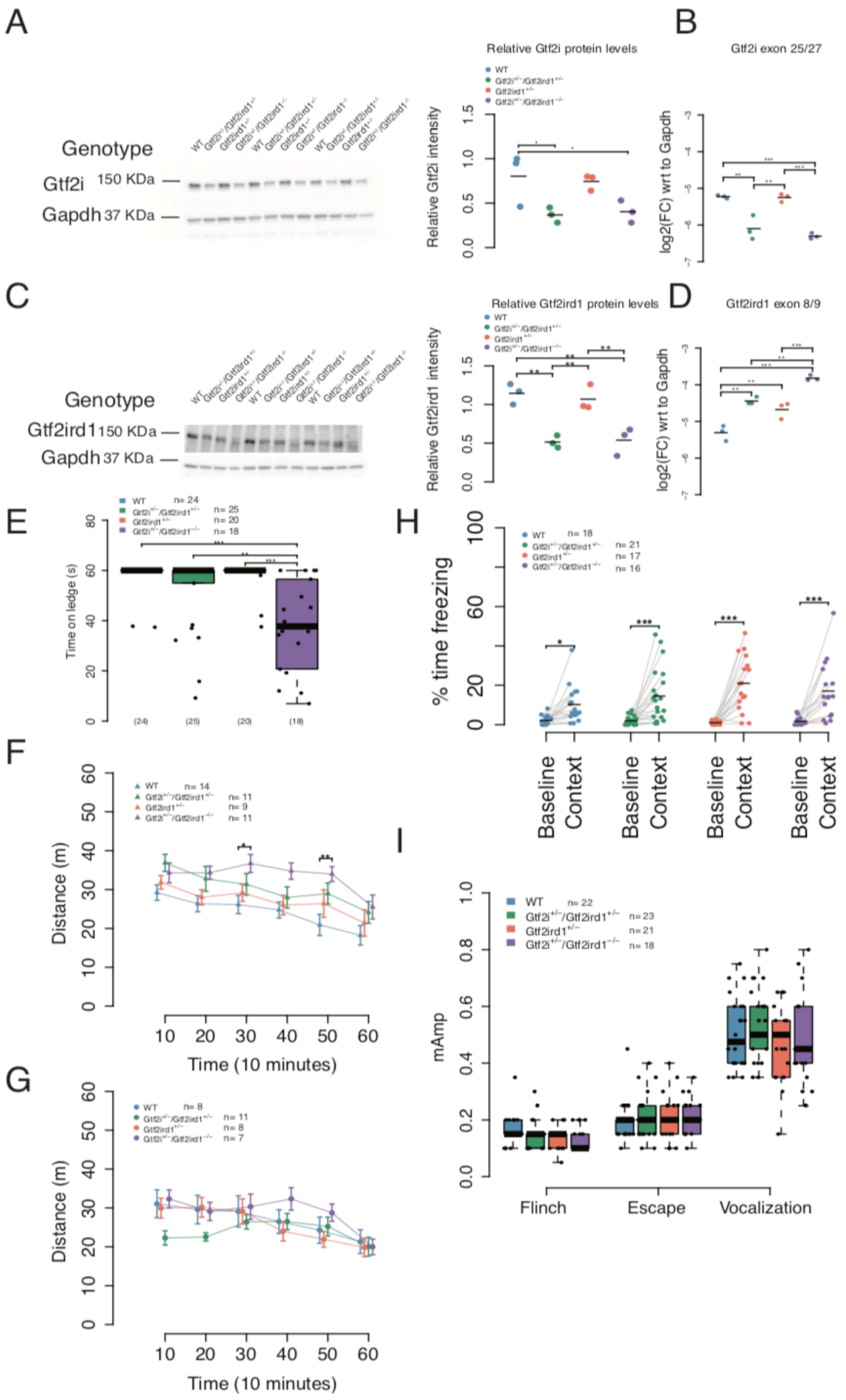
Supplemental Figure 2: RNA-seq analysis of E13.5 brain comparing the WT and *Gtf2ird1*$^{+/-}$ mutants. A Only *Gtf2ird1* showed a difference with FDR < 0.1.



Supplemental Figure 3: The effects of frameshift mutation in *Gtf2ird1*. A Female heterozygous and homozygous mutants have increased activity levels. **B** There is no difference in activity levels in male mice. **C** There is no difference between genotypes in females with respect to the time spent in the center of the apparatus. **D** There is no difference between genotypes in males with respect to the time spent in the center of the apparatus. **E** All genotypes showed a contextual fear response. Baseline refers to the first two minutes of the task on day one and context refers to the first two minutes of the task on day two. **F** There was no difference in shock sensitivity between genotypes.



Supplemental Figure 4: RNA-seq analysis of homozygous double mutant. A The homozygous double mutant showed significant changes (FDR < 0.1, highlighted in green) across many genes. Genes are labeled that had an FDR < 0.1 and a log2FC > 1 or log2FC < -1. **B** GO analysis of all nominally significant genes.



A
Genotype
WT Gtf2i+/-/Gtf2ird1+/- Gtf2ird1+/- Gtf2i+/-/Gtf2ird1-/- WT Gtf2i+/-/Gtf2ird1+/- Gtf2ird1+/- Gtf2i+/-/Gtf2ird1-/- WT Gtf2i+/-/Gtf2ird1+/- Gtf2ird1+/- Gtf2i+/-/Gtf2ird1-/-
Gtf2i 150 KDa
Gapdh 37 KDa
Relative Gtf2i protein levels
WT
Gtf2i+/-/Gtf2ird1+/-
Gtf2ird1+/-
Gtf2i+/-/Gtf2ird1-/-
Relative Gtf2i intensity
0.0 0.5 1.0 1.5
B
Gtf2i exon 25/27
log2(FC) wrt to Gapdh
-3 -4 -5 -6 -7
C
Genotype
Gtf2ird1 150 KDa
Gapdh 37 KDa
Relative Gtf2ird1 protein levels
Relative Gtf2ird1 intensity
D
Gtf2ird1 exon 8/9
E
WT n= 24
Gtf2i+/-/Gtf2ird1+/- n= 25
Gtf2ird1+/- n= 20
Gtf2i+/-/Gtf2ird1-/- n= 18
Time on ledge (s)
0 20 40 60 80
(24) (25) (20) (18)
F
WT n= 14
Gtf2i+/-/Gtf2ird1+/- n= 11
Gtf2ird1+/- n= 9
Gtf2i+/-/Gtf2ird1-/- n= 11
Distance (m)
Time (10 minutes)
10 20 30 40 50 60
G
WT n= 8
Gtf2i+/-/Gtf2ird1+/- n= 11
Gtf2ird1+/- n= 8
Gtf2i+/-/Gtf2ird1-/- n= 7
Distance (m)
Time (10 minutes)
H
WT n= 18
Gtf2i+/-/Gtf2ird1+/- n= 21
Gtf2ird1+/- n= 17
Gtf2i+/-/Gtf2ird1-/- n= 16
% time freezing
0 20 60 100
Baseline Context Baseline Context Baseline Context Baseline Context
I
WT n= 22
Gtf2i+/-/Gtf2ird1+/- n= 23
Gtf2ird1+/- n= 21
Gtf2i+/-/Gtf2ird1-/- n= 18
mAmp
0.0 0.2 0.4 0.6 0.8 1.0
Flinch Escape Vocalization

Supplemental Figure 5: Biochemical and behavioral characterization of the *Gtf2ird1*$^{+/-}$ x *Gtf2i*$^{+/-}$/*Gtf2ird1*$^{+/-}$. A, B Western blot and qPCR confirm decrease in Gtf2i protein and mRNA. **C** Western blot shows that the large Gtf2ird1 deletion decreases the protein, but adding the one base pair insertion mutation does not further decrease the protein made. **D** Gtf2ird1 mutation increases mRNA abundance. **E** Replication of ledge task in independent cohort. **F** Gtf2i+/-/Gtf2ird1-/- females have increased activity levels. **G** Gtf2i+/-/Gtf2ird1-/- males to a lesser extent have increased. **H** All genotypes showed a contextual fear memory response. **I** There is no difference between genotypes in shock sensitivity.

Table 1: Behavior and animal cohort for *Gtf2ird1*$^{+/-}$ x *Gtf2ird1*$^{+/-}$

	Male			**Female**		
Behavior	**WT**	***Gtf2ird1*$^{+/-}$**	***Gtf2ird1*$^{-/-}$**	**WT**	***Gtf2ird1*$^{+/-}$**	***Gtf2ird1*$^{-/-}$**
One hour activity	3	20	16	10	22	6
Ledge	2	20	16	11	23	6
Marble burying	3	20	16	11	22	6
Condtioned Fear	3	16	12	11	17	5
Shock Sensitivity	3	20	16	11	22	6

Table 2: Behavior and animal cohorts for the *Gtf2ird1*$^{+/-}$ x *Gtf2i*$^{+/-}$/*Gtf2ird1*$^{+/-}$

Behavior	Male				Female			
Cohort1	WT	*Gtf2ird1*$^{+/-}$	*Gtf2i*$^{+/-}$/*Gtf2ird1*$^{+/-}$	*Gtf2i*$^{+/-}$/*Gtf2ird1*$^{-/-}$	WT	*Gtf2ird1*$^{+/-}$	*Gtf2i*$^{+/-}$/*Gtf2ird1*$^{+/-}$	*Gtf2i*$^{+/-}$/*Gtf2ird1*$^{-/-}$
One hour activity	8	8	11	7	14	9	9	11
Ledge	8	8	12	7	14	9	11	11
Marble burying	8	8	12	7	14	9	11	11
Cohort2								
Ledge	13	5	12	6	11	15	13	12
Condtioned Fear	11	4	9	5	7	13	12	11
Shock Sensitivity	12	6	10	6	10	15	13	12

Supplemental Table S1: Table of summary statistics and statistical tests

Chapter 4: *Gtf2i* and *Gtf2ird1* mutation are not sufficient to reproduce mouse phenotypes caused by the Williams syndrome critical region

Nathan Kopp, Katherine McCullough, Susan E. Maloney, and Joseph D. Dougherty

From a manuscript submitted and is in review at *Human Molecular Genetics* as:

Kopp N., McCullough, K., Maloney, S.E., and Dougherty, J.D. *Gtf2i* and *Gtf2ird1* mutation are not sufficient to reproduce mouse phenotypes caused by the Williams syndrome critical region (2019) *In review.*

4.1 Abstract

Williams syndrome is a neurodevelopmental disorder caused by a 1.5-1.8Mbp deletion on chromosome 7q11.23, affecting the copy number of 26-28 genes. Phenotypes of Williams syndrome include cardiovascular problems, craniofacial dysmorphology, deficits in visual-spatial cognition, and a characteristic hypersocial personality. There are still no genes in the region that have been consistently linked to the cognitive and behavioral phenotypes, although human studies and mouse models have led to the current hypothesis that the general transcription factor 2 I family of genes, *GTF2I* and *GTF2IRD1*, are responsible. Here we test the hypothesis that these two transcription factors are sufficient to reproduce the phenotypes that are caused by deletion of the Williams syndrome critical region (WSCR). We compare a new mouse model with loss of function mutations in both *Gtf2i* and *Gtf2ird1* to an established mouse model lacking the complete WSCR. We show that the complete deletion model has deficits across several behavioral domains including social communication, motor functioning, and conditioned fear that are not explained by loss of function mutations in *Gtf2i* and *Gtf2ird1*. Furthermore, transcriptome profiling of the hippocampus shows changes in synaptic genes in the complete deletion model that are not seen in the double mutants. Thus, we have thoroughly defined a set of molecular and behavioral consequences of complete WSCR deletion, and shown that genes or combinations of genes beyond *Gtf2i* and *Gtf2ird1* are necessary to produce these phenotypic effects.

4.2 Introduction

Contiguous gene disorders provide a unique opportunity to understand genetic contributions to human biology, as their well-defined genetic etiology delimits specific genomic regions strongly affecting particular phenotypes. Williams syndrome (WS; OMIM #194050) is caused by a 1.5-1.8Mbp deletion of 26-28 genes on chromosome 7q11.23 in the Williams syndrome critical region (WSCR). Williams syndrome is phenotypically characterized by supravalvular aortic stenosis, craniofacial dysmorphology, and a distinct cognitive profile consisting of intellectual disability, severe visual-spatial deficits, yet relatively strong language skills. Other common cognitive and behavioral difficulties include high levels of anxiety, specific phobias, and a characteristic hypersocial personality manifested as strong eye contact, indiscriminate social approach, and social disinhibition (see (2, 14, 15) for reviews). Despite increased social interest, individuals with Williams syndrome have difficulties with social awareness and social cognition (20, 174). In contrast, the reciprocal duplication results in dup7q11.23 syndrome (OMIM #609757), which presents with both similar and contrasting phenotypes to WS, such as high levels of anxiety yet less social interest (175). It is also associated with autism spectrum disorders (121). The recurrent deletion and duplications of chr7q11.23 indicate that one or more genes in this region are dose sensitive and have a large effect on human cognition as well as human social behavior.

Substantial efforts have been taken to understand which genes in the WSCR contribute to different aspects of the phenotype. Three approaches have driven advances in genotype-phenotype correlations in the WSCR: phenotyping individuals with atypical deletions in the region, human induced pluripotent stem cell models, and mouse models. Patients with atypical deletions have firmly connected haploinsufficiency of the elastin (*ELN)* gene with supravalvular aortic stenosis and other elastic tissue difficulties in WS (6, 104). However, human studies have not conclusively linked other genes to specific phenotypes. Three atypical deletions that span the *ELN* gene to the typical telomeric breakpoints showed the full spectrum of the WS phenotype, suggesting that most of the phenotypes are driven by the telomeric end of the deletion, which contains genes for two paralogous transcription factors *GTF2I* and *GTF2IRD1* (34, 35). Indeed, most of the atypical deletions that have been reported that delete the centromeric end of the region and don't affect the copy number of *GTF2I* and GTF2IRD1, show mild to none of the characteristic facial features or cognitive and behavioral phenotypes of WS (31–33, 36–40, 99). While there are contrasting examples of deletions that spare *GTF2I* and still have mild facial characteristics of WS, lower IQ, and the overfriendly social phenotype (40, 176), the preponderance of evidence from these rare partial deletions have led to the dominant hypothesis being that *GTF2I* and *GTF2IRD1* mutation are necessary to cause the full extent of the social, craniofacial, visual-spatial and anxiety phenotypes. However, there are limitations to these human studies, primarily due to the rarity of partial deletions. First, because of the variable expressivity of the phenotypes even in typical WS, it can be difficult to confidently interpret any phenotypic deviation in the rare partial deletions (20, 56, 174). Second, given the rarity of WS and partial deletions, and lack of relevant primary tissue samples, it is challenging to link genetic alterations to the specific downstream molecular and cellular changes that could mediate the organismal phenotypes.

To overcome this second barrier, researchers have turned to using patient induced pluripotent stem cells to study the effects of the WSCR deletion and duplication on different disease relevant cell types (44, 45, 47–49). While linking molecular changes to organismal behavior is not possible with cell lines, this approach is amenable to studying cellular and molecular phenotypes, such as changes to the transcriptome and cellular physiology. By studying differentiated neural precursor cells from an individual with a typical WS deletion and an individual with an atypical deletion that spares the copy number of the *FZD9* gene, Chailangkarn et al. (45) showed that *FZD9* is responsible for some of the cellular phenotypes, such as increased apoptosis and morphological changes. Lalli et al. (49) used a similar approach to show that knocking down the *BAZ1B* gene in differentiated neurons was sufficient to reproduce the transcriptional differences and deficits in differentiation that were observed in WS differentiated neurons. Finally, Adamo et al. (44) studied the effects of *GTF2I* on iPSCs from typical WS deletions, dup7q11.23, and typical controls. By overexpressing and knocking down *GTF2I* in the three genotypes, they showed that *GTF2I* was responsible for 10-20% of the transcriptional changes. Overall, using iPSCs from patients with WS has highlighted a role for both the *GTF2I* family and other less appreciated genes in the molecular consequences of the WSCR mutation. This suggested the possibility that several genes may play a role in the cognitive phenotypes and *GTF2I* alone may not be sufficient for all neural molecular changes and hence cognitive phenotypes. However, iPSC studies face the limitation that they cannot be used to model whole organismal effects like anxiety or social behavior. Further, while some cellular and molecular phenotypes can be evaluated, both gene expression and cellular physiology using *in vitro* differentiation systems do not perfectly reflect the phenotype of mature neural cells, fully integrated into a functioning or dysfunctioning brain.

Mouse models have been used to link genes in WSCR to specific molecular and cellular phenotypes, as well as to the functioning of conserved organismal behavioral circuits that could be related to human cognitive phenotypes. Mouse models are particularly suitable because a region on mouse chromosome five is syntenic to the WSCR, enabling models of corresponding large deletions, including a mouse line with a complete deletion (CD) of the WSCR genes that shows both behavioral disruptions and altered neuronal morphology (93). In addition, a key advantage over human partial deletions is that researchers can easily manipulate the mouse genome to delete targeted subsets of genes in the locus, and generate large numbers of animals with identical partial mutations, enabling statistical analyses to overcome variable expressivity. For example, there are mouse models of large deletions that show that genes in the distal and proximal half of the region may contribute to separate and overlapping phenotypes (94). Likewise, many single gene knockouts exist that show some phenotypic similarities to the human syndrome, though a limitation is that some of these studies model full homozygous loss of function, rather than a hemizygous decrease in gene dose. Nonetheless, specifically for *Gtf2ird1* (92, 100, 101) and *Gtf2i* (29, 67, 96), multiple mouse models of either gene have shown extensive behavioral deficits including social and anxiety-like behaviors, some of which present contrasting evidence. However, each of these studies has been conducted in isolation, by different labs, with fairly different phenotyping assays, making it difficult to directly compare findings to other mouse models of WS.

Mouse models uniquely enable a direct way to test the sufficiency of individual mutations to recreate the organismal phenotypes detected when the entirety of the WSCR is deleted. By crossing different mutant lines together, we can create genotypes unavailable in human studies and conduct a well-powered and controlled study to directly test if specific gene mutations are sufficient to reproduce particular phenotypes of the full deletion. Since both human and mouse literature suggest that *GTF2IRD1* and *GTF2I* each contribute to the molecular, cognitive, and social phenotypes, we set out here to test if loss of function of both of these genes is sufficient to recapitulate the phenotypes of the entire WSCR deletion at both the molecular and behavioral circuit levels, or if instead, as hinted by the iPSC studies and other human mutations, other or more genes may be involved. Using CRISPR/Cas9 we generated a new mouse line that has loss of function mutations in both *Gtf2i* and *Gtf2ird1* on the same chromosome. We then crossed them to the CD full deletion model to directly compare behavior and transcriptomes of the *Gtf2i*/*Gtf2ird1* mutants to both WT and CD littermates. Examining both previously defined and newly characterized behavioral and molecular disruptions, we demonstrate that mutation of these two genes is not sufficient to replicate *any* of the CD phenotypes. In contrast to a dominant hypothesis arising from human partial deletions, this study provides strong evidence that *Gtf2i*/*Gtf2ird1* mutation alone may not be responsible for key WS cognitive and behavioral phenotypes.

4.3 Results

4.3.1 Generation and validation of *Gtf2i* and *Gtf2ird1* loss of function mutation on the same chromosome.

Prior work from comparing phenotypes of humans with partial deletions of the WSCR highlighted *GTF2I* and *GTF2IRD1* as likely involved in cognitive phenotypes in WS (34, 38, 39). Likewise, single gene mutant mouse models of both genes showed that each may contribute to relevant phenotypes (92, 96, 97, 100, 101). We wanted to test if heterozygous loss of function mutants of both *Gtf2i* and *Gtf2ird1* are sufficient to replicate the phenotypes that are caused when animals are hemizygous for the entire WSCR (**Figure 1A**).

Therefore, to test the sufficiency of these genes, we generated a mutant of *Gtf2i* and *Gtf2ird1* genes on the same chromosome using CRIPSR/Cas9. Two gRNAs were designed to target constitutive exons of *Gtf2i* or *Gtf2ird1* (**Figure 1B**) and were co-injected with Cas9 mRNA into the eggs of the FVB strain. Of the 57 pups born we detected 21 editing events using the T7 endonuclease assay. From these animals PCR amplicons around each targeted site were deeply sequenced and mutations were characterized via manual inspection of the reads in IGV. Of the founders there were five that only had mutations in *Gtf2i*, five with mutations only in *Gtf2ird1*, and 15 that had mutations in both genes (**Supplemental Figure 1A**). Most founders had more than one allele within a gene indicating high rates of mosaicism (60%, 15/25 mice). Breeding a selection of the mosaic founders to WT animals revealed that some of the founders were mosaic in the germline as well (40%, 4/10 mice), with one founder transmitting three different alleles.

To test if haploinsufficiency of both *Gtf2i* and *Gtf2ird1* is sufficient to replicate the phenotype of hemizygosity of the entire WSCR, we moved forward with characterizing a mouse line that has a G > C polymorphism followed by an eight base pair insertion in exon five of *Gtf2i* and a five base pair deletion in exon three of *Gtf2ird1*; these will be referred to as the *Gtf2i** mouse line (**Figure 1B**). These mutations are inherited together indicating that they are on the same chromosome. The mutations cause frameshifts and introduce premature stop codons in early constitutive exons (**Figure 1B**), and were thus expected to trigger nonsense mediated decay and lead to loss-of-function alleles, mimicking the effective gene dosage of WSCR region deletions for these two genes.

We first performed RT-qPCR and western blots to confirm the effects of the frameshift mutations at the transcript and protein levels in embryonic day 13.5 (E13.5) littermates that were WT, heterozygous, and homozygous mutant at the locus. We used E13.5 brains for two reasons 1) homozygosity of *Gtf2i* null mutants is embryonic lethal (87, 96) and 2) both Gtf2i and Gtf2ird1 proteins are more highly expressed during embryonic time points in the brain, with undetectable levels of Gtf2ird1 in the WT adult mouse brain (**Supplemental Figure 1B and C**).

The frameshift mutation in exon five of *Gtf2i* reduced the amount of transcript detected by qPCR, consistent with nonsense mediated decay. This mutation led to a 50% decrease of the protein in heterozygous animals and no protein in homozygous mutants (**Supplemental Figure**

1D). Indeed we were not able to recover pups that were homozygous for the *Gtf2i** mutations after birth, but we were able to harvest homozygous embryos up to E15.5. The embryos had exencephaly consistent with other *Gtf2i* mouse models (87, 96).

In contrast, the frameshift mutations in exon three of *Gtf2ird1* increased the amount of transcript, as expected. Increases in transcript of *Gtf2ird1* due to a loss of function mutation have been described in another *Gtf2ird1* mouse model, and both EMSA and luciferase reporter assays indicated that Gtf2ird1 protein represses the transcription of the *Gtf2ird1* gene (66). The increase in transcript was commensurate with the dosage of the mutation (**Supplemental Figure 1E**). However, we saw that the protein levels in our mutants did not change with dosage of the mutation and did not follow the trend of the transcript (**Supplemental Figure 1E**).

Production of detectable protein after the frameshift was surprising, especially since the increased *Gtf2ird1* mRNA levels were indeed consistent with prior studies of loss of functional Gtf2ird1 protein, so we investigated this phenomena further. We noticed that the homozygous Gtf2ird1 protein bands looked slightly shifted in the western blots. This lead us to hypothesize that there could be a translation reinitiation event at the methionine in exon three downstream of the frameshift mutation in a different open reading frame (**Supplemental Figure 1F**). In another targeted mutation of *Gtf2ird1*, where the entire exon two, which contains the conical start codon, was removed, the authors noted that there was still three percent of protein being made, and the product that was made was similarly shifted (66). From our mutation we would expect a 65aa N-terminal truncation, which corresponds to a 7KDa difference between WT. We ran a lower percentage PAGE gel to get better separation between WT and homozygous animals and we saw a slight shift, suggesting that there was reinitiation of translation at methionine-65 in a different open reading frame (**Supplemental Figure 1G**). This was indicative of the loss of the N-terminal end of the protein, which contains a leucine zipper that is thought to be important in DNA binding (66). This is consistent with the mRNA evidence that the allele is loss of function.

We therefore tested the hypothesis that we had abolished the DNA binding capacity of the truncated protein, to confirm loss of function. We performed ChIP-qPCR and pulled down DNA bound to Gtf2ird1 protein and then amplified the promoter region of *Gtf2ird1,* which has previously been shown to be bound by the Gtf2ird1 protein. We compared this to two off-target regions in the genome near *Bdnf* and *Pcbp3*. We performed this experiment in E13.5 brains of WT and homozygous *Gtf2i** embryos. There was a 15-20 fold enrichment of the on target *Gtf2ird1* promoter region compared to the off target regions in the WT animals, while the truncated protein did not show any enrichment (**Supplemental Figure 1H,I**). This suggested that while a truncated protein was still being made it did not have the DNA binding functionality of the WT protein. This indicated that the frameshift mutation in exon three of *Gtf2ird1* was a loss-of-function mutation and provided evidence that the N-terminal end of the protein, which contains a leucine zipper, is necessary for DNA binding. Thus, we confirmed we had generated a mouse line with loss of function alleles on the same chromosome for these *Gtf2i** genes.

To test the sufficiency of mutation in these two transcription factors to replicate phenotypes observed by deleting the entire WSCR, we crossed the *Gtf2i** mutant to the CD mouse (**Figure 1C**), which is hemizygous from exon five of *Gtf2i* to *Fkbp6* (**Figure 1A**). The *Gtf2i** mutants were generated on the FVB/AntJ background, whereas the CD mice were generated on the C57BL/6J background. Therefore, we only used the first generation from this cross for all experiments to ensure all mice had the same genetic background. As above, we assessed the transcript and protein levels of genotypes from this cross to confirm loss of function. Again, the CD/*Gtf2i** genotype was embryonic lethal, but we did observe that genotype up to

E15.5. The levels of *Gtf2i* transcript and protein were similar between CD heterozygous and *Gtf2i** heterozygous animals (**Figure 1D**). The levels of *Gtf2ird1* transcript increased in *Gtf2i** animals similar to what was seen in *Gtf2i** heterozygous animals on the pure FVB/AntJ background. In contrast, the CD heterozygous animals had decreased levels of *Gtf2ird1* transcript. In the CD/*Gtf2i** animals the level of transcript returned to WT levels. Again, the levels of *Gtf2ird1* transcript were not reflected in the protein levels. We saw a trend to similar slight decreases in protein levels in the both heterozygous genotypes; however, they were not significantly different from WT levels. This was interesting because in the CD animals were missing one entire copy of this gene, opposed to a frameshift mutation. This also suggested that the frameshift mutation in exon three of *Gtf2ird1* did affect the amount of protein being made, but not drastically. We did see a significant decrease in protein levels (60% of WT) in the CD/*Gtf2i** genotype (**Figure 1E**). Again suggesting that the frameshift mutation was decreasing the levels of protein.

4.3.2 Gtf2i* mutation is not sufficient to reproduce WSCR-mediated alterations of vocal communication

We next tested if haploinsufficiency for both genes would recapitulate behavioral phenotypes seen in mice hemizygous for the entire WSCR (CD mice) (**Table 1**). Since single gene knockout studies of both *Gtf2i* and *Gtf2ird1*, and larger deletion models showed evidence for disrupted social behavior we wanted to directly compare the effects of *Gtf2i** haploinsufficiency to the effects of hemizygosity of the entire WSCR on social behavior.

We first measured maternal separation induced ultrasonic vocalizations (USVs) in postnatal day three and postnatal day five pups. This is a form of developmental communication and was shown to be increased in mice that had three or four copies of *Gtf2i* compared to mice with normal copy number or only one functional copy (29). We saw a significant effect of day ($F_{1,116.00}$=5.43, p=0.021) and genotype on the call rate ($F_{2,60.7}$= 6.09, p=0.004), as well as a genotype by day interaction ($F_{2,61.64}$=6.80, p=0.002). Post hoc analysis within day showed that on day five CD mice made fewer calls than WT littermates (p<0.001) and *Gtf2i** mutant littermates (p=0.045) (**Figure 2A**). We included the weight of the mouse as a covariate to make sure the decrease in call number was not due to differences in weight. We saw that weight has a trending effect ($F_{1,75.48}$=3.95, p =0.05), but the day by genotype interaction term remained significant.

We also observed differences in the temporal and spectral features of the calls. There was a significant effect of genotype on pause length between bouts ($F_{2,60}$=11.9069, p=4.31e-5), with CD mice exhibiting longer pauses on day five compared to WT mice (p=0.0004) and *Gtf2i** mice (p=0.0014); this is correlated with fewer calls produced by CD animals (**Supplemental Figure 2A**). There was a also significant genotype by day interaction for the duration of a call bout ($F_{2,61}$=7.26, p=0.001), with CD mice exhibiting a shorter duration on day five compared to WT (p=0.046) (**Supplemental Figure 2B**). Overall, our study of vocalization provides evidence that *Gtf2i* and *Gtf2ird1* mutation alone are not sufficient to produce a CD-like deficit in this behavior.

Maternal-separation induced USVs are only produced during a transient period of development from postnatal day three to postnatal day 10, peaking at postnatal day seven and postnatal day nine in FVB/AntJ and C57BL/6J strains, respectively (177). Therefore the alteration in the CD animals could reflect an overall shift in developmental trajectory. To assess this, we checked weight gain and developmental milestones in our cohorts. No differences in developmental weights were observed between genotypes. The detachment of the pinnae at postnatal day five, a physical milestone, was similar across all genotypes (χ^2=2.593, p=0.4628,

Supplemental Table 1). However, there were weight deficits in CD animals in adulthood (**Supplemental Figure 2C**). There was a significant effect of day on weight ($F_{4,240}$=1610.9, $p <$ 2.2e-16), a significant effect of genotype ($F_{2,60}$=7.2059, p=0.001568), and a significant day by genotype interaction ($F_{8,240}$=6.9258, p=3.332e-8). These data suggest that gross developmental delay in CD animals does not explain the observed communication deficit.

4.3.3 *Gtf2i** mutation is not sufficient to reproduce WSCR-mediated alterations of social behavior

We went on to test adult social behaviors. We first applied the standard three-chamber social approach, which has not been reported in CD mice. In this task the mice are allowed to freely explore an apparatus with three chambers: a center chamber, a social chamber that contains a cup with a sex and age-matched mouse, and an empty chamber that only contains an empty cup (**Figure 2B**). This test measures the voluntary social approach of mice. We saw the expected preference for the social stimulus across all mice ($F_{1,53}$=83.2013, p=$1.894\text{x}10^{-12}$), with no impact of genotype ($F_{2,53}$=1.1516, p=0.3239) or genotype by stimulus interaction ($F_{2,53}$=0.5845, p=0.5609). Post hoc comparisons within genotypes confirmed that all genotypes spent significantly more time investigating the social stimulus than the empty cup (WT p <0.001; *Gtf2i** p < 0.001; CD p=0.00456; **Figure 2C**). Thus, sociability as measured in this task is not sensitive enough to discern a hypersocial phenotype in these animals.

In a test for social novelty, a novel stranger mouse was then placed in the empty cup. All genotypes showed the expected preference for the novel stimulus animal ($F_{1,53}$=50.3816, p=$3.137\text{x}10^{-9}$), again with no effect of genotype ($F_{2,53}$=1.3948, p=0.2568) or genotype by stimulus interaction ($F_{2,53}$=0.5642, p=0.5722). Post hoc comparisons showed that all the genotypes spent significantly more time investigating the novel stimulus (WT p < 0.001; *Gtf2i** p =0.00321; CD p=0.0012; **Supplemental Figure 2D**). Additionally in this task, we did notice a significant effect of genotype on overall distance traveled ($F_{2,53}$=3.98, p 0.024) with the *Gtf2i** mutants traveling further distance than the WT animals in the sociability trial (p=0.0305; Supplemental Figure 2E), and a corresponding trend during the social novelty trial ($F_{2,53}$=2.87, p=0.115). This suggests that the double mutants have a slight hyperactive phenotype in this task that is not seen in the CD mutants.

Previous reports on social phenotypes in mouse models of WS have described a lack of habituation to a social stimulus. To test this we repeated the three-chamber social approach task in a new cohort of animals with an extended sociability trial to test if the *Gtf2i** mutants or the CD animals showed the preference for the social stimulus after the prolonged amount of time. Similar to the classic three-chamber results we saw a significant effect of the social stimulus in the first five minutes ($F_{1,56}$=19.3683, p=4.891e-5), there was a trend of a genotype effect ($F_{2,56}$=3.098, p=0.053) and no interaction ($F_{2,56}$=0.4650, p=0.6350). Interestingly, we observed a significant preference for the social chamber in the WT and *Gtf2i** mutants, but the CD animals only trended in this direction (**Supplemental Figure 2F**). To determine if the CD mutants do indeed maintain a prolonged social interest compared to WT littermates, we examined the last five minutes of the 30 minute sociability trial. While there was a significant effect of stimulus ($F_{1,56}$=4.82, p=0.03), there was still no effect of genotype ($F_{2,56}$=0.0523, p=0.949) or an interaction ($F_{2,56}$=0.454, p=0.637). In fact, the significant effect of chamber was driven by the proportion of animals investigating the novel empty cup more than the social stimulus (**Supplemental Figure 2G**). These data lead us to conclude that the double mutants and CD animals show a WT-like habituation to social stimulus in this task.

We also tested social dominance in the tube test in these mice. Previous studies using partial deletions of the WSCR showed that the proximal deletion which contains *Gtf2i* and *Gtf2ird1* as well as deletions of both the proximal and distal regions in mice resulted in different win/loss ratios than WT mice and mice lacking just the distal end of the WSCR (94). In contrast, here, the *Gtf2i** and CD animals did not exhibit dominance behavior different than chance would predict (WT vs *Gtf2i** p=0.8318, WT vs CD p=1). *Gtf2i** and CD animals also had similar proportions of wins when paired together (*Gtf2i** vs CD p=0.6291) (**Figure 2D**).

The contrasts in our findings with those reported in prior papers could be due to differences in background strain. Different inbred mouse strains show different dominance behavior (178), and other phenotypes, such as craniofacial morphology in WS models has been shown to be strain dependent (39, 95, 101). We tested the effects of the background strain of the *Gtf2i** and CD models by performing the same task on the respective background of each line and comparing them to their WT littermates. Thi showed that the *Gtf2i** mutants had a WT-like phenotype while the CD mice had a submissive phenotype with significantly more losses to WT littermates (**Supplemental Figure 2H**). Thus, the submissive phenotype of the CD allele is dependent on strain which is not observed in the *Gtf2i** mutants.

Finally, we tested the male mice in a resident-intruder paradigm. In this task, male mice were singly housed for 10 days to establish their territory and, in a series of three test days, novel WT C57BL/6J animals were introduced into their territories as intruders. This task measures both social interactions and bouts of aggression between two freely moving animals (**Figure 2E**). In our study, only one mouse showed aggressive behavior towards the intruder mouse, so we did not further quantify this behavior. Assessment of the social interactions showed a significant main effect of genotype ($F_{2,31}$=5.241, p=0.011) with no effect of day ($F_{2,62}$=2.470, p=0.093) or day by genotyping interaction ($F_{4,62}$=0.1095, p=0.978). Post hoc tests within each day showed that the CD animals spent less total time on day two (p=0.0248) and day three (p=0.0318) engaged in anogenital sniffing compared to the WT animals (**Figure 2F**). These differences could not be explained by differences in total activity levels between the genotypes ($F_{2,31}$=1.399, p=0.262; **Supplemental Figure 2I**). The decrease in total time spent in anogenital sniffing was driven by a shorter average bout time ($F_{2,31}$=5.852, p=0.007, **Supplemental Figure 2J**) and not the number of times the animals initiated the sniffing behavior ($F_{2,31}$=2.7961, p=0.0765; **Supplemental Figure 2K**). The same differences also held for nose-to-nose sniffing (**Figure 2G**). There was a significant effect of genotype ($F_{2,31}$= 3.737, p=0.0352) and no effect of day ($F_{2,62}$=3.01, p=0.056) or day by genotype interaction ($F_{4,62}$=0.8156, p=0.520). Post hoc analysis showed that on day two the CD animals participated in nose-to-nose sniffing significantly less than the WT animals (p=0.0160), while the trend was present in the other days but was not significant. These results indicated that some aspect of social behavior was disrupted in these animals and *Gtf2i** mutants could not recapitulate the full CD phenotype. While we predicted that the WS models would show increased social interest similar to the human condition, individuals with WS have difficulties with other aspects of social behavior, such as social cognition and social awareness (20, 174), which may be reflected in these data.

4.3.4 *Gtf2i** mutation is not sufficient to reproduce WSCR mediated alterations of motor behavior

Along with a characteristic social behavior, WS also presents with other cognitive phenotypes including poor coordination, increased anxiety, specific phobias, repetitive behaviors, and mild intellectual impairment (21). Human studies and mouse models have suggested that *GTF2I* and *GTF2IRD1* contribute in aspects of the visual-spatial deficits and other cognitive phenotypes (36, 38). These genes are also highly expressed in the cerebellum, which could contribute to the coordination problems (72, 78). Therefore, we next tested if CD mice had any motor phenotypes and if haploinsufficiency of these two transcription factors were sufficient to reproduce any deficits.

We performed a sensorimotor battery to assess balance, motor coordination and strength in mutants and WT littermates. All genotypes were similar in the time to initiate walking, and reach the top of a 60 degree inclined screen or a 90 degree inclined screen. All genotypes were able to hang onto an inverted screen for the same amount of time (**Supplemental Figure 3A-D**). CD animals were significantly quicker on turning around on a pole and quicker to get off of the pole than WT animals (**Supplemental Figure 3E-F**), which may be related to body size. There was a significant effect of genotype on time to fall in the ledge task (H_2=12.505,p=0.001925), in which CD animals fell off the ledge faster than either WT (p=0.0071) or *Gtf2i** (p=0.0069) littermates (**Figure 3A**). Similarly, there was a significant effect of genotype on the time spent balancing on a platform task (H_2= 7.1578, p=0.02791) (Supplemental Figure 3G). Despite their comparable performance in strength and coordination tasks, the CD animals tended to have poorer balance, while the double mutants performed similar to WT animals. These findings suggest that other genes in the WSCR contribute to this balance deficit.

To test motor coordination in a more sensitive manner, we evaluated the mice on an accelerating rotarod. This task was performed over three days and tests coordination by quantifying how long a mouse can stay on a rotating rod. There was a main effect of day ($F_{2,339}$ = 81.58, p< $2.2\text{x}10^{-16}$) and a main effect of sex ($F_{1,63}$=10.0227, p = 0.002383), but no main effect of genotype ($F_{2,63}$=2.0394, p=0.13861). We did not observe a sex by genotype interaction ($F_{2,63}$=0.8155, p=0.447035) but did see a day by genotype interaction ($F_{4,333}$=3.6270, p=0.006558). A post hoc comparison between genotypes within each day of testing showed that

*Gtf2i** animals fell off more quickly compared to CD animals on day three ($p=0.04$) with no difference between WT and CD animals (**Supplemental figure 3H**). In contrast to the balance deficit seen on the ledge task but consistent with pole and screen performance, the rotarod results showed that all genotypes have similar motor coordination.

Marble burying is a species-specific behavior that assesses the natural tendency of mice to dig. This task also requires motor skills and has been used as a proxy for repetitive behaviors (179), which are seen in individuals with WS. It has been previously shown that CD animals bury fewer marbles than WT littermates (90, 91). Here we similarly show that there was significant effect of genotype in this task ($F_{2,66}=15.243$, $p=3.61\times10^{-6}$). CD animals buried fewer marbles than both WT ($p<0.001$), and *Gtf2i** mutants ($p=0.000265$) (**Figure 3B),** indicating that *Gtf2i** mutation is not sufficient to recapitulate CD phenotype. The differences in marble burying was not explained by any differences in activity levels between the genotypes during the task ($F_{2,65}=0.8974$, $p=0.4126$; **Supplemental Figure 3I**). However, we did see a significant effect of genotype on distance traveled in the center of the apparatus ($F_{2,66}=13$, $p=0.0015$), with CD mice traveling less distance in the center compared to WT ($p=0.0301$) and *Gtf2i** ($p=0.002$) littermates (**Figure 3C**). There was also a corresponding significant effect of genotype on time spent in the center ($F_{2,66}=14.389$, $p=0.00075$) with CD mice spending less time in the center than WT ($p=0.0079$) and *Gtf2i** ($p=0.0017$) littermates. Avoidance of the center is generally interpreted in rodents as an increase in anxiety-like behavior (Figure 3D). Thus, these results provided further support to the hypothesis that genes besides *Gtf2i** contribute to an anxiety-related phenotype. It also suggested that the decreased marbles buried may be secondary to the decreased time in center and could reflect a phenotype secondary to anxiety rather than a direct stereotypy phenotype.

Finally, to test if the mutants have normal sensorimotor gating we looked at PPI. Similar to other tasks, contrasting evidence has been observed in WS mouse models in this task. Mouse of models of just *Gtf2i* showed no phenotype (96), whereas the proximal deletion mice showed decreased PPI; however, when combined with the distal deletion the phenotype that was suppressed (94). Here we show that all genotypes exhibited the expected increased PPI with an increasing pre-pulse stimulus ($F_{2,112}$=620.61, p < 2e-16), but with no effect of genotype ($F_{2,56}$=0.7742,p=0.466) or a pre-pulse by genotype interaction ($F_{4,112}$=1.926,p=0.111) (**Supplemental Figure 3J**). A decrease was observed for overall startle response to the 120dB stimulus by CD animals, but when we included weight in the statistical model this effect disappeared (genotype $F_{2,55}$=1.48, p=0.2365; weight $F_{1,55}$=26,001, p=4.34e-6). Thus, the only phenotypic difference seen simply reflected the smaller size of the CD mice and not a change in sensorimotor gating (**Supplemental Figure 3K**).

4.3.5 WSCR mutation does not produce robust anxiety-like behaviors

WS patients have heightened anxiety (21), and mouse models of *Gtf2i* (67, 96) and *Gtf2ird1* (100, 101) mutations have produced mixed evidence to support the role of these genes in anxiety phenotypes. Larger deletion models that have either the proximal or distal regions deleted showed anxiety-like phenotypes in the open field, but not in light-dark boxes (94). Similarly the CD model has been shown to not have any differences in the open field task (93). We wanted to directly compare animals with *Gtf2i* and *Gtf2ird1* mutations to CD animals in the same tasks to test exploratory and anxiety-like phenotypes. First, we looked at the behavior of the mice in an one hour locomotor activity task. We did not see any effect of genotype on the total distance traveled ($F_{2,66}$=0.6324, p=0.53449), however there was a trend towards a time by genotype interaction ($F_{10,330}$=1.7817, p=0.06283; **Figure 3E**) with the *Gtf2i** mutants traveling further distance. This was consistent with the behavior observed during the three-chamber social approach task. In contrast to the marble burying task, here we did not see a significant main effect of genotype on the time spent in the center of the chamber ($F_{2,66}$=2.3104, p=0.10720) though we observed a trend in the first ten minutes for CD mice to spend less time in the center (**Figure 3F**). However, the *Gtf2i** mice did not show a similar trend. To further test for anxiety-like phenotypes, we performed elevated plus maze testing. Across the three days of testing, all genotypes spent similar percent time in the open arms of the apparatus ($F_{2,63}$=0.6351, p=0.5332; **Supplemental Figure 3L**). Overall, our experiments indicate there may be a subtle increase on some tasks in anxiety-like behavior in CD mice. However, if there is such a phenotype, we see no evidence that *Gtf2i** mutations are sufficient to produce it.

4.3.6 *Gtf2i** mutation is not sufficient to reproduce WSCR mediated alterations of fear conditioning

Finally, as patients with WS have both intellectual disability and increased prevalence of phobias (21, 180), we tested associative learning and memory of the mice using a contextual and cued fear conditioning paradigm. These behaviors are also mediated by brain regions that have shown to be altered in mouse models of WS and human patients, namely the amygdala and hippocampus. Individuals with WS have altered structural and functional reactivity in the hippocampus and amygdala as reviewed in (15) compared to typically developing controls. Both of these regions play a role in both contextual and cued fear conditioning (181). Likewise, CD mice have been shown to have altered morphology and physiology in the hippocampus (93, 182), thought to be important in contextual fear conditioning.

We therefore tested associative learning and memory of the animals using a three day conditioned fear task (**Figure 4A**). During the conditioning trial on day one we saw a significant difference in baseline freezing during the first two minutes, when the mice were initially exploring the apparatus. There was a main effect of genotype ($F_{2,53}$=5.31,p=0.00794) and a main effect of minute ($F_{1,53}$=7.28, p=0.009), with the CD animals freezing more than the WT animals (p=0.04) and the *Gtf2i** mutants (p=0.05) during minute one prior to any shock. By minute two of baseline, all animals showed similar levels of freezing. During the pairing of the foot shock with the context and tone during minutes three through five, we saw a significant effect of time ($F_{2,106}$=100.3071, $p < 2.2x10^{-16}$) and genotype ($F_{2,53}$=3.4304, p=0.039723) as well as a time by genotype interaction ($F_{4,106}$=3.9736, p = 0.004812). Specifically, all mice increased the amount of freezing after each foot shock, but after the last foot shock the *Gtf2i** mutants froze less than the CD animals (p=0.002; **Figure 4B**), but similarly to the WT littermates. On the subsequent day, to test contextual fear memory, mice were put back in the same apparatus and freezing behavior was measured. Comparing the average of the first two minutes of freezing during fear memory recall on day two to the baseline of the conditioning day, we saw that all genotypes exhibited contextual fear memory; indicated by the increased levels of freezing when put back in the same context they were conditioned in ($F_{1,53}$=36.4882, $p=1.56x10^{-7}$; **Supplemental Figure 4A**). Looking across time during the fear memory recall we saw a significant effect of time ($F_{7,371}$=2.7166, p=0.009291) with no main effect of genotype ($F_{2,53}$=1.2507, p=0.294625), but a time by genotype interaction ($F_{14,371}$=2.499, p=0.002085). Post hoc analysis within time showed that CD mice froze more than WT and *Gtf2i** littermates during minute three of the task (**Figure 4C**).

To test cued fear conditioning, on the subsequent day the mice were put in a different context and were played the tone that was paired with the foot shock during the conditioning day. All animals had similar freezing behavior during baseline ($F_{2,53}$=1.061, p=0.353). For the duration of the tone, there was a significant effect of time ($F_{7,371}$=21.5824, $p<2x10^{-16}$) but no effect of genotype ($F_{2,53}$=0.3014, p=0.741) or genotype by time interaction ($F_{14,371}$=0.2128, p=0.999) (**Figure 4D**). Finally, the differences in freezing behavior could not be explained by sensitivity to the foot shock as all mice showed similar behavioral responses to increasing shock doses ($F_{2,56}$=1.4521, p=0.2427; **Supplemental Figure 4B**). Overall, CD mice showed an enhancement of fear response to a contextual fear memory, and mutations in *Gtf2i** were not sufficient to reproduce this phenotype.

4.3.7 *Gtf2i** mutation is not sufficient to reproduce WSCR mediated alterations of hippocampal gene expression.

In addition to permitting behavioral phenotyping, mouse models also allow for well-powered and controlled examination of the molecular consequences of mutation in the environment of a fully developed and functioning central nervous system. Therefore, we turned from behavioral phenotyping of cognitive tasks to molecular phenotyping in the brains of these mice to 1) identify candidate molecular mediators of the behavioral phenotypes and 2) determine to what extent any transcriptional phenotype of WSCR mutation might be mediated by the haploinsufficiency of these two transcription factors. We specifically focused on the hippocampus, since we saw deficits in marble burying and differences in contextual fear memory, two behaviors thought to be mediated by hippocampal function (159, 181). Other studies in the CD animals have also shown there to be differences in LTP in the hippocampus as well as differences in Bdnf levels (91, 182). Yet the transcriptional consequences genome-wide of WSCR loss hav not been characterized in the hippocampus.

First, we conducted a targeted analysis of the genes in the WSCR locus. Of the 26 genes that make up the WSCR, only 15 were measurably expressed in the adult mouse hippocampus.

Consistent with expectation, all genes in the WSCR region showed a decrease in RNA abundance in the CD animals, and genes that lie immediately outside the region were not affected. *Gtf2i** mutants only showed disruption of *Gtf2i* and *Gtf2ird1* in directions consistent with what was previously seen in our RT-qPCR. This confirmed the genotype of the samples, and indicated that these transcription factors are not robust trans regulators of any other genes in the locus (**Figure 5A**).

Next, we conducted differential expression analysis comparing WT to CD littermates to identify the molecular consequences of WSCR loss. At an FDR < 0.1 we found 39 genes to be differentially expressed. Of the 39 genes, 15 were genes that are located in the WSCR. This small number of differentially expressed genes was surprising given that several of the WSCR genes are described as transcription factors. In addition to these differentially expressed genes, the magnitude of the changes were small (**Figure 5B** and **Supplemental Figure 5A**). Interestingly, *Slc23a1* showed to be slightly but consistently more lowly expressed in the CD animals compared to the WT animals. This is a GABA transporter, suggesting that inhibitory signaling could be altered in the hippocampus. This gene has also been shown to decreased in WS-derived cortical neurons (45). Also of note, the *Iqgap2* gene was shown to be elevated in the CD animals compared to WT animals. This gene was also upregulated in WS iPSCs (44). We also looked at genes that have been investigated previously in the CD mouse, such as *Bdnf* and *Pi3kr* (90, 91) and we show that there was little change in gene expression between genotypes (**Supplemental Figure 5B**).

To determine if *Gtf2i** loss is sufficient to drive these transcriptional changes, we next examined differential expression comparing *Gtf2i** mutants to WT littermates. In contrast to WSCR mutation, we found only *Gtf2i* and *Gtf2ird1* to be differentially expressed at an FDR <

0.1 (**Figure 5C**). To get a broader idea of how similar the transcriptomes of the two genotypes are we compared the genes that are nominally up and downregulated between each mutant line and WT controls. We saw that there was about a 9% overlap between CD and *Gtf2i** up and down regulated genes (**Figure 5D**). This is slightly below the amount of genes shown to be changed by *GTF2I* in iPSCs (44). Again this suggests that other genes in the WSCR are driving 90% of the transcriptional changes in the CD hippocampus.

To understand what role the nominally changed genes have in common we conducted a GO analysis. The biological processes that the CD genes were found to be involved in included synaptic functioning as well as nervous system differentiation. Interestingly processes that control balance were enriched and we and others have reported on balance deficits in CD animals (**Figure 5E**). When comparing these to 1000 random differential gene lists these biological processes are very specific to the genotype comparisons. For instance, out 1000 random test, positive regulation of excitatory synapses only occurred in the top 10 enriched GO terms two times (**Supplemental Table 2**). The cellular components that the genes are enriched for are extracellular, which is a similar result to the iPSC studies (44), as well as synapses. The molecular function ontologies, which are enriched for the differentially expressed genes included calcium binding (**Supplemental Figure 5**). When comparing these to randomly determined gene expression changes, all but the extracelluar components seem to be specific to the CD versus WT comparison (**Supplemental Table 2**). In contrast, the *Gtf2i** GO analysis showed that these genes are enriched for more general organ system development and are not very nervous system specific (**Figure 5F** and **Supplemental Table 3**).

Overall, we have shown that the hemizygous loss of the WSCR has a mild but significant effect on the hippocampal transcriptome. Yet, the changes that do occur point to aberrations in synapses and nervous system development. Furthermore, loss of function mutations in *Gtf2i* and *Gtf2ird1* have an even smaller effect on the transcriptome and can only account for 9% of the changes incurred by loss of the WSCR.

4.4 Discussion

Contiguous gene disorders such as WS provide insight into regions of the genome that have large effects on specific aspects of human cognition and behavior. The specific cognitive profile of WS is characterized by deficits in visual-spatial processing with relative strengths in language, and the archetypal behavioral profile consists of increased social interest, strong eye contact, high levels of anxiety, and in some cases specific phobias and hyperactivity. Here we used a new mouse model to test if loss of the paralogous transcription factors *Gtf2i* and *Gtf2ird1* are sufficient to phenocopy the behaviors and transcriptomic changes of mice that lack the entire WSCR.

Overall, CD mice consistently have more severe phenotypes than the *Gtf2i** mutants. We saw that the CD animals have a deficit in social communication as measured by maternal separation induced pup ultrasonic vocalizations. The *Gtf2i** mutants on average make fewer calls than the WT littermates, however not significantly so, but this may suggest that these two transcription factors contribute slightly to this phenotype but other genes in the region are necessary to produce the full phenotype seen in the CD animals. Previously it was shown that animals that have increased copy number of *Gtf2i* increased the number of pup USVs emitted while animals with only one copy produced similar number of calls to WT animals (29). This was interpreted as increased separation anxiety. Here we see that lower copy number of the entire region produces the opposite effect of increased *Gtf2i* copy number. Decreased USVs could mean there is a lack of motivation to make the calls or an inability to make as many calls.

A model of *Gtf2ird1* mutant animals was shown to have different USV production due to a difference in the muscle composition of the larynx (92). This has not been shown in the CD animals but it could contribute to the phenotype as well as differences in the skull morphology (93). Another possible explanation is that since the production of USVs is a developmentally regulated trait, it could be that deleting 26 genes could disrupt typical developmental trajectories. While we do not see any gross developmental problems such as lower weight or delayed detachment of pinnae, the deletion could have a more severe effect on brain development, thus affecting developmentally regulated behavioral traits.

To our surprise, there was no detectable social phenotype in the *Gtf2i** mutants or CD animals in the classical three-chamber social approach assay. Our results showed that all genotypes on average prefer to investigate the social stimulus for a similar amount of time. The preference for social novelty is also intact across all the groups. In an attempt to test if the WS models fail to habituate to a social stimulus we showed that after thirty minutes of having the opportunity to investigate an unfamiliar mouse or an empty cup, all genotypes habituate to the social stimulus and by the end of the thirty minutes seem to have a small preference for the empty cup. The three-chamber social approach task has been done in the larger partial deletion models where they have shown that the proximal deletion and the trans full deletion models have a significant preference for the social stimulus, and the WT and distal deletion mice do not show a preference, suggesting that the proximal deletion, which harbors genes such as *Gtf2i* and *Gtf2ird1*, are involved in this social task (94). Mouse models that are haploinsufficient for only *Gtf2i* have shown in the three-chamber approach task that after eight minutes WT animals investigate a novel object the same amount as a social stimulus, but the *Gtf2i* mutants still have a significant preference suggesting a lack of habituation (96). In another *Gtf2i* model, Martin et al.

compared animals with one, two, three, and four copies of *Gtf2i* in the three-chamber social approach task, and showed that only animals with one and three copies of *Gtf2i* displayed a significant preference for the social stimulus (97), but WT animals did not. These three-chamber social approach tests are interpreting a lack of significance as evidence for increased social behavior and not directly comparing the levels of investigation between genotypes (183). Furthermore, in some cases the WT controls are not showing the expected preference for the social stimulus, thus, possibly confounding interpretation of the mutant preference.

The three-chamber social approach assay has come under recent criticism due to how dependent it is on activity levels of mice and its lower heritability compared to tests of direct social interaction (184). The CD animals had not previously been tested in this procedure exactly, but have been tested in a modified social approach where the time spent investigating a mouse in a cup is measured but with no competing non-social stimulus (90, 91, 93). The data showed that the CD animals investigated the social stimulus for more time than the WT animals and delivery of *Gtf2i* cDNA by AAV9 via the magna cisterna can return the investigation time to normal levels (90). Here, we showed that all animals preferred the social stimulus. It is possible that the standard social approach suffers from several confounding factors, such as lower heritability, as well as activity and anxiety-like components that make this task less sensitive to detect a hypersocial phenotype in WS models. It could also be that the three-chamber social task does not test the specific aspects of social behavior that are disrupted in WS models. For example, newer tasks, such as social operant tasks that test motivation to receive a social stimulus may more directly test the aspects of social behavior that are affected in WS. This task has been performed on *Gtf2i* mutants and mice that have only one copy of *Gtf2i* will work harder to receive a social reward (97).

Direct social tasks have higher heritability than the three-chamber social approach and offer a more natural social experience (184), which may make them a more sensitive assay for testing social behaviors. Direct tasks have shown that *Gtf2i* models have increased nose-to-nose investigation time (97), mouse models lacking the proximal end of the region have increased investigation frequency (94), and *Gtf2ird1* mutants make fewer aggressive actions but show increased following time (101). We employed the resident-intruder paradigm as a full contact social assay. While we did not see bouts of aggression from any of the genotypes, we could see differences in social investigation. To our surprise, the CD animals spent less time overall in anogenital sniffing and nose-to-nose sniffing of the intruder animals when compared to WT littermates. The double mutants were not significantly different from the WT animals but had intermediate values between the WT and CD animals. This phenotype was being driven by the decreased time per bout of investigation in the CD animals, as all genotypes had a similar frequency of the sniffing behavior. This result was contrary to what would be predicted from the human condition and previous mouse results. However, while individuals with WS are described as having prosocial behavior in terms of increased social approach and friendliness (19), they also have difficulties maintaining long term relationships because of deficits in other aspects of social behavior (20, 27, 28, 174), and on scales measuring social reciprocity often score in the autistic range (174). In addition, there is a high co-morbidity with ADHD which has features of impulsiveness (22). While the CD animals did not show the expected increase in social interest, this may be a manifestation of attention deficits that are present from deleting the 26 genes in the WSCR, but this needs to be examined. Loss-of-function mutations in *Gtf2i* and *Gtf2ird1* were not sufficient to produce as strong an effect in these investigative behaviors. However, the somewhat intermediate effect suggests they could contribute to it.

One limitation of our study is that some aspects of the social phenotype in the models tested here could be masked by the mouse background strain. While we have controlled for mouse background strain in our experiments by only using the F1 generation of the FVB/AntJ and C57BL/6J cross, the hybrid background may prevent the manifestation of a social phenotype caused by the mutations tested. For example, it has been documented that craniofacial phenotypes in *Gtf2ird1* models are sensitive to background strain (39, 78, 95, 101). Here, the double mutants and CD animals on the hybrid background showed no dominance phenotype in the tube test. However, when we tested each mutation on the respective mouse background strain, we saw that the CD animals had a submissive phenotype, but the double mutants did not. Studies done in the larger partial deletions have shown altered win/loss ratios in the tube test in the proximal deletion and full trans deletion models (94), suggesting that the CD models on the C57BL/6J background can replicate this phenotype, but other genes in the proximal region besides *Gtf2i* and *Gtf2ird1* are also required.

In this study, we have replicated several of the phenotypes previously seen in the CD animals, such as marble burying and balance deficits (91, 93, 182). It was shown that CD animals bury fewer marbles than WT animals and rescuing the *Gtf2i* levels in the hippocampus did not rescue this phenotype. Both the results presented here and in Borralleras et al. suggest that other genes in the region beyond *Gtf2i* and *Gtf2ird1* are important in this behavior. Here we have extended the results to suggest that there could be an anxiety-like component to the marble burying deficit. By tracking the animals during the task we see that CD animals spend less time and travel less distance in the center of the apparatus. This could preclude them from burying as many marbles in the center. It could also be that the CD animals do not show the normal motivation to dig.

CD animals showed difficulty in balancing tasks, but normal motor coordination. Motor coordination of WS has been tested using the rotarod. The larger partial deletion models showed that the distal deletion and proximal deletion mice had intermediate phenotypes with the full trans deletion mice falling off the rotarod sooner (94). Similarly the CD mice have shown deficits in the rotarod and addition of *Gtf2i* coding sequence does not rescue this phenotype (182). The CD mice in this study did not show a deficit in the rotarod despite having poor balance on the ledge and platform tasks. CD animals were not able to balance on a ledge or platform as long as their WT and *Gtf2i** mutant littermates. This suggests that motor coordination, as tested by our rotarod paradigm, is intact in these WS models, but balance is specifically affected in the CD animals. The discrepancy could be due to body size. The adult CD animals are significantly smaller than the WT and *Gtf2i** mutants, which could make staying on the wider rotarod less challenging. This study also used a different accelerating paradigm where the rod itself is continuously accelerating until the mouse falls off while other paradigms test the mice at different continuous rotation speeds.

Along with balance and coordination problems, individuals with WS tend to have specific phobias and high levels of non-social anxiety (21). We showed that CD animals had an altered fear conditioning response. We saw that the CD animals have an increased fear response in contextual fear but not cued fear. It was previously reported that CD animals showed a slight decrease in freezing but was not significant (93). Two separate *Gtf2ird1* mutations have shown contrasting results, one showed an increased fear response (99) while another showed decreased fear response (101). It could be that this hybrid background used here is more sensitive to see increases in freezing because FVB/AntJ do not exhibit as much freezing in conditioned fear tasks as C57BL/6J animals (185). The observed increased contextual fear response could be due to differences in the hippocampus and amygdala, both regions that have been shown to be disrupted in WS. We did not see a robust anxiety-like behavior phenotypes in one hour locomotor task or the elevated plus maze, which is consistent with previous findings in the CD model (93). However, we did see reduced time and distance traveled in the center during the marble burying task. Perhaps suggesting that the novel environment in combination with the novel marbles can induce slightly higher levels of anxiety in the CD model.

Given the behavioral differences in marble burying and contextual fear, two behaviors thought to be mediated by the hippocampus (159, 181), we examined the transcriptomes of the hippocampus of the *Gtf2i** mutants and CD animals and compared them to WT littermates. This provided the first transcriptional profile documenting the consequences of the 26 gene deletion in a mature brain, and allowed us to determine what portion of that was driven by Gtf2i* proteins. Surprisingly, we did not see any significantly differentially expressed genes between the *Gtf2i** mutants and WT littermates, besides the mutated genes themselves. Looking at the overlap of nominally differentially expressed genes between CD-WT and *Gtf2i**-WT comparisons, showed a small overlap of about 9%. This is slightly less than the estimate from Adamo *et al.*, of 15-20% of genes dysregulated in WS iPSCs being attributed to reduced levels of *GTF2I.* Perhaps these general findings suggest that *Gtf2i* and *Gtf2ird1* contribute to small transcriptional changes broadly across the genome, and in combination with other genes in the WSCR more profound neural specific gene disruptions become apparent.

Our transcriptional studies overall showed limited impact of *Gtf2i** mutation in the brain. The global brain transcriptome of *Gtf2i* mutants has not been investigated, but brain transcriptome studies of *Gtf2ird1* knockout mouse models have not found any evidence of differentially expressed genes (88). These data suggest that in the adult hippocampus these two transcription factors do not greatly affect the transcriptome. There are some limitations to this negative result. It could be that we are diluting some of the signal because we are studying the effects on the transcriptome of the whole hippocampus, which has a diverse cellular composition. Larger effect sizes might be detected in more homogenous cellular populations. Likewise, if these genes regulate dynamics of gene expression rather than baseline values, greater differences might become apparent after experimental manipulations that activate transcription.

One additional limitation of our study is that the mutated *Gtf2ird1* allele is still producing an N-terminally truncated protein. However, we show that N-truncated Gtf2ird1 does not bind to its known target, the promoter region of *Gtf2ird1*, and this absence leads to increased RNA from the locus, consistent with a loss of its transcriptional repressor function. Thus, we confirmed this truncated protein is a loss of function for the only known roles for Gtf2ird1. However, it is possible that the protein does have other unknown functions we could not assay here. It has also been proven to be a remarkably challenging gene to completely disrupt, across multiple studies (66, 101). The combination of the upregulation of its RNA upon deletion with the ability to re-initiate at a variety of downstream codons is intriguing. One possibility is that Gtf2ird1 has an unusual amount of homeostatic regulation at both transcriptional and translational levels that are attempting to normalize protein levels. Another possibility is that these kinds of events are actually quite common across genes, but that they are detected in Gtf2ird1 because the WT protein is at such low abundance it is on par with what is actually an infrequent translation re-initiation event. Our detection of Gtf2ird1 protein in the brain required substantial optimization and is still only apparent in younger brains. Indeed, in validations of mutations of more abundant proteins, the immunoblots may not be routinely developed long enough to see a trace re-initiation event that might occur. Regardless, future studies aimed at understanding the transcriptional and translational regulation of this unusual gene would be of interest.

Examining the profile of CD mutants compared to WT littermates, we do define a number of transcriptionally dysregulated genes. Of the genes in WSCR that are expressed in the hippocampus all had decreased expression in the CD animals. In addition, there were 24 genes outside the WSCR that had a FDR < 0.1 between CD and WT controls. Among these genes is *Slc23a1*, the GABA vesicle transporter, which is down regulated in CD animals. Interestingly this gene was also found to be down regulated in human iPSC derived neurons from individuals with WS (45). This points to aberrant inhibitory activity in the CD brain, which could lead to functional deficits. Also consistent with other human WS derived iPSC studies, the gene *Iqgap2* was shown to be upregulated in the CD hippocampus (44), and has the potential to interact with the cytoskeleton through actin binding (186). Broadening the analysis to include nominally differentially expressed genes and conducting systems-level analyses, the CD-WT comparison highlighted genes involved in the positive regulation of excitatory postsynaptic potential. Chailangkarn et al. showed that WS derived iPSC neurons had increased glutamatergic synapses. Our data also showed some signal in the GO term for postsynaptic density assembly. Taken together these data suggest abnormal synapse functioning in the CD animals and potentially altered inhibitory/excitatory balance. This also suggests pharmacological agents that increase GABA tone may be of use in reversing some WS phenotypes. The RNA-seq data also had signal in neuromuscular processes controlling balance. Altered gene expression in the CD animals could be contributing to the balance deficits. In contrast to the synapse and neural specific GO term enrichment seen in the CD-WT comparison, comparing the transcriptomes of the *Gtf2i** mutants and WT shows signal in more general organ development, such as ossification and eye development.

Taken together, our results support the hypothesis that other genes in the WSCR besides *Gtf2i* and *Gtf2ird1* are necessary to produce some phenotypes that are seen when the entire WSCR is deleted. While these two transcription factors have been highlighted in the human literature as large contributors to the WS phenotype, the literature is also consistent with a model where most genes contribute to aspects of different phenotypes in WS, but the full phenotypic effects occur when all the genes are deleted (**Figure 6**). Studying patients with atypical deletions highlights the variability of the region. Even within families that have inherited small deletions some of the cardiovascular, cognitive, and craniofacial phenotypes have incomplete penetrance (31, 32, 40). Comparing the deletion sizes and corresponding phenotypes shows a large overlap of genes that are deleted, but no clear pattern of which specific phenotypes are affected. Many of atypical deletions described to date that do not have *Gtf2i* and *Gtf2ird1* deleted show no overfriendly phenotype, but there are examples where this is not true. Recent work in zebrafish that was done to dissect which genes in the 16p11.2 region contribute to craniofacial dysmorphology led to a similar conclusion, that multiple genes in the region contribute to the phenotype but in combination some have synergistic effects and others have additive effects (102). Sanders *et al.* also suggested that copy number variations with higher gene content are more likely to have several genes of smaller effect sizes suggesting an oligogenic pattern of contribution (121). Our data suggests that looking beyond the general transcription factor 2I family at possible combinations of more genes in the region may more completely reproduce the WS phenotype. Given the ease of making new mouse models with current genome editing technology, a combinatorial dissection of the region is feasible and could lead to interesting new insight into the underlying mechanisms that contribute to the phenotypic spectrum of WS.

4.5 Materials and Methods

Generating genome edited mice

sgRNAs were designed to target early constitutive exons of the mouse *Gtf2i* and *Gtf2ird1* genes. The gRNAs were cloned into the pX330 Cas9 expression plasmid (Addgene) and transfected into N2a cells to validate the cutting ability of each gRNA using the T7 enzyme assay. Primers used to amplify target regions tested by the T7 enzyme assay are in Supplemental Table 4. One guide was selected for each gene based on cutting activity (Supplemental Table 4). The gRNAs were in vitro transcribed using MEGAShortScript (Ambion) and Cas9 mRNA was in vitro transcribed, G-capped, and poly-A tailed using the mMessageMachine kit (Ambion). The mouse genetics core at Washington University School of Medicine co-injected the Cas9 mRNA (25ng/ul) along with both gRNAs (13ng/ul of each gRNA) into FVB/NJ fertilized eggs and implanted the embryos into recipient mothers. This resulted in 57 founders. Founders were initially checked for any editing events using the T7 assay. There were 36 animals with no editing events. We deep sequenced the expected cut sites, as described below, in the remaining 21 founders to identify which alleles were present.. Founders were crossed to wild type (WT) FVB/AntJ (https://www.jax.org/strain/004828) animals, which are different from FVB/NJs at two loci; Tyr$^{c\text{-}ch}$ results in a chinchilla coat color and they are homozygous WT for the 129P2/OlaHSd *Pde6b* allele, which prevents them from developing blindness due to retinal degeneration. Coat color was visually genotyped and the functional FVB/AntJ *Pde6b* allele was genotyped using primers recommended by Jackson labs (**Supplemental Table 5**). The mice were crossed to FVB/AntJ until the mutations were on a background homozygous for the FVB/AntJ coat color and *Pde6b* alleles.

Genotyping

Initial founder genotyping was performed by deep sequencing amplicons around the expected cuts sites of each gRNA. Primers were designed around the cut sites using the NCBI primer blast tool. To allow for Illumina sequencing we concatenated the Illumina adapter sequences to the designed primers (**Supplemental Table 5**). The regions surrounding the cut sites were amplified using the following thermocycler conditions: 95° C 4 minutes, 95° C 35 seconds, 58.9° C 45 seconds, 72° C 1 minute 15 seconds, repeat steps 2 through 4 35 times, 72° C for 7 minutes, hold at 4° C. A subsequent round of PCR was performed to add the requisite Illumina P5 and P7 sequences as well as sample specific indexes using the following thermocycler conditions: 98° C 3 minutes, 98° C 10 seconds, 64° C 30 seconds, 72° C 1 minute, repeat steps 2 through 4 20 times, 72° C 5 minutes, hold 4° C. The PCR amplicons were pooled and run on a 2% agarose gel and the expected band size was gel extracted using the NucleoSpin gel extraction kit (Macherye-Nagel). The samples were sequenced on a MiSeq. The raw fastq files were aligned to the mm10 genome using bwa v0.7.17 –mem with default settings (140), and the bam files were visualized using the integrated genome visualizer (IGV)v2.3.29 to determine the genotype.

Once the alleles of the founder lines were shown to be in the germline, we designed PCR genotyping assays that can distinguish mutant and WT alleles. Since the *Gtf2i* mutation and the *Gtf2ird1* mutation are in linkage and are always passed on together, primers were designed that would only amplify the five base pair deletion in exon three of *Gtf2ird1*. The primer was designed so that the three prime end of the forward primer sits on the new junction formed by the mutation with an expected size of 500bp. Beta actin primers, with an expected size of 138bp, were also used to help ensure specificity of the mutation specific *Gtf2ird1* primers as well as act as a PCR control (**Supplemental Table 5**). The CD animals were genotyped using primer sequences provided by Dr. Victoria Campuzano and primers that amplify the WT *Gtf2ird1* allele as a PCR control (**Supplemental Table 5**).

PCR was performed on toe clippings that were incubated overnight at 55° C in tail lysis buffer (10mM Tris pH 8, 0.4M NaCl, 2mM EDTA, 0.1% SDS, 3.6U/mL Proteinase K (NEB)). The proteinase K was inactivated by incubation at 99° C for 10 minutes. 1ul of lysate was used in the PCR reactions. Two bands indicated a heterozygous mutation in *Gtf2i* and *Gtf2ird1*. The cycling conditions for the 5bp *Gtf2ird1* deletion were: 95° C 4 minutes, 95° C 35 seconds, 66.1° C 45 seconds, 72° C 1 minute 15 seconds, repeat steps 2 through 4 35 times, 72° C for 7 minutes, hold at 4° C. The cycling conditions for the CD genotyping were: 95° C 4 minutes, 95° C 35 seconds, 58° C 45 seconds, 72° C 1 minute 15 seconds, repeat steps 2 through 4 35 times, 72° C for 7 minutes, hold at 4° C.

Western blotting

E13.5 whole brains were dissected in cold PBS and immediately frozen in liquid nitrogen and stored at -80°C until genotyping was performed. Frozen brains were homogenized in 500ul of 1x RIPA buffer (10mM Tris HCl pH 7.5, 140mM NaCl, 1mM EDTA, 1% Triton X-100, 0.1% DOC, 0.1% SDS, 10mM Na_3V0_4, 10mM NaF, 1x protease inhibitor (Roche)) and RNAase inhibitors (RNasin (Promega) and SUPERase In (Thermo Fisher Scientific) and incubated on ice for 20 minutes. Lysates were cleared by centrifugation at 10,000g for 10 minutes at 4° C. The lysate was split into two 100ul aliquots for protein analysis and 250ul of lysate was added to 750ul of Tizol LS (Thermo Fisher Scientific) for RNA analysis. Protein concentration was quantified using a BCA assay and loaded at 25-50ug in 1x Lamelli Buffer with B-mercaptoethanol onto a 4-15% TGX protean gel (Bio-Rad). In some experiments to achieve greater separation to detect the N-truncation, the protein lysates were instead run on a 7.5% TGX protean gel (Bio-Rad). The protein was transferred to PVDF 0.2um membrane by wet transfer. The membrane was blocked for one hour at RT with TBST 5% milk. The membranes were cut at 75KDa, and the top of the membrane was probed for either Gtf2i or Gtf2ird1, and the bottom of the membrane was probed for Gapdh, with the following primary antibodies: Rabbit anti-GTF2IRD1 (1:500, Novus, NBP1-91973), Mouse anti-GTF2I (1:1000 BD Transduction Laboratories, BAP-135), and Mouse anti-Gapdh (1:10,000, Sigma Aldrich, G8795). Primary antibodies were incubated overnight at 4° C in TBST 5% milk. We used the following secondary antibodies: HRP-conjugated Goat anti Rabbit IgG (1:2000, Sigma Aldrich, AP307P) and HRP-conjugated Goat anti Mouse IgG (1:2000, Bio Rad, 1706516) and incubated for 1 hour at room temperature. Signal was detected using Clarity Western ECL substrate (Bio-Rad) in a MyECL Imager (Thermo Scientific). Quantification of bands was performed using Fiji (NIH) (187) normalizing to Gapdh levels and a WT reference sample.

Transcript measurement using RT-qPCR

Total RNA from E13.5 brains lysates was extracted from Trizol LS using the Zymo Clean and Concentrator-5 with on column DNAase I digestion and eluted in 30ul of water. RNA quantity and purity was determined using a Nanodrop 2000 (Thermo Scientific). cDNA was prepared using 1ug of total RNA and the qscript cDNA synthesis kit (Quanta Biosciences). 25ng of cDNA was used in a 10ul RT-qPCR reaction with 2x PowerUP SYBR Green Master Mix

(Applied Biosystems) and 500nM primers that would amplify constitutive exons of *Gtf2ird1* (exons 8/9), *Gtf2i* (exons 25/27) or *Gapdh* (**Supplemental Table 5**). The RT-qPCR was carried out in a QuantStudio6Flex machine (Applied Biosystems) with the following cycling conditions: 95° C 20 seconds, 95° C 1 second, 60° C 20 seconds, repeat steps 2 through 3 40 times. There were three biological replicates per genotype in all experiments and each cDNA was assessed in triplicate technical replicates. Relative transcript abundance of *Gtf2i* and *Gtf2ird1* was determined using the deltaCT method normalizing to *Gapdh*.

ChIP-qPCR

Chromatin preparation

Chromatin was prepared by homogenizing one frozen E13.5 brain in 10mL of 1x cross-linking buffer (10mM HEPES pH7.5, 100mM NaCl, 1mM EDTA, 1mM EGTA, 1% Formaldehyde (Sigma)) using the large clearance pestle in a Dounce homogenizer and allowed to crosslink for 10 minutes at room temperature with end-over-end rotation. The formaldehyde was quenched with 625ul of 2M glycine. The cells were spun down at 200g at 4° C and the pellet was washed with 10mL 1x PBS 0.2mM PMSF and spun again. The pellet was resuspended in 5mL L1 buffer (50mM HEPES pH 7.5, 140 mM NaCl, 1mM EDTA, 1mM EGTA, 0.25% Triton X-100, 0.5% NP40, 10.0% glycerol,1mM BGP (Sigma), 1x Na Butyrate (Millipore), 20mM NaF, 1x protease inhibitor (Roche)) and homogenized using the low clearance pestle in a Dounce homogenizer to lyse the cells and leave the nuclei intact. The homogenate was spun at 800g for 10 minutes at 4° C to pellet the nuclei. The pellet was washed in 5mL of L1 buffer and spun again and resuspended in 5mL of L2 buffer (10mM Tris-HCl pH 8.0, 200mM NaCl, 1mM BGP, 1x Na Butyrate, 20mM NaF, 1x protease inhibitor) and incubated at room temperature for 10 minutes while shaking. The nuclei were pelleted by spinning at 800g for 10 minutes and resuspended in 950ul of L3 buffer (10mM Tris-HCl pH 8.0, 1mM EDTA, 1mM EGTA, 0.3% SDS, 1mM BGP, 1x Na Butyrate, 20mM NaF, 1x protease inhibitor) and transferred to a milliTUBE 1mL AFA Fiber (100)(Covaris). The sample was then sonicated to a DNA size range of 100-500bp in a Covaris E220 focused-ultrasonicator with 5% duty factor, 140 PIP, and 200cbp. The sonicated samples were diluted to 0.1% SDS using 950ul of L3 buffer and 950ul of 3x Covaris buffer (20mM Tris-HCl pH 8.0, 3.0% Triton X-100, 450mM NaCl, 3mM EDTA). The samples were spun at max speed in a tabletop centrifuge for 10 minutes at 4° C to pellet any insoluble matter. The supernatant was pre-cleared by incubating with 15ul of protein G coated streptavidin beads (ThermoFisher) for two hours at 4° C.

Chromatin IP

GTF2IRD1 antibody (Rb anti GTF2IRD1 NBP1-91973 LOT:R40410) was conjugated to protein G coated streptavidin beads by incubating 6ug of antibody (10ul) with 15ul of beads in 500ul TBSTBp (1x TBS, 0.1% Tween 20, 1%BSA, .2mM PMSF) and end-over-end rotation for one hour at room temperature. The antibody-conjugated beads were washed three times with 500ul of TBSTBp. 400ul of the pre-cleared lysate was added to the antibody-conjugated beads and rotated end-over-end at 4° C overnight. 80ul of the pre-cleared lysate was added to 120ul of 1x TE buffer with 1% SDS and frozen overnight to be the input sample.

The IP was washed two times with a low salt buffer (10mM Tris-HCl pH 8.0, 2mM EDTA, 150mM NaCl, 1.0% Triton X-100, 0.1% SDS), two times with a high salt buffer (10mM Trish-HCl pH 8.0, 2mM EDTA, 500mM NaCl, 1.0% Triton X-100, 0.1% SDS), two times with LiCl wash buffer (10mM Tris-HCl pH 8.0, 1mM EDTA, 250mM LiCl (Sigma), 0.5%

NaDeoxycholate, 1.0% NP40), and one time with TE (10mM Tris-HCl pH 8.0, 1mM EDTA) buffer. The samples were eluted from the beads by incubating with 100ul of 1x TE and 1% SDS in an Eppendorf thermomixer R at 65° C for 30 minutes, mixing at 1400rpm. This was repeated for a total of 200ul of eluate. The samples and input were then de-crosslinked by incubating in a thermocycler (T1000 Bio-Rad) for 16 hours at 65° C. The samples were incubated with 10ug of RNAseA (Invitrogen) at 37° C for 30 minutes. The samples were then incubated with 140ug of Proteinase K (NEB) at 55° C in a thermomixer mixing at 900rpm for two hours. The DNA was extracted using phenol/chloroform/isoamyl alcohol (Ambion) and cleaned up using Qiagen PCR purification kit and eluted two times using 30ul of buffer EB for a total of 60ul. The concentration was assessed using the highsensitivity DNA kit for qubit (Thermo Fisher Scientific). A portion of the input DNA was run on a 2% agarose gel post stained with ethidium bromide to check the DNA fragmentation.

ChIP qPCR

Primers were designed to amplify the region around the *Gtf2ird1* transcription start site (TSS), which has been shown to be a target of Gtf2ird1 binding (66). Two primer sets were also designed to amplify off target regions, one 10kb upstream of the *Bdnf* TSS and one 7Kbp upstream of the *Pcbp3* TSS. These were far enough away from any TSS that it would be unlikely that there would be a promoter region. The primers can be found in Supplemental Table 5. A standard curve was made by diluting the input sample for each IP sample 1:3, 1:30, and 1:300. The input, the input dilutions, and the IP samples for each genotype condition were run in triplicate using the Sybr green Power UP mastermix (AppliedBiosystems) and primers at a final concentration of 250nM. The PCR was carried out in a QuantStudio6Flex machine (Applied Biosystems) with the following cycling conditions: 50° C for 2 minutes, 95° C for 10 minutes, 95° C 15 seconds, 60° C for 1 minute, repeat steps 3 through 4 40 times. Relative concentrations for the IP samples were determined from the standard curves for that sample and primer set. The on target relative concentration for each genotype was divided by either off target relative concentration to determine the enrichment of Gtf2ird1 binding.

Hippocampus RNA-sequencing

Library preparation

The hippocampus was dissected from adult animals of the second behavior cohort (**Table1**). We used six animals of each genotype, three males and females of the WT and CD animals and two males and four females of the *Gtf2i** genotype. The hippocampus was homogenized in 500ul of 1x RIPA supplemented with two RNAse inhibitors, RNAsin and SUPERase In, and 250ul of the homogenate was added to 750ul of Trizol LS and stored at -80° C until RNA extraction. RNA was extracted using the Zymo clean and concentrator-5 kit following the on column DNAse I digestion protocol and eluted in 30ul of water. The quality and concentration of the RNA was determined using a nanodrop 2000 and Agilent RNA Highsenstivity Tape screen ran on the TapeStation 2000 (Agilent). All RINe scores were above seven.

1ug of RNA was used as input and rRNA was depleted using the NEBNext rRNA Depletion kit (Human/Mouse/Rat). RNA-seq libraries were prepared using the NEB Next Ultra II RNA library Prep Kit for Illumina. The final uniquely indexed libraries for each sample were amplified using the following thermocycler conditions: 98° C for 30 seconds, 98° C 10 seconds, 65° C 1 minute and 15 seconds, 65° C 5 minutes, hold at 4° C, repeat steps 2 through 3 6 times. Each sample had a unique index. Samples were pooled at equal molar amounts and 1x50 reads were sequenced on one lane of a HiSeq3000 at the Genome Technology Access Center at Washington University School of Medicine. The RNA-seq data is available at GEO with accession number (submitted, waiting on accession number).

RNA-seq analysis

The raw reads were trimmed of Illumina adapters and bases with base quality less than 25 using the Trimmomatic Software (161). The trimmed reads were aligned to the mm10 mouse genome using the default parameters of STARv2.6.1b (169). Samtools v1.9 (141) was used to sort and index the aligned reads. Htseq-count v0.9.1 (170) was used to count the number of reads that aligned to features in the Ensembl GRCm38 version 93 gtf file.

The htseq output was analyzed for differential gene expression using EdgeR v3.24 (165). Lowly expressed genes were defined as genes that had a cpm less than two across all samples. Lowly expressed genes were then filtered out of the dataset. We used the exactTest function to make pairwise comparisons between the three groups: WT versus *Gtf2i**, WT versus CD, and CD versus *Gtf2i**. Genes were considered differentially expressed if they had an FDR< 0.1.

GO analysis was performed using the goseq R package (188). Nominally significant up and down regulated genes for each comparison were considered differentially expressed genes and the background gene set included all expressed genes after filtering out the lowly expressed genes. The top 10 most significant go terms for each ontology category were reported. To test how unlikely it is to see these go terms given the differentially expressed genes from the genotype comparisons, we shuffled the genotypes among the samples and repeated the differential expression analysis and go term analysis 1000 times and counted how many times the same go terms were identified in the top ten most significant go terms.

Behavioral tasks

Animal statement

All animal testing was done in accordance with the Washington University in St. Louis animal care committee regulations. Mice were same sex and group housed with mixed genotypes in standard mouse cages measuring 28.5 x 17.5 x 12cm with corn cob bedding and ad libitum access to food and water in a 12 hour light dark cycle, 6:00am-6:00pm light. The temperature of the colony rooms was maintained at 20-22° C and relative humidity at 50%. Two cohorts of mice were used in the behavior and RNA-seq experiments. The CD animals (Del (5Gtf2i-Fkbp6)1Vcam) were a gift from Dr. Victoria Campuzano and have been previously described (93) and were maintained on the C57BL/6J strain (https://www.jax.org/strain/000664). The first behavior cohort (**Table 1**) used *Gtf2i** and CD females as breeders. The second behavior cohort (**Table 1**) used just CD female breeders as male CD animals were frequently not successful at breeding. Male and female mice were included in the behavior tasks. Experimenters were blind to genotype during all testing. Besides the maternal separation induced pup ultrasonic vocalization, all behaviors were done in adult animals older than 60 days and less than 150 days old. Mice were moved to the testing facility at least 30 minutes before the test to allow the mice to habituate to the room. The male experimenter was present during this habituation so the mice could also acclimate to the experimenter. Sex differences were assessed in all experiments, and are discussed when they were significant. Otherwise, the data is presented with the males and females pooled. Animals were removed from analysis if they were outliers, defined as having values greater than 3.5 standard deviations above or below the mean for their genotype group. Animals were also removed if the video and tracking quality were too poor to be analyzed. All filtering was conducted blind to genotype.

Maternal separation induced pup ultrasonic vocalization

To assess early communicative behaviors we performed maternal separation induced pup ultrasonic vocalization (USVs). Animals were recorded on postnatal day three and postnatal day five, days when FVB/AntJ animals begin to make the most calls (177). The parents were placed in a new cage, and the home cage containing the pups was placed in a warming box (Harvard Apparatus) set at 33° C for at least 10 minutes prior to the start of recording. Pups were individually placed in an empty standard-mouse cage (28.5 x 17.5 x 12cm) located in a MDF sound-attenuating box (Med Associates) that measures 36 x 64 x 60cm. Prior to recording, the pup's skin temperature was recorded using a noncontact HDE Infrared Thermometer, as it has been shown that decreased body temperature elicits increased USVs (189). There was no difference in body temperature between genotypes ($F_{2,61}$= 2.521, p=0.089)(**Supplemental Table 1**). USVs were detected using an Avisoft UltraSoundGate CM16 microphone placed 5cm above the bottom of the cage, Avisoft UltraSoundGate 416H amplifier, and Avisoft Recorder software (gain=3dB, 16bits, sampling rate =250kHz). Animals were recorded for 3 minutes, weighed, checked for detachment of pinnae, and then placed back into the home cage in the warming chamber. After all animals had been recorded the parents were returned to the home cage. Sonograms of the recordings were prepared in MATLAB (frequency range =25-120kHz, FFT [Fast Fourier Transform] size=512, overlap=50%, time resolution =1.024ms, frequency resolution = 488.2Hz) along with number of syllables and spectral features using a previously published protocol (177, 190) based on validated methods (191).

Sensorimotor battery

We assessed motoric initiation, balance, coordination, and strength as described in (171, 192) over two days using the following tasks: day 1) walking initiation, ledge, platform, pole; day 2) 60 screen, 90 screen, and inverted screen. Each task was performed once then the animals were allowed a 20 minute break then the tests were repeated in reverse order to control for practice effects. The two trials for each task were then averaged to be used in analysis. Walking initiation was tested by recording the time it takes for the mouse to exit a demarcated 24 x 24cm square on top of a flat surface. To assess balance, the mice were placed on a plexiglass ledge with a width of 0.5cm and a height of 38cm. We recorded how long the mouse balanced on the ledge up to 60 seconds. Another test of balance required the mouse to balance on a wooden platform measuring 3.0cm in diameter, 3.5cm thick and was 25.5cm high. The amount of time the animal balanced on the platform was recorded up to 60 seconds. Motor coordination was tested by placing the mouse at the top of a vertical pole with the head facing upward. The time it took the mouse to turn so the head was facing down was recorded as well as the time it took the mouse to reach the bottom of the pole up to 120 seconds. On day two the mice performed screen tasks that assessed coordination and strength. Mice were placed head facing downward in the center of a mesh wire grid that had 16 squares per 10cm and was 47cm off the ground and inclined at 60 degrees. The time it took the mice to turn and reach the top of the screen was recorded up to 60 seconds. Similarly the mice were placed in the center facing downward of mesh wire screen with 16 squares per 10cm, elevated 47cm from the surface of a utility cart, and inclined at 90 degrees. The time it took the mice to turn around and reach the top was recorded up to 60 seconds. To test strength, the mice were placed in the center of a mesh wire grid used for the 90 screen task and then inverted so the mouse was hanging from the screen that was elevated 47cm. The time the mouse was able to hang onto the screen up to 60 seconds was recorded.

One hour locomotor activity

We tested the animals' general exploratory activity and emotionality in an one hour locomtor activity task (171). Animals were placed in the center of a standard rat cage (47.6 x 25.4 x 20.6cm) and allowed to explore the cage for one hour in a sound-attenuating enclosure with the lightening set to 24 lux. The one hour sessions were video recorded and the animals position and horizontal movements were tracked using the ANY-maze software (Stoelting Co.: RRID: SCR_014289). The apparatus was split into two zones: a 33 x 11cm center zone, and a bordering 5.5cm edge zone. ANY-maze recorded total distance traveled in the apparatus, and total distance traveled, time spent, and entries into each zone. The mouse was considered to have entered a zone when 80% of the body was detected within the zone. The rat cages are thoroughly cleaned with 70% ethanol between mice.

Marble burying

Marble burying is a task that measures the natural digging behavior of mice and is correlated to compulsive behaviors and hippocampal function (179). Following our previously published methods (171), a standard rate cage (47.6 x 25.4 x 20.6cm) was filled with autoclaved aspen bedding to a depth of 3cm and placed in a sound-attenuating enclosure with lighting set to 24 lux. 20 glass marbles were arranged in 5 x 4 grid on the surface of the bedding. Mice were placed in the center of the rat cage and allowed 30 minutes to explore and bury the marbles. The session was recorded using a digital camera and the animals horizontal movements and position in the apparatus were tracked using ANY-maze with the same center and edge zones as described in the one hour activity task. After 30 minutes mice were put back in their home cage and the number of marbles not buried was counted by two observers. A marble was considered buried if 2/3 of the marble was underneath the bedding. The average of the two scorers was used to calculate the average number of marbles buried. The marbles and rat cages were thoroughly cleaned with 70% ethanol between mice.

Three-chamber social approach

To assess voluntary sociability and preference for social novelty we used the three-chamber social approach assay as previously described (171, 193, 194). The task took place in a plexiglass arena with two partitions with rectangular openings (5 x 8cm) dividing the arena into three chambers that each measure 19.5 x 39 x 22cm. The openings could be closed using plexiglass doors that slide into the openings. The task consisted of four consecutive 10 minute trials. During trial one the animals were habituated to the middle chamber with the openings to the side chambers closed. In trial two the animals were allowed to explore the entire apparatus. Trial three was the sociability trial. In one side chamber there was an empty steel pencil cup (Galaxy Pencil/Utility Cup, Spectrum Diversified Designs, Inc.) that was placed upside with an upside clear drinking cup secured to the top to prevent animals from climbing on top of the cup; this was the empty side. In the other side chamber there was an identical pencil cup that housed an age- and sex-matched, sexually naive, unfamiliar C57BL/6J stimulus animal; this was the social side. The pencil cups allowed sniffing behavior to occur and exchange of odor cues, but limited physical contact to prevent aggressive behaviors. The experimental animal was allowed to explore the whole apparatus. The side of the empty cup and social cup were counterbalanced across all the samples. In trial four we tested preference for social novelty. A new stranger stimulus animal was placed in the formerly empty cup. All stimulus animals were habituated to the apparatus and the cups for 10 minutes one day prior to testing. Each stimulus animal was used only once per day. During all trials the task was video recorded and the animal's position, animal's head, and movement was tracked with ANY-maze software. We quantified how much time the animal spent in each chamber, as well as distance traveled and number of entries. A 2cm area around the cups was defined as the investigation zone, and the animal's head was used to determine when it was investigating the stimulus animals or the empty cup. The first five minutes of the social and novelty trials were analyzed because this is when the majority of the social investigation occurs (195). The entire apparatus was thoroughly cleaned after each animal using 2% chlorhexidine (Zoetis). The stimulus cups were cleaned using 70% ethanol.

Modified social approach

To test for habituation to social stimuli over extended amounts of time, we slightly modified the social approach task. We used the same apparatus as described above. We included an additional day of habituation to the apparatus for the experimental animals on the day prior to the actual test to ameliorate novelty induced exploration of the apparatus and to potentiate exploration of the investigation zones. During the habituation day the animals were placed in the center chamber for 10 minutes with the doors to the side chambers closed. Next, the animals were allowed to explore the whole apparatus for 20 minutes. The stimulus animals were habituated to the cups in the apparatus for 30 minutes prior to the test day. Trial one and trial two were the same as the social approach described above. For trial three, the sociability trial, the experimental animals were placed in a cylinder in the center chamber, while the empty cup and stimulus animal cup were being placed in the side chambers. This ensures a random starting direction for the experimental mouse so we could make an unbiased measure of which chamber the experimental mouse chose to enter first. The sociability trial lasted for 30 minutes, in which the experimental animal was allowed to freely explore the apparatus and investigate the empty cup and social cup. The social novelty trial was not conducted.

Tube test of social dominance

The tube test of social dominance tests for social hierarchy behaviors in mice (171, 196). This task took place over five days. Days one and two were habituation trials. During day one, the animals were placed in the left entrance of a clear acrylic tube measuring 3.6cm in diameter and 30cm in length and allowed to walk through the entire tube and exit the tube on the right side. Day two was the same but the mice started on the opposite side of the tube. These two habituation days allow the mice to acclimate to the tube, and potentiates task performance. On each of three consecutive test days, two mice of different genotypes were placed in the entrances to the tube and allowed to meet in the middle, at a clear acrylic partition. When both mice were at the acrylic partition, it was removed and the trial began. The trial ended when one mouse was pushed out or backed out of the tube so that all four paws were out of the tube, or two minutes had passed. The mouse that remained in the tube was considered the dominant winner and the mouse that was no longer in the tube was considered the submissive loser. If both mice were still in the tube after two minutes it was considered a tie. Each mouse was tested only once each day, and the mice were tested against a novel mouse each day. After each test, the tube was cleaned with 2% chlorhexidine (Zoetis) solution. All of the test sessions were recorded using a USB camera connected to a PC laptop (Lenovo). The observer scored the test from the videos.

Rotarod

The accelerating rotarod (Rotamex-5; Columbus Instruments, Columbus, OH) tests motor coordination, motor learning, and balance. We used a previously published rotarod paradigm (172, 197, 198) that tests animals on three conditions: 1) stationary rod 2) continuous rotation and 3) accelerating rotation during three different sessions that were separated by three days to minimize motor learning. During each day the animals had five trials; one stationary trial, two continuous trials, and two accelerating trials. During the stationary trial, the animals were placed on the stationary rod and the time that the animals stayed on the rod was recorded up to 60 seconds. During the continuous trials, the animals were placed on the rod rotating at three rotations per minute. The time the animals stayed on the rotating rod was recorded up to 60 seconds. In the accelerating trials, the animals were placed on the rod that was rotating at two rotations per minute. Once the animals were on the rotating rod, the rod began to accelerate at 0.1rpm and reached 17rpm at 180 seconds. The time the animals stayed on the rod up to 180 seconds was recorded. The two trials for the continuous rotation and accelerating rotation during each session were averaged for analysis. If an animal fell off the rod during any session within the first five seconds, the animal was placed back on the rod and the time was reset up to two times. If the mouse fell off within five seconds on the third try that time was recorded.

Elevated Plus Maze

The elevated plus maze was used to assess anxiety-like behaviors in mice using previously published protocols (152, 194, 199). The apparatus had two closed arms that measured 36 x 6.1 x 15cm, two open arms, and a central platform that measured 5.5 x 5.5cm. The time spent in the open arms was used as a measure of anxiety-like behavior in mice, since mice prefer to be in an enclosed area. Each mouse was tested once per day for three consecutive days. During the test the animals had five minutes to freely explore the apparatus. The animals position, movement, entries into each arm, and time spent in each arm were determined by beam breaks of pairs of photocells arranged in a 16 (x-axis) x 16 (y-axis) grid. Beam breaks were monitored by the Motor Monitor software (Kinder scientific). The test was conducted in the dark with black lights, and was recorded by an overhead digital camera using the night vision setting.

Pre-pulse inhibition (PPI)

To test for normal sensorimotor gating and normal acoustic startle response we performed PPI on the animals. Mice were placed in a cage located on top of a force transducer inside of a sound-attenuating box with a house light on (Kinder Scientific). The force transducer measured the startle response of the animals in Newtons. We used a protocol adapted from (194, 200). The protocol was run using the Startle Monitor II software (Kinder scientific). The protocol started with five minutes of acclimation to the 65dB background white noise, which is played continuously throughout the procedure. After acclimation there were 65 trials that pseudo-randomly alternated between different stimulus conditions, beginning with five consecutive trials of the startle stimulus, which was a 40msec 120dB pulse of white noise. The middle trials cycled through blocks of pre-pulse conditions, blocks of non-startle conditions, where only the background noise is played, and two blocks of startle conditions. Each block consisted of five trials. The testing ended with single trials of pulses played at 80dB, 90dB, 100dB, 110dB, followed by five more startle trials of 120dB. There were three different pre-pulse conditions, where a pulse of 4dB, 8dB, or 16dB white noise above the background sound was played 100msec preceding the 120dB startle stimulus. The average startle response during the middle two blocks of startle trials was considered to be the animal's acoustic startle response(ASR). Each trial measured the startle of the animal for 65msec after the stimulus, and the average force in Newtons across this time was used as the startle response. The pre-pulse inhibition was calculated as the difference of the average ASR and the startle response during the respective pre-pulse trial (PP) divided by the ASR of the startle trials multiplied by 100: ((ASR – PP)/ASR)*100.

Contextual and Cued Fear Conditioning

Contextual and cued fear conditioning were used to assess associative learning and memory. We followed a previously published method (172, 201). The test occurred over three days. A camera placed above the apparatus recorded the session. Freezing behavior during each minute was detected in .75s intervals using the FreezeFrame (Actimetrics, Evanston, IL) software. Freezing behavior was defined as no movement except for normal respiration, and is presented as percent time freezing per minute. During day one, animals were allowed to explore the Plexiglas chamber (26cm x 18cm x 18cm; Med Associates Inc.) with a metal grid floor and a peppermint scent that was inaccessible to the animals. A trial light in the chamber turned on for the duration of the five minute trial. During the first two minutes animals were habituated to the apparatus, and freezing during this time was considered the baseline. An 80db white noise tone was played for 20 seconds at 100 seconds, 160 seconds, and 220 seconds during the test. During the last two seconds of the tone (conditioned stimulus CS) a 1.0mA foot shock (unconditioned stimulus UCS) was delivered. The mice were returned to their home cage at the end of the five minute trial. On day two contextual fear memory was tested. The animals were placed into the same chamber with peppermint scent and the illuminated light and no tone or shock was delivered. Freezing behavior was measured over the eight minute task. The amount of time freezing in the first two minutes on day two was compared to the baseline freezing on day one to test the effects of the contextual cues associated with the UCS from day one. On day threed the animals were placed in a new context, a chamber with black walls, and a partition that creates a triangle shaped area and an inaccessible coconut odor. During this 10 minute task, the trial light was on for the entire duration. The animals explored the apparatus for the first two minutes to determine baseline freezing and then the same 80dB (CS) tone from day one was played for eight minutes. The freezing behavior during this time tested the effects of the CS associated with the UCS shock from day one. Shock sensitivity was tested for each mouse three days after the cued fear test following the procedure previously described in (172). Mice were placed in the chamber with the wire grid floor and delivered a two second shock of 0.05mA. The mA of the shock was increased by 0.05mA up to 1.0mA. At each shock level the animal's behavior was observed and the current level at which the animal flinched, exhibited escape behavior, and vocalized was recorded. Once the animal had exhibited each of the behaviors the test ended. Shock sensitivity assessment served to confirm differences in conditioned fear freezing were not confounded by differences in reactivity to the shock current.

Resident intruder

The resident-intruder paradigm, as described previously (202), was used as a direct social interaction test. Only males were used in this experiment. Male mice were individually housed in standard mouse cages for 10 days. Cages were not changed so the mice could establish a territory. The testing took place over three days in which the home cage of the experimental animal was placed in a sound-attenuating box in the dark with two infrared illuminators placed in the box. A clear Plexiglas covering with holes was placed over the cage to prevent animals from jumping out of the cage. A digital camera using the night vision setting recorded the task. On each day a WT C57BL/6J stimulus animal (intruder), age and sex matched was introduced into the experimental animal's (resident) home cage. The animals were allowed to interact for 10 minutes after which the stimulus animal was removed from the cage. A stimulus animal was only used once per day. The testing was repeated for two more days, during which the experimental animals were paired with novel intruders.

The videos were tracked using Ethovision XT 13 software (Noldus Information Technology) using the social interaction module. This module allows for simultaneous tracking of two unmarked animals. The initial tracking was further corrected manually using the track editing tools, to ensure the head and the tail points were oriented correctly. All of the video tracks were smoothed first with the loess method and then with the minimal distance moved method. The variables of interest were the mean bout of time, frequency, and the cumulative duration of time that the experimental animal's nose was less than 0.6cm from the stimulus animal's nose, interpreted as nose-to-nose sniffing, or when the experimental animal's nose was less than 0.45cm from the tail base of the stimulus animal, interpreted as anogenital sniffing. These distance thresholds were determined by an experimenter blind to genotype, examining the videos using the plot integrated view functionality to ensure that the events called by the software accurately defined the social behavior.

Statistical Analysis

All statistical tests were performed in R v3.4.2. Western blots and qPCR were analyzed using a one factor ANOVA and the post hoc Tukey all pairwise comparison test was used determine differences between groups using the multcomp package (173).

For all behavior tests the data was assessed for univariate testing assumptions of normality and equal variances. Normality was assessed using the Shapiro-Wilkes test as well as manual inspection of qq plots. Equality of variances was tested using the Levene's test. Behaviors that violated these assumptions were analyzed using non parametric tests. Repeated measures were analyzed using linear mixed models with the animal as the random effect. Significance of fixed effects were tested using the Anova function from the Car (203) package in R. Post hoc testing was done using the Tukey HSD test from the multcomp package. Tukey HSD test 'within time point' was used for post hoc repeated measures comparisons, as appropriate. See Supplemental Tables 1 and 6 for descriptions of all statistical tests.

4.6 Acknowledgments

This work was supported by 1R01MH107515 (JDD), and the Autism Science Foundation, and the National Science Foundation Graduate Research Fellowship DGE-1745038 to NDK. We would also like to thank Dr. Victoria Campuzano for sharing the CD mouse line, and the Genome Technology Access Center for technical support, as well as Dr. Beth Kozel for critical advice on this project. We would also like to thank Dr. David Wozniak and the Animal Behavior Core at the Washington University School of Medicine for their time and resources. We would like to thank Rena Silverman for her contribution to the resident-intruder analysis.

4.7 Figures

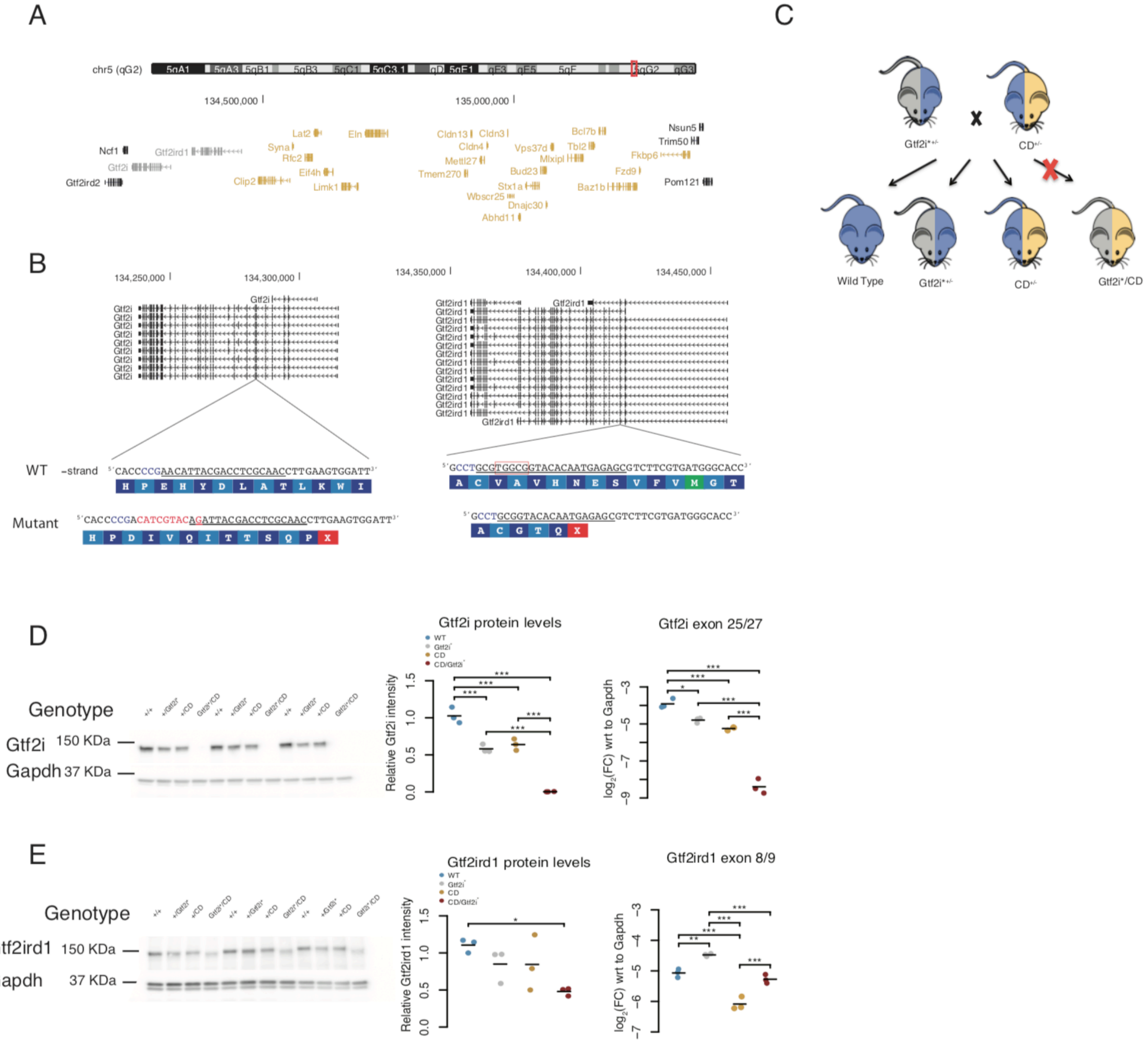


Figure 1. Generation of double mutant *Gtf2i model**. **A** Schematic of the syntenic WSCR in mouse on chromosome 5. The two transcription factors being tested here are highlighted in grey and the genes that are deleted in the CD animals are highlighted in yellow. **B** Gene models of *Gtf2i* and *Gtf2ird1* showing the multiple isoforms of each gene. The WT sequences with the gRNA target underlined and the PAM highlighted in blue with the mutant sequences below along with the corresponding amino acid sequence. **C** Breeding scheme for the behavior tasks **D.** E13.5 whole brain Gtf2i western and qPCR of *Gtf2i** x CD. Gtf2i protein and transcript are similarly reduced in the *Gtf2i** and CD animals. **E** E13.5 whole brain Gtf2ird1 western and qPCR of *Gtf2i** x CD. Gtf2ird1 protein is slightly reduced in the *Gtf2i**/CD brain compared to WT. *Gtf2ird1* transcript is increased in the *Gtf2i** genotype, decreased in the CD genotype, and returns to WT levels in *Gtf2i**/CD genotype. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

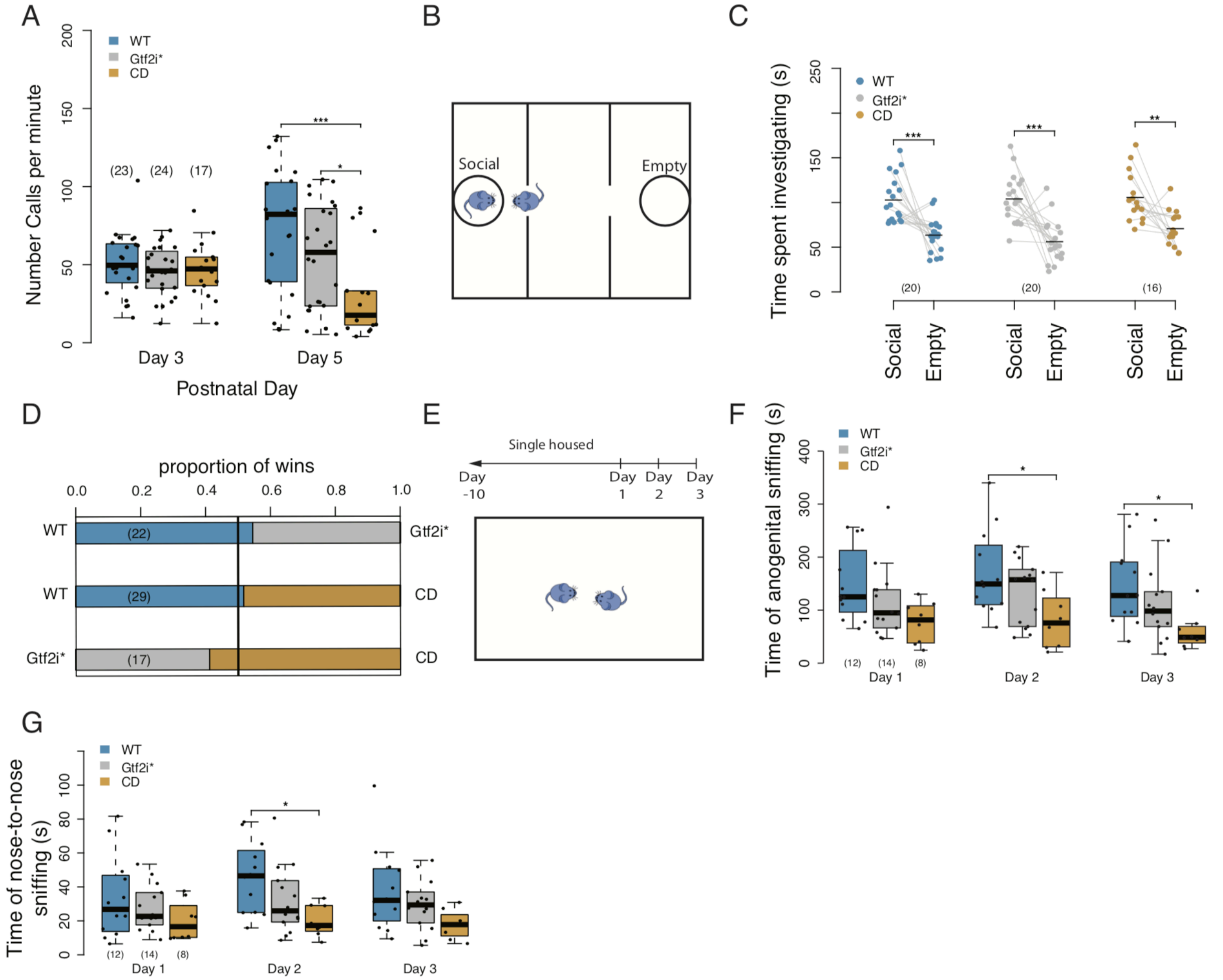


Figure 2. CD mice have deficits in ultrasonic vocalizations and decreased social investigation. **A** Callrate across two days shows that on postnatal day 5 CD animals produce fewer ultrasonic vocalizations than either WT or *Gtf2i** littermates. **B** Schematic of the three-chamber social approach task. **C** All genotypes show preference for social stimulus in three-chamber social approach assay. **D** *Gtf2i** and CD animals show similar dominance behavior to WT animals in the tube test for social dominance. **E** Schematic of the resident intruder paradigm. **F** CD animals show decreased time engaged in anogential sniffing in resident intruder task. **G** CD animals show decreased time engaged in nose-to-nose sniffing in resident intruder task. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ Sample sizes are shown as numbers in parentheses

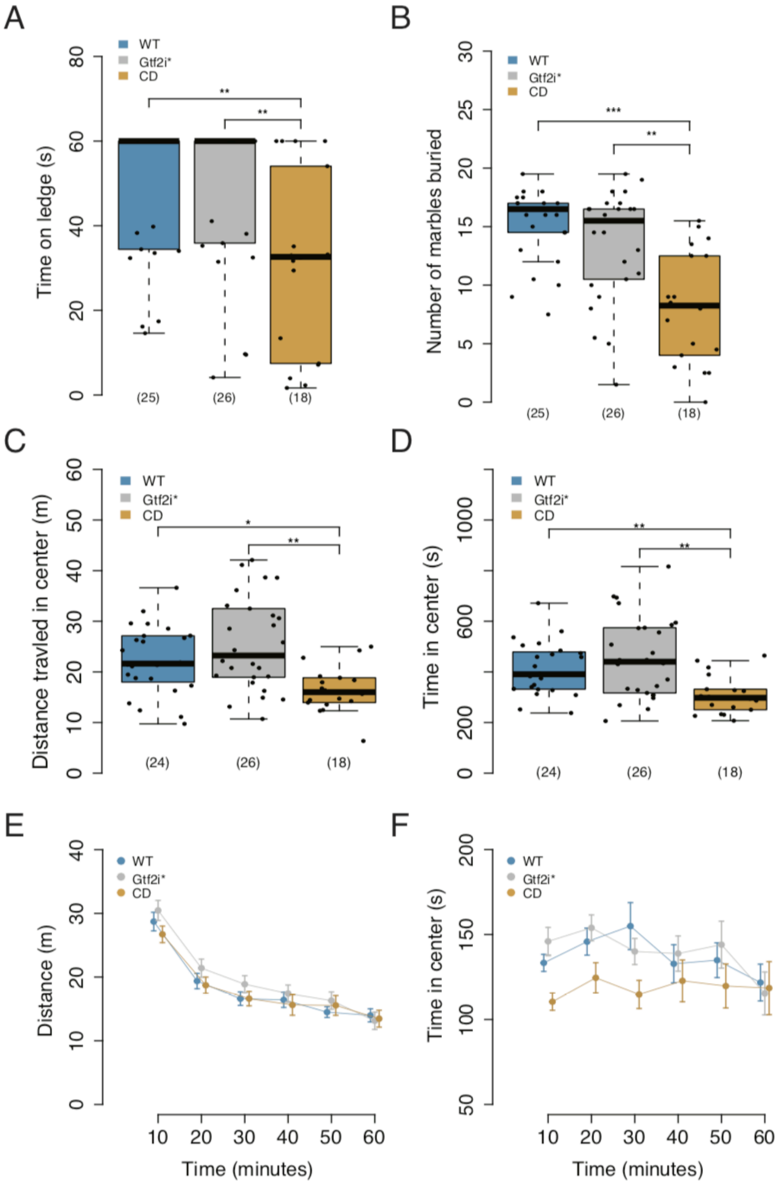


Figure 3. CD mice have motor deficits. **A** CD mice fall off a ledge sooner than WT or *Gtf2i** mutants. **B** CD mice bury fewer marbles than either the WT or *Gtf2i** mutants. **C** CD mice travel less distance in the center during marble burying task **D** CD animals spend less time in the center during marble burying task. **E** All genotypes travel similar distance in open field. **F** All genotypes spend similar time in the center during open field. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ Sample sizes are shown as numbers in parentheses

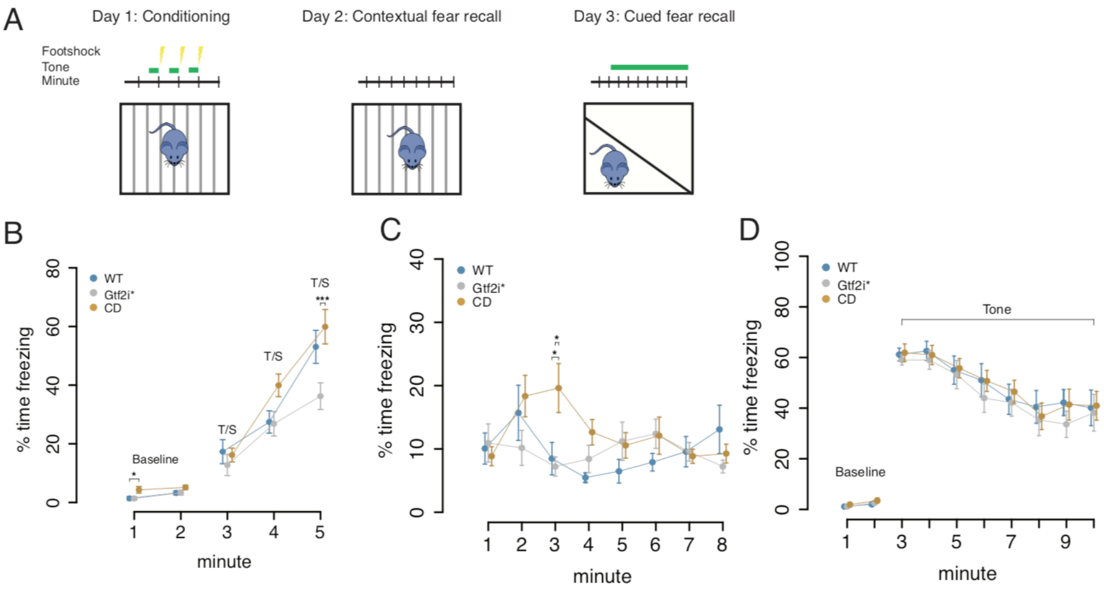


Figure 4. CD mice have more severe contextual fear phenotypes than double mutants. A The conditioned fear task design. Day one animals are delivered a tone and then a footshock throughout the five minute task. Day twp the animals are put in the same context without a footshock to measure contextual fear memory. Day three animals are put in a new chamber and delivered the tone to measure cued fear memory **B** Percent time freezing during conditioned fear acquisition. CD mice have increased baseline freezing during minute one and *Gtf2i** mutants show decreased freezing during minute five **C** Percent time freezing during contextual fear memory recall. CD mice show elevated freezing during fear memory recall. **D** Percent time freezing during cued fear memory recall. All animals show increased freezing when the tone is played. * p < 0.05, ** p < 0.01, *** p < 0.001 Sample sizes are shown as numbers in parentheses

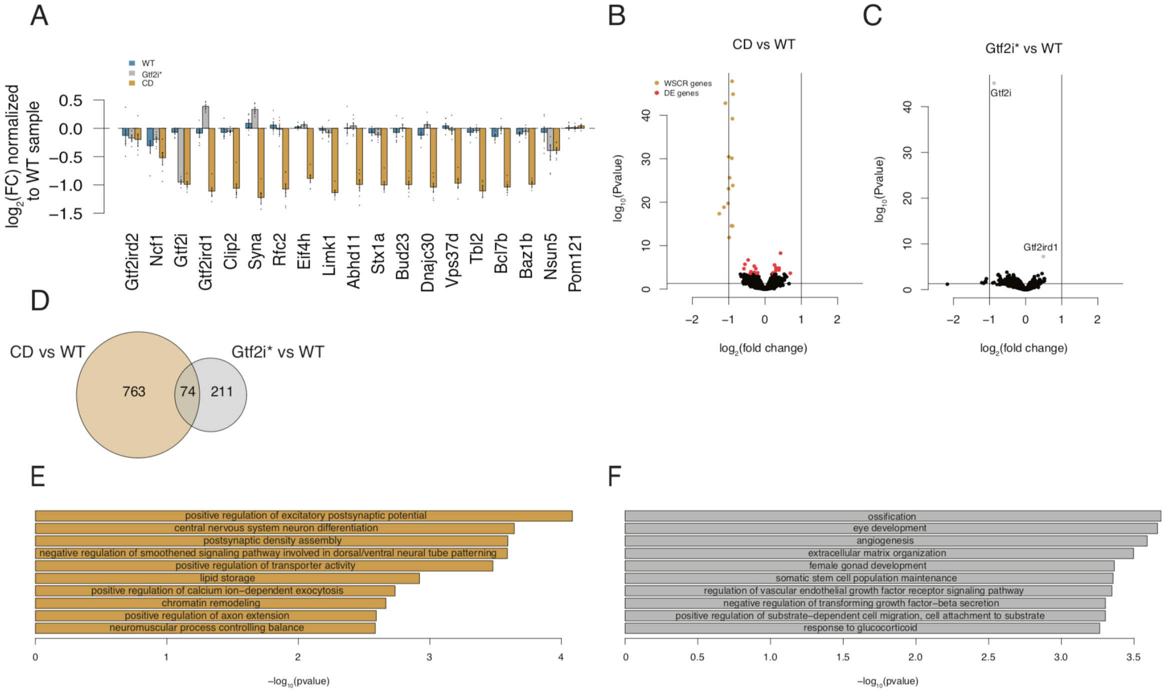
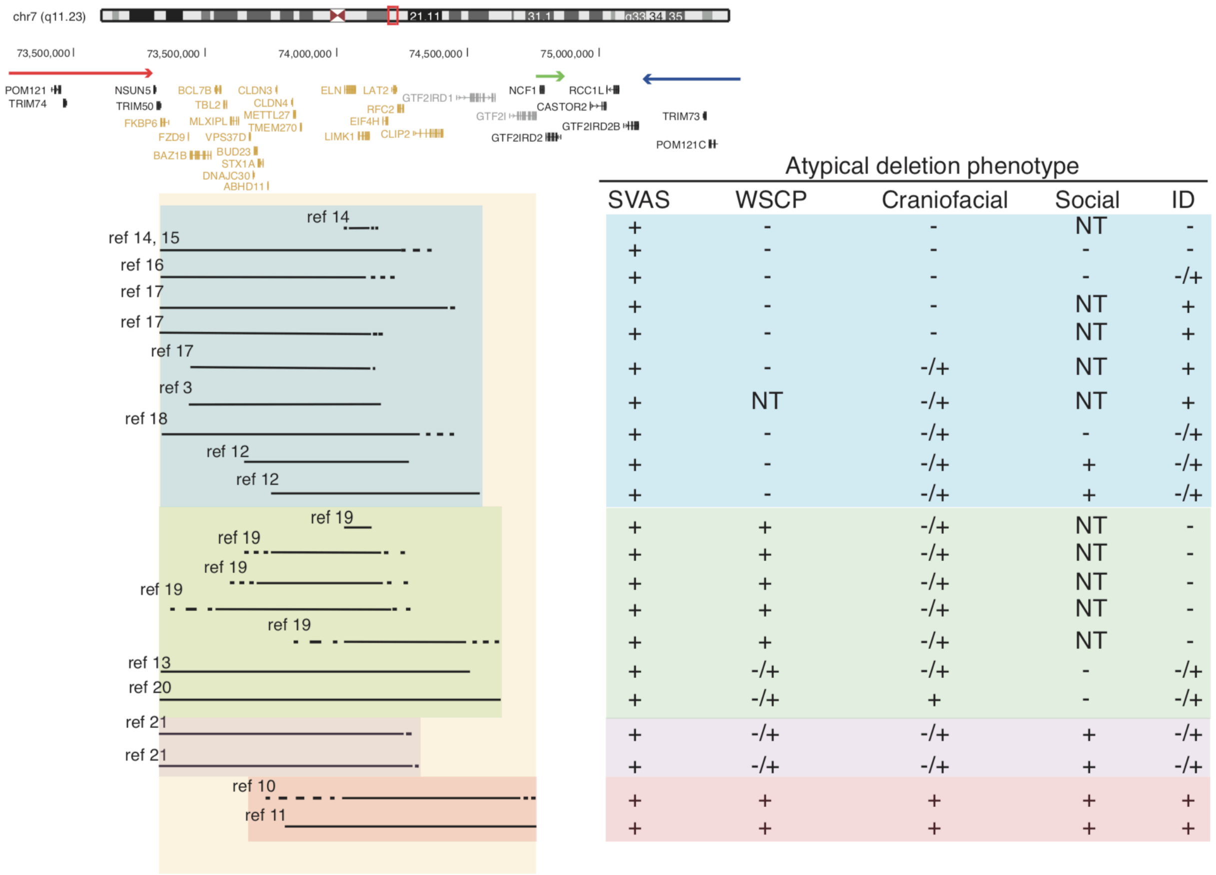
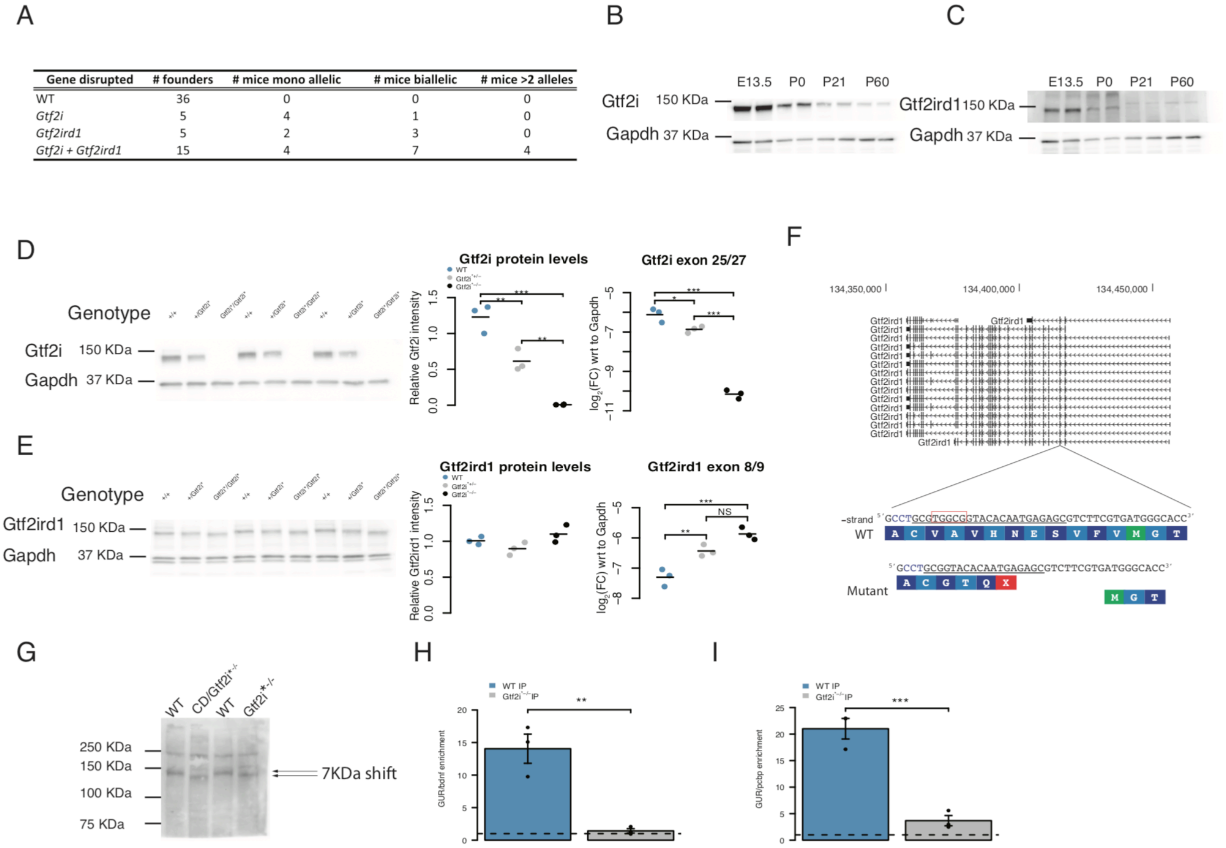


Figure 5. CD mice have altered mRNA for synaptic genes in a hippocampus transcriptome. **A** CD animals show decreased expression of the WSCR that are expressed in the hippocampus. **B** volcano plot comparing CD and WT differentially expressed genes.WSCR genes are highlighted in yellow and genes with FDR < 0.1 are highlighted in red. **C** Besides *Gtf2i* and *Gtf2ird1* there are no significantly differentially expressed genes **D** There is a 9% overlap between nominally significantly up and down regulated genes between CD and *Gtf2i** comparisons to WT controls. **E** CD differentially expressed genes are enriched for GO biological processes involved in synapses and nervous system development. **F** *Gtf2i** differentially expressed genes are enriched for GO biological processed involved in more general organ development.



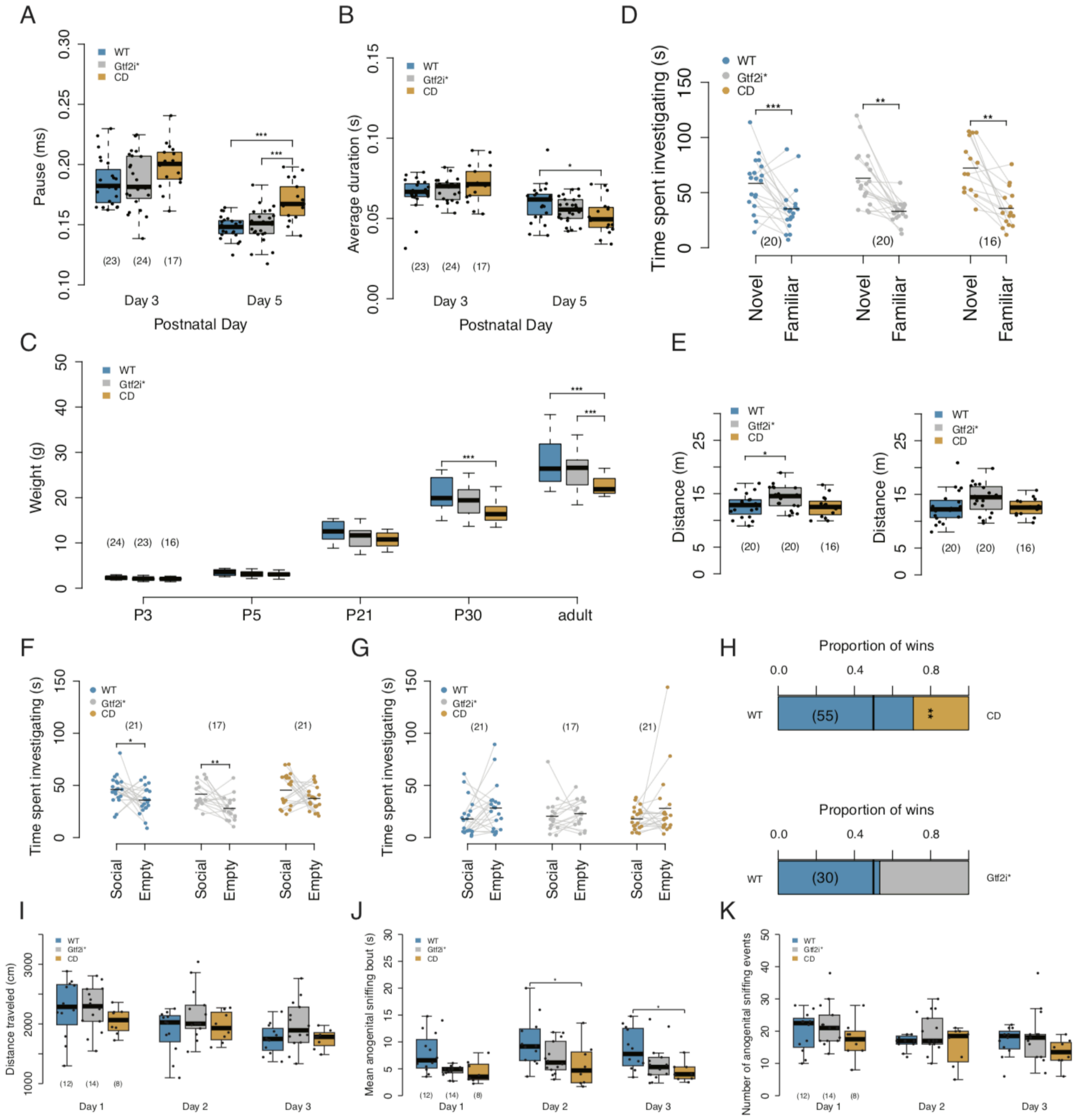
Atypical deletion phenotype				
SVAS	WSCP	Craniofacial	Social	ID
+	-	-	NT	-
+	-	-	-	-
+	-	-	-	-/+
+	-	-	NT	+
+	-	-	NT	+
+	-	-/+	NT	+
+	NT	-/+	NT	+
+	-	-/+	-	-/+
+	-	-/+	+	-/+
+	-	-/+	+	-/+
+	+	-/+	NT	-
+	+	-/+	NT	-
+	+	-/+	NT	-
+	+	-/+	NT	-
+	+	-/+	NT	-
+	-/+	-/+	-	-/+
+	-/+	+	-	-/+
+	-/+	-/+	+	-/+
+	-/+	-/+	+	-/+
+	+	+	+	+
+	+	+	+	+

Figure 6. **Human atypical deletions support oligogenic contribution of genes in the WSCR to phenotypes.** Schematic of the WSCR on chr7q11.23. The arrows indicate the regions of low copy repeats. The typical deletion is demarcated using the yellow box. Atypical deletions demarcated in blue show no contribution to the WSCP. Atypical deletions demarcated in green show contribution to the WSCP. Atypical deletions demarcated in purple provide evidence of deletions that spare *GTF2I* and *GTF2IRD1* that show contributions to across phenotypic domains including social behavior. Atypical deletions demarcated in red provide evidence that the telomeric region is sufficient to produce the full spectrum of phenotypes. The large amount of overlap of all deleted regions and the mild phenotypes present across the atypical deletions suggests an oligogenic pattern. SVAS (supravalvular aortic stenosis), WSCP (Williams syndrome cognitive pfofile) ID (intellectual disability) NT (Not tested), - absent, + present, -/+ milder than typical WS.

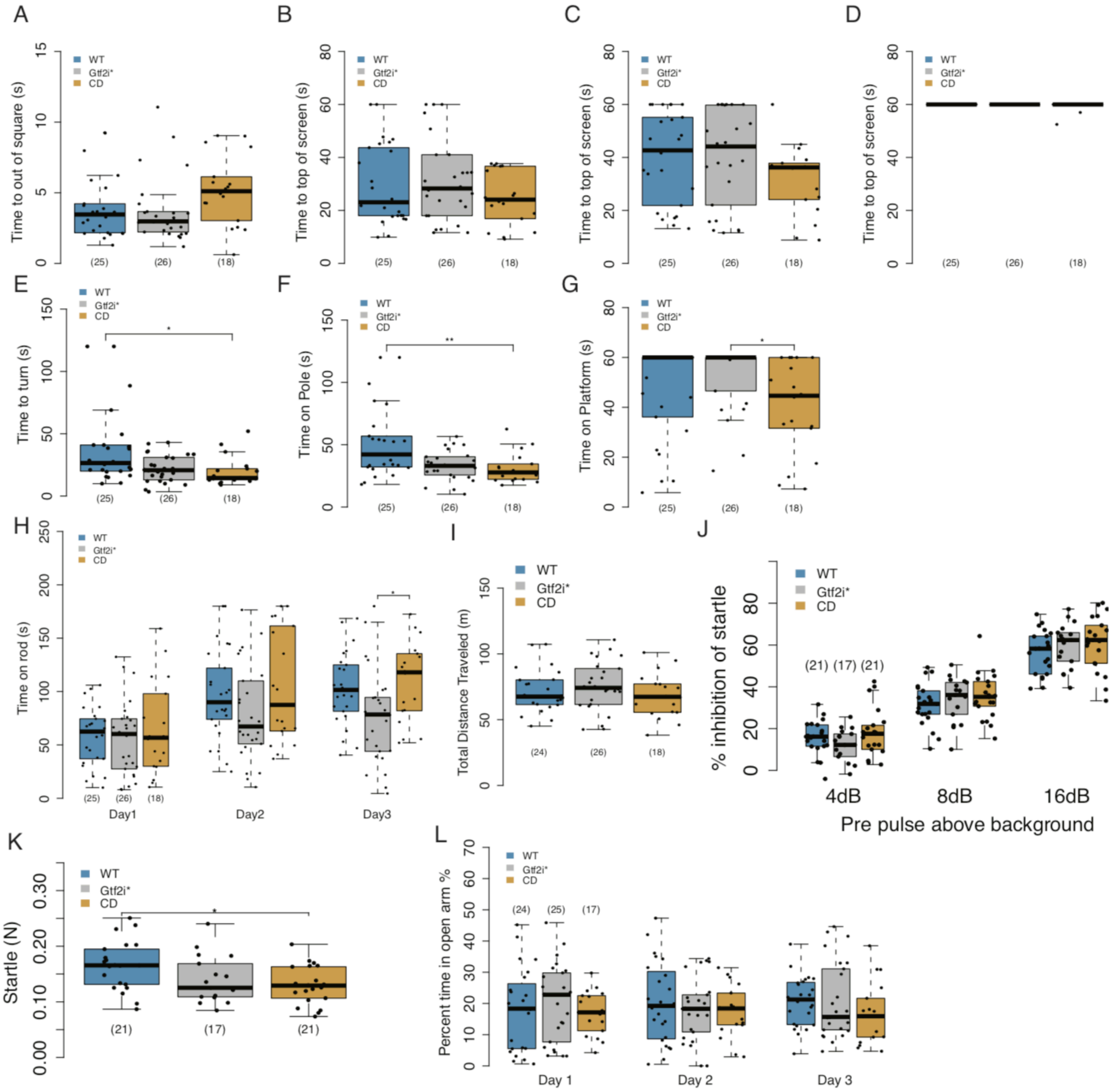


Gene disrupted	# founders	# mice mono allelic	# mice biallelic	# mice >2 alleles
WT	36	0	0	0
Gtf2i	5	4	1	0
Gtf2ird1	5	2	3	0
Gtf2i + Gtf2ird1	15	4	7	4

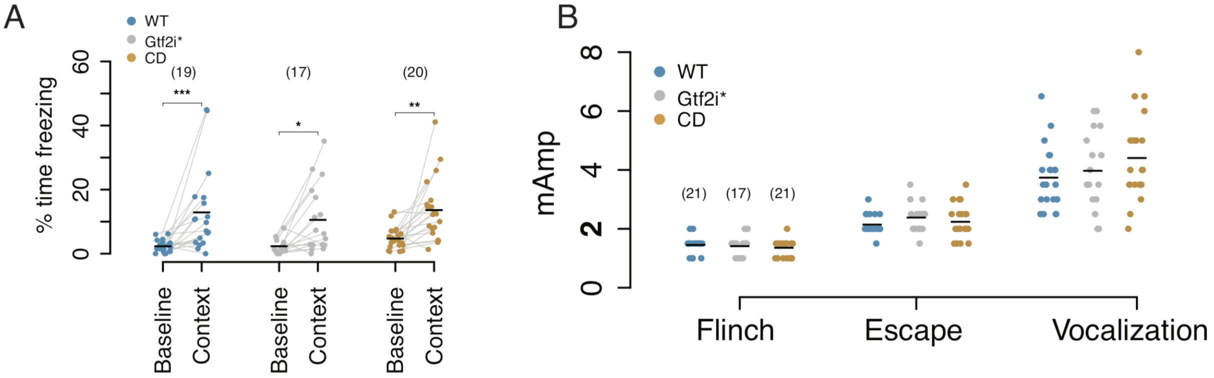
Supplemental Figure 1. Generation of loss of function mutations in Gtf2i and Gtf2ird1. A The number of founders from gRNA injection shows that two gRNAs are efficient at mutating both targets and have high rates of mosaicism. **B** Gtf2i protein is more highly expressed in the embryonic brain and is detectable in the adult brain, each time point includes two biological replicates. **C** Gtf2ird1 protein is more highly expessed in the embryonic brain and not detectable in the adult brain, each time point includes two biological replicates. **D** Gtf2i protein and transcript levels are decreased in the heterozyous Gtf2i* mice and not detectable in the homozygous Gtf2i* E13.5 brain. **E** Gtf2ird1 protein is not decreased in heterozygous or homozygous Gtf2i* E13.5 brain, but the transcript is increased in heterozygous and homozygous animals. **F** Schematic of the consequences of the 5 bp deletion in Gtf2ird1 showing the potential translation re-initation methionine in a new open reading frame. **G** A slight shift of Gtf2ird1 protein in animals homozygous and hemizgyous for the 5 bp deletion in exon 3 of Gtf2ird1, suggesting an N-terminal truncation of Gtf2ird1. **H** ChIP qPCR of the enrichment of the Gtf2ird1 upstream regulatory sequence (GUR) over an off target sequence 7kbp upstream of Bdnf transcription start site in WT versus Gtf2i* homozygous E13.5 brain. **I** ChIP qPCR of the enrichment of the Gtf2ird1 upstream regulatory sequence (GUR) over an off target sequence 10kbp upstream of Pcbp3 transcription start site in WT versus Gtf2i* homozygous E13.5 brain. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$



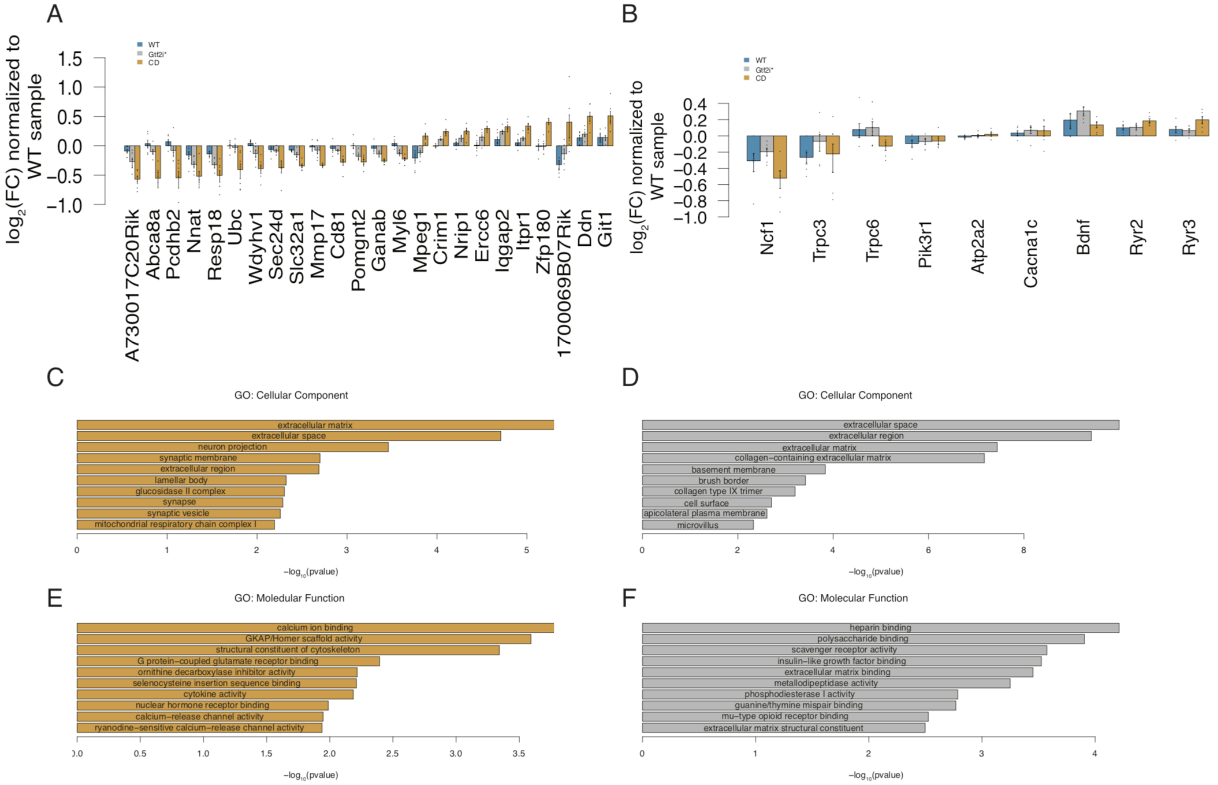
Supplemental Figure 2. Social behaviors in CD and Gtf2i* mutants. A CD animals have increased pauses between bouts of USVs. **B** CD animals have decreased duration of USVs. **C** CD animals have decreased weight in adulthood, developmental weight does not explain differences in USV. **D** All genotypes show preference for social novelty. **E** Double mutants show increased activity in the social approach and social novelty trials of three chambered social appraoch. **F** WT and double mutants show social preference in the first 5 minutes of the extended social approach, but the CD mice are trending. **G** None of the genotypes show preference for social stimulus during the last 5 minutes of the extended social approach. **H** CD mice on C57BL6/J background show a submissive phenotype in tube test of social dominance while the double mutants show no phenotype on FVB/ANTJ background. **I** All genotypes travel similar distance in the resident intruder task. **J** CD animals have decreased mean bout time of anogenital sniffing in the resident intruder task. **K** However all genotypes have similar frequencies of anogential sniffing. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.0001$. Sample sizes are shown as numbers in parentheses



Supplemental Figure 3. Motor and anxiety phenotypes in double mutants and CD animals. A All animals show similar time to initiate walking. **B** All animals reach the top of a 60 degree inverted screen in similar amounts of time. **C** All animals reach the top of a 90 degree inverted screen in similar amounts of time. **D** All animals can hang onto an inverted screen for similar amounts of time. **E** CD animals are able to turn their bodies 180 degrees on a pole quicker than WT animals. **F** CD animals are able to reach the bottom of a pole quicker than WT littermates. **G** CD animals tend to fall off a platform more than double mutants. **H** On day 3 of the rotorod task double mutants fall off sooner than the CD animals. **I** All genotypes travel similar total distances in the marble burying assay. **J** All genotypes show normal PPI. **K** CD animals have decreased startle to 120dB stimulus overall but this is due to decreased weight. **L** All genotypes spend similar amounts of time in the open arm during elevated plus maze. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.0001$. Sample sizes are shown as numbers in parentheses



Supplemental Figure 4. Contextual fear and shock sensitivity in WS mutant models. A All genotypes show a contextual fear response. **B** The response to foot shock is similar across all genotypes. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ Sample sizes are shown as numbers in parentheses



Supplemental Figure 5. Small changes in hippocampal transcriptomes of WS models. A Fold change of differentially expressed genes between WT and CD animals at an FDR < 0.1 normalized to WT levels. **B** Fold change of genes previously tested in CD hippocampus RNAseq from Ortiz-Romero et al. 2018. **C** The top ten enriched Cellular Component gene ontologies for genes that are nominally up or down regulated between CD and WT animals. **D** The top ten enriched Cellular Component gene ontologies for genes that are nominally up or down regulated between Gtf2i* and WT animals. **E** The top ten enriched Molecular Function gene ontologies for genes that are nominally up or down regulated between CD and WT animals. **F** The top ten enriched Molecular Function gene ontologies for genes that are nominally up or down regulated between Gtf2i* and WT animals

Table 1: Behavior and animal cohorts for the *Gtf2i** x CD

Behavior	**Male**			**Female**		
Cohort 1	**WT**	***Gtf2i****	**CD**	**WT**	***Gtf2i****	**CD**
Pup USV P3 and P5	11	12	8	12	12	9
Sensorimotor battery	12	15	7	13	11	11
Elevated plus maze	12	13	7	12	12	10
1 hour locomotor activity	12	14	8	13	12	10
Marble burying	12	14	8	13	12	10
Rotarod	12	14	8	13	12	10
Three-chamber social approach	10	12	6	10	8	10
Resident intruder	12	14	8	NA	NA	NA
Cohort 2						
Modified three-chamber social approach	10	3	9	11	14	12
Tube test of social dominance	11	3	9	11	14	12
Pre-pulse inhibition	10	3	9	11	14	12
Conditioned fear	9	3	8	10	14	12
Shock sensitivity	10	3	9	11	14	12

Supplemental Table 1: Supplemental figures statistic table

Figure		Assay Performed	Parameter (unit)	Comparison			Descriptive Statistics		Statistical Analysis		
				Independent Variables	Age	n(animals)	Average ± SEM	Median (2Q, 3Q)	Statistical Test	Significance	
1	d	Western blotting	Gtf2i protein level in hom and het double mutants compared to WT relative to Gapdh	Gtf2i band density relative to Gapdh	E13.5	WT: 3	WT: 1.23 ± 0.12	WT: 1.16 (0.839, 1.14)	One-way ANOVA; Tukey's HSD multiple comparison test	genotype: F(2,6)=52.949, p=0.0002	WT-HET: p=0.00491**
						HET: 3	HET: 0.61 ± 0.089	HET: 0.54 (0.53, 0.67)			WT-HOM: p <0.001***
						HOM: 3	HOM: 0.007 ± 0.003	HOM: 0.006 (0.004, 0.009)			HET-HOM: p 0.00479**
		qPCR	deltaCT	Gtf2i transcript level relative to Gapdh	E13.5	WT: 3	WT: -6.12 ± 0.204	WT: -5.99 (-6.26, -5.92)	One-way ANOVA; Tukey's HSD multiple comparison test	genotype: F(2,6)=204.73, p=3.012x10-6	WT-HET: p=0.0284**
						HET: 3	HET: -6.87 ± 0.099	HET: -6.83 (-6.94, -6.78)			WT-HOM: p <0.001***
						HOM: 3	HOM: -10.16 ± 0.128	HOM: -10.12 (-10.26, -10.04)			HET-HOM: p < 0.001***
	e	Western blotting	Gtf2ird1 protein level in hom and het double mutants compared to WT relative to Gapdh	Gtf2ird1 band density relative to Gapdh	E13.5	WT: 3	WT: 1.00 ± 0.032	WT: 1.00 (0.977, 1.033)	One-way ANOVA; Tukey's HSD multiple comparison test	genotype: F(2,6)=3.6179, p=0.09315	
						HET: 3	HET: 0.897 ± 0.052	HET: 0.900 (0.853, 0.943)			
						HOM: 3	HOM: 1.1 ± 0.070	HOM: 1.08 (1.046, 1.156)			
		qPCR	deltaCT	Gtf2ird1 transcript level relative to Gapdh	E13.5	WT: 3	WT: -7.30 ± 0.17	WT: -7.25 (-7.43, -7.15)	One-way ANOVA; Tukey's HSD multiple comparison test	genotype: F(2,6)=28.764, p=0.0008	WT-HET: p=0.0092**
						HET: 3	HET: -6.43 ± 0.11	HET: -6.49 (-6.54, -6.35)			WT-HOM: p <0.001***
						HOM: 3	HOM: -5.87 ± 0.12	HOM: -5.91 (-5.98, -5.78)			HET-HOM: p = 0.057 .
	h	qPCR	relative enrichment	Gtf2ird1 on target primer compared to Bdnf off target primer in IP	E13.5	WT: 3	WT: 14.05 ± 2.24	WT: 15.08 (12.42, 16.20)	One-way ANOVA; Tukey's HSD multiple comparison test	genotype: F(2,6) = 22.953, p=0.00155	WT-Gtf2i*: p=0.00168**
						Gtf2i*: 3	Gtf2i*: 1.43 ± 0.34	Gtf2i*: 1.26 (1.1018, 1.70)			
						WT: 3			One sample t-test u=1	t(2)=5.824, p=0.028	
						Gtf2i*: 3			One sample t-test u=1	t(2)=1.262, p=0.3341	
	i	qPCR	relative enrichment	Gtf2ird1 on target primer compared to Pdcp3 off target primer in IP	E13.5	WT: 3	WT: 21.01 ± 1.94	WT: 22.9 (20.03, 22.95)	One-way ANOVA; Tukey's HSD multiple comparison test	genotype: F(2,6) = 69.347, p=7.13e-5	WT-Gtf2i*: p < 0.001**
						Gtf2i*: 3	Gtf2i*: 3.68 ± 0.97	Gtf2i*: 2.96 (2.72, 4.28)			
						WT: 3			One sample t-test u=1	t(2)=10.33, p=0.0092	
						Gtf2i*: 3			One sample t-test u=1	t(2)=2.76, p=0.1094	
2	not shown	Pup Ultrasonic Vocalization	surface body temperature	surface body tempearture (C)	P3	WT: 23	WT: 34.26 ± 0.304	WT: 34.40 (33.10, 35.60)	linear mixed model; Animal id random effect; Anova to test fixed effects	Day: F(1,61)=1.37, p = 0.247 Genotype: F(2,61)=2.521, p=0.089 Interaction:Day*Genotype: F(2,61)=0.261, p=0.771	
						Gtf2i*: 24	Gtf2i*: 34.25 ± 0.234	Gtf2i*: 34.45 (33.58, 35.10)			
						CD: 17	CD: 33.64 ± 0.335	CD: 33.25 (32.60, 34.45)			
					P5	WT: 23	WT: 34.33 ± 0.302	WT: 34.50 (33.20, 35.40)			
						Gtf2i*: 24	Gtf2i*: 34.71 ± 0.264	Gtf2i*: 34.3 (33.78, 35.78)			
						CD: 17	CD: 33.92 ± 0.34	CD: 33.80 (33.2, 34.50)			
	a	Pup Ultrasonic Vocalization	pause between calls (ms)	pause between calls (ms)	P3	WT: 23	WT: 0.1855 ± 0.0041	WT: 0.182 (0.1683, 0.1958)	linear mixed model; Animal id random effect; Anova to test fixed effects; Tukey's HSD multiple comparison within day	Day: F(1,61)=131.94, p< 2.2e-16 Genotype: F(2,61)=11.907, p=3.31e-5 Interaction:Day*Genotype: F(2,61)=0.679, p=0.511	Day3: WT-Gtf2i*: p=1
						Gtf2i*: 24	Gtf2i*: 0.1872 ± 0.0063	Gtf2i*: 0.1814 (0.1731, 0.2068)			Day3 WT-CD: p=0.079
						CD: 17	CD: 0.1990 ± 0.0045	CD: 0.2005 (0.1878, 0.2102)			Day3: Gtf2i*-CD: p=0.09
					P5	WT: 23	WT: 0.1496 ± 0.002	WT: 0.1481 (0.1425, 0.1534)			Day5: WT-Gtf2i*: p=1
						Gtf2i*: 24	Gtf2i*: 0.1496 ± 0.0030	Gtf2i*: 0.1512 (0.1428, 0.1582)			Day5: WT-CD: p <0.0004***
						CD: 17	CD: 0.1699 ± 0.0040	CD: 0.1672 (0.1576, 0.1816)			Day5: Gtf2i*-CD: p=0.0014**
	b	Pup Ultrasonic Vocalization	Averge duration of call bout (s)	duration of call bout	P3	WT: 23	WT: 0.0654 ± 0.0022	WT: 0.066 (0.0636, 0.0717)	linear mixed model; Animal id random effect; Anova to test fixed effects; Tukey's HSD multiple comparison within day	Day: F(1,61)=65.8621, p=2.784e-11 Genotype: F(2,61)=0.1046, p=0.901 Interaction:Day*Genotype: F(2,61)=7.26, p=0.001	Day3: WT-Gtf2i*: p=0.846
						Gtf2i*: 24	Gtf2i*: 0.0681 ± 0.0014	Gtf2i*: 0.070 (0.062, 0.0718)			Day3 WT-CD: p=0.222
						CD: 17	CD: 0.0716 ± 0.0028	CD: 0.0714 (0.0642, 0.0793)			Day3: Gtf2i*-CD: p=0.741
					P5	WT: 23	WT: 0.0598 ± 0.0024	WT: 0.0619 (0.0523, 0.0651)			Day5: WT-Gtf2i*: p=0.473
						Gtf2i*: 24	Gtf2i*: 0.0554 ± 0.0016	Gtf2i*: 0.055 (0.0501, 0.0617)			Day5: WT-CD: p =0.046*
						CD: 17	CD: 0.0515 ± 0.0027	CD: 0.0495 (0.0445, 0.0569)			Day5: Gtf2i*-CD: p=0.659
	c	developmental weight	weight (g)	weight	P3	WT: 24	WT: 2.29 ± 0.0781	WT: 2.315 (1.9, 2.58)	linear mixed model; Animal id random effect; Anova to test fixed effects; Tukey's HSD multiple comparison within day	Day: F(4,240)=161.069, p < 2.2e-16 Genotype: F(2,60)=7.21, p=0.0016 Interaction:Day*Genotype: F(8,240)=6.91, p=3.332e-8	Day3: WT-Gtf2i*: p=1
						Gtf2i*: 23	Gtf2i*: 2.06 ± 0.080	Gtf2i*: 2.13 (1.78, 2.3)			Day3 WT-CD: p=1
						CD: 16	CD: 2.07 ± 0.077	CD: 2.05 (1.95, 2.20)			Day3: Gtf2i*-CD: p=1
					P5	WT: 24	WT: 3.495 ± 0.124	WT: 3.57 (2.82, 4.08)			Day5: WT-Gtf2i*: p=1
						Gtf2i*: 23	Gtf2i*: 3.08 ± 0.13	Gtf2i*: 3.14 (2.61, 3.63)			Day5: WT-CD: p =1
						CD: 16	CD: 3.06 ± 0.13	CD: 3.055 (2.79, 3.32)			Day5: Gtf2i*-CD: p =1
					P21	WT: 24	WT: 12.5 ± 0.41	WT: 12.54 (10.84, 14.71)			Day21: WT-Gtf2i*: p=0.505
						Gtf2i*: 23	Gtf2i*: 11.08 ± 0.457	Gtf2i*: 11.66 (9.255, 12.63)			Day21: WT-CD: p=0.285
						CD: 16	CD: 10.64 ± 0.437	CD: 10.77 (9.42, 11.96)			Day21: Gtf2i*-CD: p=1
					P30	WT: 24	WT: 20.70 ± 0.71	WT: 19.895 (18.4, 24.41)			Day30: WT-Gtf2i*: p=0.405
						Gtf2i*: 23	Gtf2i*: 19.17 ± 0.69	Gtf2i*: 19.41 (16.61, 21.76)			Day30: WT-CD: p < 0.001***
						CD: 16	CD: 17.06 ± 0.738	CD: 16.35 (15.01, 17.72)			Day30: Gtf2i*-CD: p=0.159
					adult	WT: 24	WT: 27.88 ± 1.04	WT: 26.395 (23.90, 31.58)			Day_adult: WT-Gtf2i*: p=0.225
						Gtf2i*: 23	Gtf2i*: 26.11 ± 0.84	Gtf2i*: 23.58 (22.83, 28.31)			Day_adult: WT-CD: p < 0.001***
						CD: 16	CD: 22.05 ± 1.01	CD: 21.855 (21.11, 23.5)			Day_adult: Gtf2i*-CD: p < 0.001***
	not shown	pinnae development	counts of animals	presence or absence of pinae	P3	WT	number present: 1		Chi square test of independence, with p-value simulation for low expected cell number	χ^2=0.7480, p=1	
							number absent: 22				
						Gtf2i*	number present: 1				
							number absent: 23				
						CD	number present: 0				
							number absent: 17				
					P5	WT	number present: 23		Chi square test of independence, with p-value simulation for low expected cell number	χ^2=2.59, p=0.4378	
							number absent: 0				
						Gtf2i*	number present: 22				
							number absent: 2				
						CD	number present: 15				
							number absent: 2				
	d	Three Chamber Assay	Investigation Time (s)	Novel Social Chamber	adult	WT: 20	WT: 57.05 ± 5.36	WT: 61.75 (40.650, 69.350)	linear mixed model; Animal id random effect; Anova to test fixed effects; Tukey's HSD multiple comparison within genotype	Chamber: F(1,53)=50.3816, p=3.137e-9 Genotype: F(2,53)=1.39, p=0.2568 Interaction:Chamber*Genotype: F(2,53)=0.5642, p=0.5722	WT:social-empty: p < 0.001***
						Gtf2i*: 20	Gtf2i*: 62.05 ± 6.49	Gtf2i*: 57.10 (46.12, 77.275)			Gtf2i*: social-empty: p =0.00321**
						CD: 16	CD: 70.65 ± 6.05	CD: 66.80 (50.37, 90.775)			CD: social-empty: p = 0.00121**
				Familiar Social Chamber		WT: 20	WT: 34.01 ± 4.78	WT: 32.15 (24.13, 37.68)			
						Gtf2i*: 20	Gtf2i*: 34.38 ± 2.53	Gtf2i*: 34.20 (28.65, 38.88)			
						CD: 16	CD: 37.22 ± 4.57	CD: 31.65 (23.18, 49.45)			
	e	Three Chamber Assay: social approach trial	Total distance traveled (m)	distance traveled in apparatus	adult	WT: 20	WT: 12.59 ± 0.50	WT: 12.49 (10.62, 13.85)	One-way ANOVA; Tukey's HSD multiple comparison test	Genotype: F(2,53) = 3.98, p=0.024	WT-Gtf2i*: p =0.0305*
						Gtf2i*: 20	Gtf2i*: 14.43 ± 0.51	Gtf2i*: 14.67 (14.67, 16.16)			WT-CD: p=0.95
						CD: 16	CD: 12.82 ± 0.53	CD: 12.73 (11.23, 14.10)			Gtf2i* - CD: 0.0872
	e	Three Chamber Assay: social novelty trial	Total distance traveled (m)	distance traveled in apparatus	adult	WT: 20	WT: 12.55 ± 0.67	WT: 12.12 (10.72, 13.74)	One-way ANOVA; Tukey's HSD multiple comparison test	Genotype: F(2,53) = 2.287, p=0.1115	WT-Gtf2i*: p =0.111
						Gtf2i*: 20	Gtf2i*: 14.29 ± 0.62	Gtf2i*: 14.55 (12.56, 16.35)			WT-CD: p=0.906
						CD: 16	CD: 12.93 ± 0.52	CD: 12.62 (11.41, 13.87)			Gtf2i* - CD: 0.295
	f	Long Three Chamber Assay	Investigation Time (s) First 5 minutes	Social Chamber	adult	WT: 21	WT: 46.14 ± 2.85	WT: 45.80 (39.60, 53.40)	linear mixed model; Animal id random effect; Anova to test fixed effects; Tukey's HSD multiple comparison within genotype	Chamber: F(1,53)=83.20, p=1.89x10-12 Genotype: F(2,53)=1.15, p=0.3239 Interaction:Chamber*Genotype: F(2,53)=0.5845, p=0.561	WT:social-empty: p = 0.033*
						Gtf2i*: 17	Gtf2i*: 41.62 ± 2.87	Gtf2i*: 36.00 (33.60, 52.50)			Gtf2i*: social-empty: p = 0.00726**
						CD: 21	CD: 45.48 ± 3.36	CD: 50.6 (33.50, 57.90)			CD: social-empty: p = 0.134
				Empty Chamber		WT: 21	WT: 35.95 ± 2.78	WT: 36.20 (28.50, 45.30)			
						Gtf2i*: 17	Gtf2i*: 27.98 ± 2.86	Gtf2i*: 25.70 (19.90, 35.70)			
						CD: 21	CD: 37.53 ± 2.47	CD: 37.40 (29.90, 44.20)			
	g	Long Three Chamber Assay	Investigation Time (s) Last 5 minutes	Social Chamber	adult	WT: 21	WT: 17.83 ± 3.70	WT: 9.10 (5.8, 23.0)	linear mixed model; Animal id random effect; Anova to test fixed effects; Tukey's HSD multiple comparison within genotype	Chamber: F(1,56)=4.8227, p=0.03225 Genotype: F(2,56)=0.0523, p=0.949 Interaction:Chamber*Genotype: F(2,56)=0.454, p=0.637	WT:social-empty: p = 0.248
						Gtf2i*: 17	Gtf2i*: 20.42 ± 3.99	Gtf2i*: 17.60 (11.70, 25.00)			Gtf2i*: social-empty: p =0.975
						CD: 21	CD: 17.88 ± 2.31	CD: 14.60 (9.50, 24.700)			CD: social-empty: p = 0.279
				Empty Chamber		WT: 21	WT: 28.35 ± 4.87	WT: 28.60 (8.60, 35.50)			
						Gtf2i*: 17	Gtf2i*: 23.00 ± 3.16	Gtf2i*: 23.00 (13.40, 34.40)			
						CD: 21	CD: 28.00 ± 6.86	CD: 18.80 (11.6, 24.30)			
	h	Tube Test of Social Dominance	Proportion of wins	Proportion of wins	adult	WT vs CD: 55	WT wins: 39		two tailed binomial exact test, null hypothesis probability of success=0.5	p=0.0027	
	h	Tube Test of Social Dominance	Proportion of wins	Proportion of wins	adult	WT vs Gtf2i*: 30	WT wins: 16		two tailed binomial exact test, null hypothesis probability of success=0.5	p=0.856	
	i	Resident Intruder	Total distance traveled (cm)	Day1	adult	WT: 12	WT: 12266.11 ± 140.41	WT: 2286.14 (2083.66, 2652.30)	linear mixed model; Animal id random effect; Two-way Anova to test fixed effects; Tukey's HSD multiple comparison within day	Genotype: F(2,31) = 1.399, p=0.26202 Day: F(2,62)=28.9543, p=1.318e-9 Interaction: geontype*day: F(4,62)=2.2204, p=0.07699	
						Gtf2i*: 14	Gtf2i*: 2258.54 ± 101.55	Gtf2i*: 2296.34 (2046.78, 2562.98)			
						CD: 8	CD: 2049.54 ± 77.08	CD: 2065.12 (1912.83, 2200.41)			
				Day2	adult	WT: 12	WT: 1876.62 ± 107.27	WT: 2026.08 (1758.73, 2129.49)			
						Gtf2i*: 14	Gtf2i*: 2154.05 ± 112.32	Gtf2i*: 2004.34 (1938.57, 2298.71)			
						CD: 8	CD: 1950.31 ± 90.35	CD: 1930.40 (1761.16, 2171.50)			
				Day3	adult	WT: 12	WT: 1752.85 ± 71.10	WT: 1749.29 (1580.85, 1900.69)			
						Gtf2i*: 14	Gtf2i*: 1974.90 ± 107.39	Gtf2i*: 1893.18 (70.53, 133.51)			
						CD: 8	CD: 1751.59 ± 56.56	CD: 1783.10 (1666.36, 1836.28)			
	j	Resident Intruder	Mean bout of anogenital sniffing (s)	Day1	adult	WT: 12	WT: 7.80 ± 1.04	WT: 6.58 (5.27, 9.84)	linear mixed model; Animal id random effect; Two-way Anova to test fixed effects; Tukey's HSD multiple comparison within day	Genotype: F(2,31) = 5.8520, p=0.00698 Day: F(2,62)=4.477, p=0.01526 Interaction: geontype*day: F(4,62)=0.1101, p=0.97855	Day1 WT-Gtf2i*: p=0.308
						Gtf2i*: 14	Gtf2i*: 5.22 ± 0.72	Gtf2i*: 4.91 (4.28, 5.19)			Day1: WT-CD p=0.176
						CD: 8	CD: 4.36 ± 0.70	CD: 3.53 (3.01, 5.68)			Day1: Gtf2i*-CD p=0.993
				Day2	adult	WT: 12	WT: 9.97 ± 107.27	WT: 9.15 (6.57, 12.23)			Day2 WT-Gtf2i*: p=0.170
						Gtf2i*: 14	Gtf2i*: 6.99 ± 0.80	Gtf2i*: 6.16 (4.91, 9.51)			Day2: WT-CD p=0.0464*
						CD: 8	CD: 5.70 ± 1.43	CD: 4.68 (2.42, 7.91)			Day2: Gtf2i*-CD p=0.948
				Day3	adult	WT: 12	WT: 8.67 ± 1.11	WT: 7.76 (5.81, 12.40)			Day3 WT-Gtf2i*: p=0.306
						Gtf2i*: 14	Gtf2i*: 6.08 ± 0.96	Gtf2i*: 5.32 (3.91, 6.96)			Day3: WT-CD p=0.0491*
						CD: 8	CD: 4.43 ± 0.63	CD: 3.98 (3.17, 5.34)			Day3: Gtf2i*-CD p=0.855
				Day1	adult	WT: 12	WT: 19.92 ± 1.76	WT: 22.5 (15.5, 23.5)			
						Gtf2i*: 14	Gtf2i*: 21.79 ± 1.78	Gtf2i*: 21.0 (17.0, 24.25)			

						CD:8	CD: 17.38 ± 2.09	CD: 17.5 (14.0, 19.50)	linear mixed model; Animal id random effect; Two-way Anova to test fixed effects; Tukey's HSD multiple comparison within day	Genotype: F(2,31) = 2.7961, p=0.07648 Day: F(2,62)=4.0732, p=0.02178 Interaction: geontype*day: F(4,62)=0.0819, p=0.98766	
	k	Resident Intruder	Number of anogenital sniffing event	Day2	adult	WT: 12	WT: 17.25 ± 1.00	WT: 17.0 (16.0, 18.25)			
						Gtf2i*:14	Gtf2i*: 19.21 ± 1.58	Gtf2i*: 17.0 (16.00, 23.25)			
						CD:8	CD: 15.5 ± 2.13	CD: 18.5 (11.25, 20.00)			
				Day3	adult	WT: 12	WT: 16.92 ± 1.32	WT: 18.5 (14.75, 19.50)			
						Gtf2i*:14	Gtf2i*: 18.50 ± 2.09	Gtf2i*: 18.0 (13.00, 19.00)			
						CD:8	CD: 13.38 ± 1.48	CD: 13.5 (11.0, 16.25)			
3	a	Sensorimotor battery	Time to leave the area (s)	Walking intitiation	adult	WT: 25	WT: 3.93± 0.445	WT: 3.46 (2.17, 4.225)	One-way ANOVA; Tukey's HSD multiple comparison test	Genotype: F(2,66) = 2.13, p=0.1271	
						Gtf2i*:26	Gtf2i*: 3.73 ± 0.46	Gtf2i*: 2.98 (2.25, 3.65)			
						CD:18	CD: 5.13± 0.576	CD: 5.11 (3.34, 6.04)			
	b	Sensorimotor battery	Time to reach top of screen (s)	60 degree screen	adult	WT: 25	WT: 29.60 ± 3.04	WT: 23.08 (18.02, 43.71)	One-way ANOVA; Tukey's HSD multiple comparison test	Genotype: F(2,66) = 1.19, p=0.312	
						Gtf2i*:26	Gtf2i*: 31.24 ± 3.07	Gtf2i*: 28.29 (18.84, 39.36)			
						CD:18	CD: 24.62 ± 2.43	CD: 24.08 (17.39, 36.27)			
	c	Sensorimotor battery	Time to reach top of screen (s)	90 degree screen	adult	WT: 25	WT: 40.05 ± 3.45	WT: 42.67 (21.86, 55.16)	One-way ANOVA; Tukey's HSD multiple comparison test	Genotype: F(2,66) = 1.59, p=0.2116	
						Gtf2i*:26	Gtf2i*: 39.85 ± 3.50	Gtf2i*: 44.12(24.24, 58.02)			
						CD:18	CD: 31.88 ± 3.08	CD: 36.23 (24.33, 37.62)			
	d	Sensorimotor battery	Time on screen (s)	inverted screen	adult	WT: 25	WT: 60.00 ± 0.00	WT: 60.00 (60, 60)	One-way ANOVA; Tukey's HSD multiple comparison test	Genotype: F(2,66) = 2.52, p=0.088	
						Gtf2i*:26	Gtf2i*: 60.00 ± 0.00	Gtf2i*: 60.00(60, 60)			
						CD:18	CD: 59.414 ± 0.44	CD: 56.75 (31.25, 92.38)			
	e	Sensorimotor battery	Time to reach bottom of pole (s)	Pole	adult	WT: 25	WT: 51.73 ± 5.81	WT: 42.205 (32.14, 56.94)	Kruskal Wallis; Nemenyi tests for multiple comparisons of rank sums	Genotype: H(2)=10.335, p=0.0057	WT-Gtf2i*: p=0.0724
						Gtf2i*:26	Gtf2i*: 33.41 ± 2.30	Gtf2i*: 33.08 (26.40, 40.07)			WT-CD: p=0.0093*
						CD:18	CD: 31.06 ± 2.83	CD: 27.76 (22.32, 34.03)			Gtf2i*-CD:p=0.6126
	f	Sensorimotor battery	Time to turn around on pole (s)	Pole	adult	WT: 25	WT: 37.86 ± 6.11	WT: 26.5 (20, 41)	Kruskal Wallis; Nemenyi tests for multiple comparisons of rank sums	Genotype: H(2)=8.68, p=0.013	WT-Gtf2i*: p=0.1268
						Gtf2i*:26	Gtf2i*: 22 ± 2.15	Gtf2i*: 20.75 (13.63, 30.88)			WT-CD: p=0.018*
						CD:18	CD: 19.81 ± 2.77	CD: 14.5 (13, 21.63)			Gtf2i*-CD:p=0.6087
	g	Sensorimotor battery	Time to fall (s)	Platform	adult	WT: 25	WT: 46.35 ± 3.75	WT: 60.00 (36.11, 60)	Kruskal Wallis; Nemenyi tests for multiple comparisons of rank sums	Genotype: H(2)=7.158, p=0.02791	WT-Gtf2i*: p=0.4879
						Gtf2i*:26	Gtf2i*: 52.90 ± 3.78	Gtf2i*: 60 (49.70, 60)			WT-CD: p=0.2924
						CD:18	CD: 40.07 ± 4.43	CD: 44.64 (31.80, 58.91)			Gtf2i*-CD:p=0.0279*
	h	Accelerating Rotorod	Time on rotorod (s)	Accelerating Rotorod Trial 1	adult	WT: 25	WT: 57.96 ± 5.50	WT: 62.50 (37.00, 74.50)	linear mixed model; Animal id random effect; Three-way Anova to test fixed effects; Tukey's HSD multiple comparison within day	Genotype: F(2,63) = 2.0394, p=0.138 Day: F(2,333)=82.09, p < 2.2e-16 Sex: F(1,63) = 10.023, p=0.0023 Interaction: genotype*sex: F(2,63)=0.8155, p=0.447 Interaction: geontype*day: F(4,333)=3.627, p=0.0066 Interaction: sex*day: F(2,333)=3.174, p=0.0431 Interaction: genotype*sex*day:F(4,333)=0.4374, p=0.782	Trial1:WT-Gtf2i*: p=1
						Gtf2i*:26	Gtf2i*: 55.48 ± 6.97	Gtf2i*: 60.25(28.00, 74.13)			Trial1: WT-CD:p=0.98
						CD:18	CD: 64.64 ± 10.31	CD: 60.00 (58.50, 60)			Trial1: Gtf2i*-CD:p=0.93
				Accelerating Rotorod Trial 2	adult	WT: 25	WT: 96.72 ± 8.44	WT: 90.00 (74.00, 122.00)			Trial2:WT-Gtf2i*: p=0.40
						Gtf2i*:26	Gtf2i*: 78.54 ± 8.69	Gtf2i*: 67.25 (51.88, 107.13)			Trial2: WT-CD:p=0.96
						CD:18	CD: 104.28 ± 12.59	CD: 87.50 (63.63, 160.00)			Trial2: Gtf2i*-CD:p=0.17
				Accelerating Rotorod Trial 3	adult	WT: 25	WT: 102.52 ± 6.83	WT: 101.50 (81.50, 125.00)			Trial3:WT-Gtf2i*: p=0.09
						Gtf2i*:26	Gtf2i*: 75.54 ± 8.68	Gtf2i*: 78.50 (44.13, 94.38)			Trial3: WT-CD:p=0.98
						CD:18	CD: 109.00 ± 8.87	CD: 118.00 (82.125, 134.75)			Trial3: Gtf2i*-CD:p=0.04*
	i	Marble burying	Total distance traveled (m)	Total Distance traveled (m)	adult	WT: 24	WT: 71.60 ± 3.32	WT: 67.55 (61.58, 80.21)	One-way ANOVA; Tukey's HSD multiple comparison test	Genotype: F(2,65) = 0.8974, p=0.4126	
						Gtf2i*:26	Gtf2i*: 75.34 ± 8.68	Gtf2i*: 74.19 (62.42, 88.23)			
						CD:18	CD: 68.29 ± 4.01	CD: 67.38 (55.64, 77.15)			
	j	Pre pulse inhibition	percent startle inhibition	ppi 4dB	adult	WT: 21	WT: 15.63 ± 1.79	WT: 16.16 (11.57, 21.79)	linear mixed model; Animal id random effect; Anova to test fixed effects	Genotype: F(2,56)=0.7742, p=0.4659 PrePulse: F(2,122)=620.61, p < 2e-16 Interaction: Genotype*PrePulse: F(4,112)=1.926, p=0.111	
						Gtf2i*:17	Gtf2i*: 11.04 ± 2.37	Gtf2i*: 12.22 (6.61, 17.46)			
						CD:21	CD: 16.96 ± 2.90	CD: 18.37 (10.05, 21.57)			
				ppi 8dB	adult	WT: 21	WT: 31.39 ± 2.24	WT: 31.84 (26.87, 38.08)			
						Gtf2i*:17	Gtf2i*: 34.01 ± 2.60	Gtf2i*: 36.09 (26.81, 42.09)			
						CD:21	CD: 35.50 ± 2.34	CD: 25.21 (30.72, 42.52)			
				ppi 16dB	adult	WT: 21	WT: 56.38 ± 2.30	WT: 58.26 (46.08, 64.27)			
						Gtf2i*:17	Gtf2i*: 59.67 ± 2.59	Gtf2i*: 62.39 (52.25, 65.92)			
						CD:21	CD: 61.05 ± 2.84	CD: 62.40 (51.25, 69.43)			
	k	Startle	Startle (N)	Startle at 120dB	adult	WT: 21	WT: 0.1651 ± 0.0097	WT: 0.1656 (0.1313, 0.1948)	Two-way ANOVA	Genotype: F(2,55) = 1.48, p=0.2365 Weight: F(1,55)=26.01, p=4.34e-6	
						Gtf2i*:17	Gtf2i*: 0.1403 ± 0.0102	Gtf2i*: 0.1253(0.109, 0.1686)			
						CD:21	CD: 0.1296 ± 0.0343	CD: 0.1292 (0.1066, 0.1631)			
	l	Elevated Plus Maze	Percent time in open arm	Elevate pluse maze Trial 1	adult	WT: 24	WT: 18.05 ± 2.67	WT: 18.34 (5.56, 25.23)	linear mixed model; Animal id random effect; Anova to test fixed effects	Genotype: F(2,63)=0.6351, p=0.5332 Trial: F(2,126)=0.3463, p=0.708 Interaction: Genotype*Trial: F(4,126)=0.4285, p=0.787	
						Gtf2i*:25	Gtf2i*: 20.08 ± 2.49	Gtf2i*: 22.79 (7.66, 29.78)			
						CD:17	CD: 16.97 ± 1.73	CD: 17.14 (11.26, 22.49)			
				Elevate pluse maze Trial 2	adult	WT: 24	WT: 20.16 ± 2.65	WT: 19.23 (8.88, 29.63)			
						Gtf2i*:25	Gtf2i*: 18.11 ± 2.02	Gtf2i*: 18.30 (10.86, 22.76)			
						CD:17	CD: 17.44 ± 2.05	CD: 18.40 (13.14, 23.34)			
				Elevate pluse maze Trial 3	adult	WT: 24	WT: 20.68 ± 1.65	WT: 21.27 (13.29, 26.67)			
						Gtf2i*:25	Gtf2i*: 20.81 ± 2.43	Gtf2i*: 15.68 (11.55, 31.08)			
						CD:17	CD: 16.82 ± 2.40	CD: 15.91 (9.22, 21.66)			
4	a	Conditioned Fear	Contextual Fear memory	average % freezing baseline	adult	WT:19	WT: 2.31 ± 0.44	WT: 2.01 (0.67, 3.44)	linear mixed model; Animal id random effect; Anova to test fixed effects; post hoc comparison of genotypes between context	Genotype: F(2,53)=1.0410, p=0.3602 Context: F(1,53)=36.49, p=1.56e-7 Intearction: genotype*context: F(2,53)=0.2032, p=0.8168	WT: baseline-context:p=0.00035***
						Gtf2i*:17	Gtf2i*: 2.32 ± 0.54	Gtf2i*: 1.34 (0.67, 3.56)			Gtf2i*: baseline-context:p=0.0117*
						CD:20	CD: 4.71 ± 0.77	CD: 4.10 (2.55, 6.43)			CD: baseline-context:p=0.0024**
				average % freezing context first two minutes	adult	WT:19	WT: 12.88 ± 3.00	WT: 9.75 (4.22, 16.63)			
						Gtf2i*:17	Gtf2i*: 10.55 ± 2.56	Gtf2i*: 6.21 (2.67, 17.56)			
						CD:20	CD: 13.60 ± 2.2	CD: 12.33(7.43, 15.64)			
	b	shock sensitivity	mAmp at which behavior occurred	Flinch	adult	WT:21	WT: 1.45 ± 0.06	WT: 1.5 (1.5, 1.5)	Kruskal Wallis	H(2)=1.191, p=0.5513	
						Gtf2i*:17	Gtf2i*: 1.41 ± 0.08	Gtf2i*: 1.5 (1, 1.5)			
						CD:21	CD: 1.36 ± 0.06	CD: 1.5 (1, 1.5)			
				Escape	adult	WT:21	WT: 2.14 ± 0.07	WT: 2 (2, 2.5)	Kruskal Wallis	H(2)=2.615, p=0.2705	
						Gtf2i*:17	Gtf2i*: 2.38 ± 0.12	Gtf2i*: 2.5 (2, 2.5)			
						CD:21	CD: 2.24 ± 0.12	CD: 2 (2, 2.5)			
				Vocalization	adult	WT:21	WT: 3.74 ± 0.23	WT: 3.5 (3, 4)	One-way ANOVA	Genotype:F(2,56)=1.45, p=0.243	
						Gtf2i*:17	Gtf2i*: 3.97 ± 0.32	Gtf2i*: 3.5 (3, 5)			
						CD:21	CD: 4.40 ± 0.32	CD: 4 (3.5, 5)			

Supplemental Table 2: Random GO enrichments for CD-WT comparison

ontology	CD_go_terms	CD_log_p	number of times seen in 1000 random DE lists
CC	extracellular matrix	5.299763321	184
CC	extracellular space	4.709714152	343
CC	neuron projection	3.459856842	28
CC	synaptic membrane	2.699392237	6
CC	extracellular region	2.688790717	337
CC	lamellar body	2.324190197	9
CC	glucosidase II complex	2.304084642	2
CC	synapse	2.286313736	22
CC	synaptic vesicle	2.258909614	4
CC	mitochondrial respiratory chain complex I	2.193353178	28
MF	calcium ion binding	3.774528557	70
MF	GKAP/Homer scaffold activity	3.593061599	59
MF	structural constituent of cytoskeleton	3.342895657	55
MF	G protein-coupled glutamate receptor binding	2.394858721	4
MF	ornithine decarboxylase inhibitor activity	2.217525626	3
MF	selenocysteine insertion sequence binding	2.21096167	1
MF	cytokine activity	2.186438203	20
MF	nuclear hormone receptor binding	1.98850255	4
MF	calcium-release channel activity	1.947608066	8
MF	ryanodine-sensitive calcium-release channel activity	1.939022155	3
BP	positive regulation of excitatory postsynaptic potential	4.084162987	2
BP	central nervous system neuron differentiation	3.642462175	4
BP	postsynaptic density assembly	3.593042668	1
BP	negative regulation of smoothened signaling pathway involved in dorsal/ventral neural tube patterning	3.590087957	3
BP	positive regulation of transporter activity	3.479515877	1
BP	lipid storage	2.921534374	2
BP	positive regulation of calcium ion-dependent exocytosis	2.735510193	6
BP	chromatin remodeling	2.665586516	9
BP	positive regulation of axon extension	2.592558674	1
BP	neuromuscular process controlling balance	2.585323612	4

Supplemental Table 3: Random GO enrichments for *Gtf2i**-WT comparison

ontology	Gtf2i*_go_terms	Gtf2i*_log_p	number of times seen in 1000 random DE lists
CC	extracellular space	9.998750285	343
CC	extracellular region	9.414445338	337
CC	extracellular matrix	7.439719124	184
CC	collagen-containing extracellular matrix	7.170561483	248
CC	basement membrane	3.833411616	163
CC	brush border	3.42003544	18
CC	collagen type IX trimer	3.200408002	16
CC	cell surface	2.708057445	110
CC	apicolateral plasma membrane	2.609972709	11
CC	microvillus	2.328482848	16
MF	heparin binding	4.212498792	69
MF	polysaccharide binding	3.905475505	19
MF	scavenger receptor activity	3.572999245	20
MF	insulin-like growth factor binding	3.524711392	47
MF	extracellular matrix binding	3.450116983	53
MF	metallodipeptidase activity	3.249120219	16
MF	phosphodiesterase I activity	2.785797036	5
MF	guanine/thymine mispair binding	2.769210919	0
MF	mu-type opioid receptor binding	2.526203229	10
MF	extracellular matrix structural constituent	2.49719731	151
BP	ossification	3.685648093	23
BP	eye development	3.661194591	2
BP	angiogenesis	3.590308052	14
BP	extracellular matrix organization	3.496345482	14
BP	female gonad development	3.363933991	7
BP	somatic stem cell population maintenance	3.355276999	8
BP	regulation of vascular endothelial growth factor receptor signaling pathway	3.347530325	4
BP	positive regulation of substrate-dependent cell migration; cell attachment to substrate	3.302955283	25
BP	negative regulation of transforming growth factor-beta secretion	3.302955283	23
BP	response to glucocorticoid	3.263966662	0

Supplemental Table 4: Primers for CRISPR sgRNA, validation, and IVT

target	cloning oligos	PAM
Gtf2i_exon5_up_b	CACCGGTTGCGAGGTCGTAATGTTC	CGG
Gtf2i_exon5_lw_b	AAACGAACATTACGACCTCGCAACC	
Gtf2ird1_exon3_up	CACCGCTCATTGTGTACCGCCACGC	AGG
Gtf2ird1_exon3_lw	AAACGCGTGGCGGTACACAATGAGC	

target	T7 endonuclease assay primers
Gtf2i_exon5_b_F	AGCATAACAGCGTCTGCATT
Gtf2i_exon5_b_R	CACGCGTGGGTCATGCTAAT
Gtf2ird1_exon3_F	TATTGGGCCTCAGTGTTCCC
Gtf2ird1_exon3_R	GTTCCAGGCTGGTCTTGACT

target	IVT primer
T7-gRNA-gtf2iex5b-For	TTAATACGACTCACTATAGGGGGTTGCGAGGTCGTAATGTTC
T7-gRNA-IRD1ex3-For	TTAATACGACTCACTATAGGGGCTCATTGTGTACCGCCACGC
Zhang-IVT-gRNA-Rev	AAAAGCACCGACTCGGTGCC
T7-Zhang-C9WT-For	TAATACGACTCACTATAGGGAGAATGGACTATAAGGACCACGAC
T7-Zhang-C9WT-Rev	GCGAGCTCTAGGAATTCTTAC

Supplemental Table 5: Genotyping and RT-qPCR primers

Target	Pde6b genotyping primers
oIMR2093	AAGCTAGCTGCAGTAACGCCATTT
oIMR2094	ACCTGCATGTGAACCCAGTATTCTATC
oIMR2095	CTACAGCCCCTCTCCAAGGTTTATAG

Target	Gtf2ird1 exon 3 5bp deletion
Gtf2ird1_5bp_del_F	GCTCTCATTGTGTACCGCAGGC
Gtf2ird1_wt_R	ACGCTTTGCTGCAAATGCTTG
Bactin_F	AGAGGGAAATCGTGCGTGAC
Bactin_R	CAATAGTGATGACCTGGCCGT

Target	CD genotyping primer
hprtVcam_F	CTCTGAGGCTTCAAAGGTTC
hprtVcam_R	AATCCAGCTTGTTTGGGCTA

Target	qPCR primers
gapdh_F	AGGTCGGTGTGAACGGATTTG
gapdh_R	GGGGTCGTTGATGGCAACA
Gtf2ird1_ex8/9_F	TTTAACAGCAGATACGCGGAAG
Gtf2ird1_ex8/9_R	CGTAAGTACAGGGTCGCTTGAA
Gtf2i_ex25/27_R	GCACCTCTTCCAAAAGCCCTCCA
Gtf2i_ex25/27_R	GGTCGTTGACCTGCTCCCGC

Target	ChIP enrichment qPCR primers	condition
Gtf2ird1_GUR_F	GGTTCTAATCCGTGGCTGGGG	on target
Gtf2ird1_GUR_R	TTGGCTGTCATTTACATACGGGA	on target
bdnf_us_F	GGCCAAGGTGAATTGGGTAT	off target
bdnf_us_R	TGATGGCAGCAATGTTTCTC	off target
pcbp3_us_F	CCCAAAGGATGATGTGGTTT	off target
pcbp3_us_R	AGGGCACTACACATGCACAC	off target

target	amplicon-sequencing primer
Gtf2i_exon5_seq_F	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTcacatgaacaatctgtgacggg
Gtf2i_exon5_seq_R	ACACTCTTTCCCTACACGACGCTCTTCCGATCTcctgtgccatatgagaagatgc
Gtf2ird1_exon3_seq_F	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTcatagggtactcacggcagaa
Gtfi2rd1_exon3_seq_R	ACACTCTTTCCCTACACGACGCTCTTCCGATCTtccaggctggtcttgacttag

Supplemental Table 6: Main figures statistic table

Figure		Assay Performed	Parameter (unit)	Comparison			Descriptive Statistics		Statistical Analysis		
				Independent Variables	Age	n (animals)	Average ± SEM	Median (2Q,3Q)	Statistical Test	Significance	
1	d	Western blotting	Gtf2i protein level in mutants compared to WT relative to Gapdh	Gtf2i band density relative to Gapdh	E13.5	WT:3	WT: 1.02 ± 0.062	WT: 1.00 (0.97,1.07)	One-way ANOVA; Tukey's HSD multiple comparison test	genotype: F(3,8)=104.65, p=9.293x10-7	WT-Gtf2i*: p< 0.001***
						Gtf2i*:3	Gtf2i*:0.58 ± 0.032	Gtf2i*: 0.551 (0.549, 0.598)			WT-CD: p< 0.001***
						CD:3	CD:0.64 ± 0.044	CD: 0.641 (0.601, 0.678)			WT-Gtf2i*/CD: p <0.001***
						Gtf2i*/CD:3	Gtf2i*/CD 0.002 ± 0.001	Gtf2i*/CD: 0.001 (0.001,0.003)			Gtf2i*-CD: p=0.758
											Gtf2i*-Gtf2i*/CD: p <0.001***
											CD - Gtf2i*/CD: p< 0.001***
		qPCR	delta CT	Gtf2i transcript level relative to Gapdh	E13.5	WT:3	WT: -3.91 ± 0.15	WT: -4.04 (-4.06, -3.83)	One-way ANOVA; Tukey's HSD multiple comparison test	genotype: F(3,8)=180.53, p=1.098x10-7	WT-Gtf2i*: p=0.0119*
						Gtf2i*:3	Gtf2i*: -4.79 ± 0.080	Gtf2i*: -4.73 (-4.84, -4.71)			WT-CD: p <0.001***
						CD:3	CD: -5.25 ± 0.056	CD: -5.21 (-5.28, -5.19)			WT-Gtf2i*/CD: p <0.001***
						Gtf2i*/CD:3	Gtf2i*/CD -8.40 ± 0.23	Gtf2i*/CD: -8.49 (-8.62, -8.23)			Gtf2i*-CD: p=0.1992
											Gtf2i*-Gtf2i*/CD: p <0.001***
											CD - Gtf2i*/CD: p< 0.001***
	e	Western blotting	Gtf2ird1 protein level in mutants compared to WT relative to Gapdh	Gtf2ird1 band density relative to Gapdh	E13.5	WT:3	WT: 1.11 ± 0.053	WT: 1.15 (1.07,1.16)	One-way ANOVA; Tukey's HSD multiple comparison test	genotype: F(3,8)=3.9081, p=0.05467	WT-Gtf2i*: p=0.5373
						Gtf2i*:3	Gtf2i*:0.849 ± 0.13	Gtf2i*: 0.978 (0.783, 0.980)			WT-CD: p=0.5232
						CD:3	CD:0.844 ± 0.22	CD: 0.788 (0.644, 1.02)			WT-Gtf2i*/CD: p=0.0378*
						Gtf2i*/CD:3	Gtf2i*/CD 0.480 ± 0.031	Gtf2i*/CD: 0.505 (0.462,0.510)			Gtf2i*-CD: p=1.00
											Gtf2i*-Gtf2i*/CD: p=0.2616
											CD - Gtf2i*/CD: p=0.2708
		qPCR	delta CT	Gtf2ird1 transcript level relative to Gapdh	E13.5	WT:3	WT: -5.07 ± 0.080	WT: -5.05 (-5.13, -5.00)	One-way ANOVA; Tukey's HSD multiple comparison test	genotype: F(3,8)=60.833, p=7.524x10-6	WT-Gtf2i*: p=0.00516**
						Gtf2i*:3	Gtf2i*: -4.48 ± 0.03	Gtf2i*: -4.47 (-4.50, -4.45)			WT-CD: p <0.001***
						CD:3	CD: 609 ± 0.12	CD: -6.19 (-6.21, -6.02)			WT-Gtf2i*/CD: p=0.38529
						Gtf2i*/CD:3	Gtf2i*/CD -5.27 ± 0.083	Gtf2i*/CD: -5.30 (-5.35, -5.21)			Gtf2i*-CD: p< 0.001***
											Gtf2i*-Gtf2i*/CD: p <0.001***
											CD - Gtf2i*/CD: p< 0.001***
2	a	Pup Ultrasonic Vocalization	USV calls/minute		P3	WT: 23	WT: 50.84 ± 4.11	WT: 49.58 (38.39, 63.36)	linear mixed model; Animal id random effect; Anova to test fixed effects; Tukey's HSD multiple comparison within day	Day: F(1, 116.00)=5.43, p=0.021 Genotype: F(2,60.7)=6.09, p=0.004 Interaction:Day*Genotype: F(2, 61.64)=6.80, p=0.002 Weight: F(1,75.48)=3.95, p=0.05	Day3: WT-Gtf2i*: p=0.952
						Gtf2i*:24	Gtf2i*: 45.57 ± 3.21	Gtf2i*: 46.02 (35.09, 57.56)			Day3: WT-CD: p=0.988
						CD:17	CD: 46.83 ± 4.21	CD: 47.27 (36.55, 54.85)			Day3: Gtf2i*-CD: p=1.00
					P5	WT: 23	WT: 72.17 ± 8.21	WT: 82.24 (39.05, 102.69)			Day5: WT-Gtf2i*: p=0.154
						Gtf2i*:24	Gtf2i*: 54.96 ± 6.89	Gtf2i*: 57.87 (23.61, 85.12)			Day5: WT-CD: p <0.001***
						CD:17	CD: 31.94 ± 7.06	CD: 17.60 (11.28, 33.24)			Day5: Gtf2i*-CD: p=0.045*
	c	Three Chamber Assay	Investigation Time (s)	Social Chamber	adult	WT: 20	WT: 100.31 ± 6.07	WT: 91.30 (79.98, 115.73)	linear mixed model; Animal id random effect; Anova to test fixed effects; Tukey's HSD multiple comparison within genotype	Chamber: F(1,53)=83.20, p=1.89x10-12 Genotype: F(2,53)=1.15, p=0.3239 Interaction Chamber*Genotype: F(2,53)=0.5845, p=0.561	WT social-empty: p< 0.001***
						Gtf2i*:20	Gtf2i*: 104.27 ± 5.82	Gtf2i*: 102.70 (84.125, 120.23)			Gtf2i*: social-empty: p <0.001***
						CD:16	CD: 106.88 ± 6.86	CD: 101.05 (89.03, 126.45)			CD: social-empty: p= 0.00456**
				Empty Chamber		WT: 20	WT: 62.14 ± 4.24	WT: 62.20 (46.68, 73.10)			
						Gtf2i*:20	Gtf2i*: 56.40 ± 5.12	Gtf2i*: 51.90 (42.20, 67.83)			
						CD:16	CD: 69.8 ± 4.97	CD: 65.25 (55.23, 83.83)			
	d	Tube Test of Social Dominance	Proportion of wins		adult	WT vs Gtf2i*: 22	WT wins: 12		two tailed binomial exact test, null hypothesis probability of success=0.5	p=0.8318	
						WT vs CD: 29	WT wins: 15			p=1	
						CD vs Gtf2i*:17	Gtf2i* wins: 7			p=0.6291	
	f	Resident Intruder	Day1	Time of ano-genital sniffing (s)	adult	WT: 12	WT: 148.20 ± 20.03	WT: 125.07 (103.71, 194.42)	linear mixed model; Animal id random effect; Two-way Anova to test fixed effects; Tukey's HSD multiple comparison within day	Genotype: F(2,31)=5.241, p=0.01095 Day: F(2,62)=2.470, p=0.093 Interaction: geontype*day: F(4,62)=0.1095, p=0.978	Day1: WT-Gtf2i*: p=0.709
						Gtf2i*:14	Gtf2i*: 115.13 ± 17.76	Gtf2i*: 94.94 (70.53, 138.49)			Day1: WT-CD: p=0.105
						CD:8	CD: 76.28 ± 13.89	CD: 81.60 (39.24, 106.32)			Day1: Gtf2i*-CD: p=0.671
			Day2	Time of ano-genital sniffing (s)	adult	WT: 12	WT: 169.06 ± 23.12	WT: 149.35 (111.50, 213.62)			Day2: WT-Gtf2i*: p=0.627
						Gtf2i*:14	Gtf2i*: 132.96 ± 16.40	Gtf2i*: 157.42 (71.16, 174.12)			Day2: WT-CD: p=0.0248*
						CD:8	CD: 81.40 ± 19.34	CD: 75.84 (31.57, 114.98)			Day2: Gtf2i*-CD: p=0.374
			Day3	Time of ano-genital sniffing (s)	adult	WT: 12	WT: 144.78 ± 21.51	WT: 127.52 (92.20, 189.48)			Day3: WT-Gtf2i*: p=0.707
						Gtf2i*:14	Gtf2i*: 111.65 ± 19.15	Gtf2i*: 98.26 (70.53, 133.51)			Day3: WT-CD: p=0.0318*
						CD:8	CD: 59.55 ± 12.26	CD: 48.75 (41.22, 66.70)			Day3: Gtf2i*-CD: p=0.363
	g	Resident Intruder	Day1	Time of nose-to-nose sniffing (s)	adult	WT: 12	WT: 33.54 ± 7.05	WT: 26.79 (14.49, 45.73)	linear mixed model; Animal id random effect; Two-way Anova to test fixed effects; Tukey's HSD multiple comparison within day	Genotype: F(2,31)=3.737, p=0.03516 Day: F(2,62)=3.01, p=0.05643 Interaction: geontype*day: F(4,62)=0.8156, p=0.520	Day1: WT-Gtf2i*: p=0.930
						Gtf2i*:14	Gtf2i*: 27.58 ± 3.57	Gtf2i*: 22.69 (18.58, 36.04)			Day1: WT-CD: p=0.435
						CD:8	CD: 19.86 ± 4.10	CD: 16.58 (10.30, 25.93)			Day1: Gtf2i*-CD: p=0.878
			Day2	Time of nose-to-nose sniffing (s)	adult	WT: 12	WT: 45.64 ± 6.11	WT: 46.53 (24.97, 59.57)			Day2: WT-Gtf2i*: p=0.300
						Gtf2i*:14	Gtf2i*: 32.21 ± 5.34	Gtf2i*: 25.91 (19.82, 42.67)			Day2: WT-CD: p=0.0160*
						CD:8	CD: 20.16 ± 3.26	CD: 17.30 (14.62, 28.98)			Day2: Gtf2i*-CD: p=0.537
			Day3	Time of nose-to-nose sniffing (s)	adult	WT: 12	WT: 37.97 ± 7.22	WT: 32.10 (21.91, 50.25)			Day3: WT-Gtf2i*: p=0.750
						Gtf2i*:14	Gtf2i*: 29.51 ± 3.93	Gtf2i*: 29.40 (20.44, 36.13)			Day3: WT-CD: p=0.0950
						CD:8	CD: 17.83 ± 2.96	CD: 17.78 (12.17, 21.87)			Day3: Gtf2i*-CD: p=0.569
3	a	Ledge	Time to fall (s)		adult	WT: 25	WT: 48.83 ± 3.25	WT: 60.00 (34.43, 60)	Kruskal Wallis; Nemenyi tests for multiple comparisons of rank sums	Genotype: H(2)=12.51, p=0.001925	WT-Gtf2i*: p=0.9998
						Gtf2i*:26	Gtf2i*: 48.37 ± 3.57	Gtf2i*: 60 (36.44, 60)			WT-CD: p=0.0071**
						CD:18	CD: 30.98 ± 5.04	CD: 32.64 (8.94, 49.33)			Gtf2i*-CD: p=0.0069**
	b	Marble Burying	Number marbles Buried		adult	WT: 25	WT: 15.14 ± 0.62	WT: 16.50 (14.50, 17.00)	One-way ANOVA; Tukey's HSD multiple comparison test	genotype: F(2,66)=15.243, p=3.61x10-6	WT-Gtf2i*: p=0.381
						Gtf2i*:26	Gtf2i*: 13.56 ± 0.92	Gtf2i*: 15.50 (10.625, 16.50)			WT-CD: p< 0.0001***
						CD:18	CD: 8.11 ± 1.15	CD: 8.25 (4.125, 12.50)			Gtf2i*-CD: p=0.000265***
	c		Distance traveled in the center (m)	Distance traveled in the center (m)	adult	WT: 24	WT: 22.38 ± 1.41	WT: 21.64 (18.33, 27.12)	Kruskal Wallis; Nemenyi tests for multiple comparisons of rank sums	Genotype: H(2)=13, p=0.0015	WT-Gtf2i*: p=0.6617
						Gtf2i*:26	Gtf2i*: 25.45 ± 1.80	Gtf2i*: 23.23 (18.97, 32.17)			WT-CD: p=0.0301*
						CD:18	CD: 16.51 ± 1.08	CD: 16.00 (14.01, 18.73)			Gtf2i* - CD: p=0.002*
	d		Time spent in center (s)	Time spent in center (s)	adult	WT: 24	WT: 411.78 ± 21.26	WT: 390.75 (332.28, 476.50)	Kruskal Wallis; Nemenyi tests for multiple comparisons of rank sums	Genotype: H(2)=14.389, p=0.00075	WT-Gtf2i*: p=0.9058
						Gtf2i*:26	Gtf2i*: 455.24 ± 31.56	Gtf2i*: 440.25 (319.60, 574.18)			WT-CD: p=0.0079*
						CD:18	CD: 309.14 ± 18.00	CD: 297.75 (253.03, 330.80)			Gtf2i* - CD: p=0.0017
	e	One hour activity in Open Field	minutes 1:10	Total Distance traveled (m)	adult	WT: 25	WT: 28.71 ± 1.46	WT: 26.52 (24.83, 33.43)	linear mixed model; Animal id random effect; Anova to test fixed effects	Genotype: F(2,66) = 0.6324, p=0.53449 Time: F(5,330)= 218.1739, p <2e-16 Interaction: genotype*time: F(10,330)=1.7817, p=0.06283	
						Gtf2i*:26	Gtf2i*: 30.46 ± 1.58	Gtf2i*: 29.87 (24.84, 35.21)			
						CD:18	CD: 26.71 ± 1.29	CD: 26.86 (24.67, 29.14)			
			minutes 11:20		adult	WT: 25	WT: 19.37 ± 1.20	WT: 18.62 (15.96, 24.34)			
						Gtf2i*:26	Gtf2i*: 21.40 ± 1.42	Gtf2i*: 20.60 (17.26, 26.32)			
						CD:18	CD: 18.72 ± 1.28	CD: 18.85 (14.32, 21.66)			
			minutes 21:30		adult	WT: 25	WT: 16.61 ± 1.06	WT: 15.35 (12.58, 19.66)			
						Gtf2i*:26	Gtf2i*: 18.88 ± 1.34	Gtf2i*: 19.31 (13.81, 23.53)			
						CD:18	CD: 16.62 ± 1.12	CD: 17.42 (12.67, 19.66)			
			minutes 31:40		adult	WT: 25	WT: 16.43 ± 1.19	WT: 15.93 (12.17, 20.70)			
						Gtf2i*:26	Gtf2i*: 17.41 ± 1.34	Gtf2i*: 15.96 (12.75, 22.52)			
						CD:18	CD: 15.62 ± 1.61	CD: 15.03 (9.86, 19.48)			
			minutes 41:50		adult	WT: 25	WT: 14.50 ± 0.85	WT: 13.99 (11.94, 16.32)			
						Gtf2i*:26	Gtf2i*: 16.31 ± 1.35	Gtf2i*: 14.76 (11.45, 20.81)			
						CD:18	CD: 15.55 ± 1.55	CD: 14.82 (10.75, 18.64)			
			minutes 51:60		adult	WT: 25	WT: 13.98 ± 1.03	WT: 13.81 (9.21, 16.12)			
						Gtf2i*:26	Gtf2i*: 13.13 ± 1.35	Gtf2i*: 12.49 (8.40, 18.05)			
						CD:18	CD: 13.47 ± 1.33	CD: 13.94 (9.25, 15.20)			
	f		minutes 1:10	Time spent in center (s)	adult	WT: 25	WT: 133.32 ± 5.00	WT: 126.1 (118, 149.1)	linear mixed model; Animal id random effect; Anova to test fixed effects	Genotype: F(2,66) = 2.3104, p=0.10720 Time: F(5,330)= 2.9187, p = 0.01356 Interaction: genotype*time: F(10,330)=1.0105, p=0.43410	
						Gtf2i*:26	Gtf2i*: 145.98 ± 8.21	Gtf2i*: 136.75 (121.23, 159.3)			
						CD:18	CD: 110.47 ± 5.01	CD: 111.1 (94.03, 128.4)			
			minutes 11:20		adult	WT: 25	WT: 145.83 ± 7.97	WT: 148.1 (114.1, 168.6)			
						Gtf2i*:26	Gtf2i*: 153.96 ± 7.62	Gtf2i*: 153.65 (129.7, 171.57)			
						CD:18	CD: 124.53 ± 8.89	CD: 120.9 (102.43, 148.65)			
			minutes 21:30		adult	WT: 25	WT: 154.97 ± 13.82	WT: 140 (117.2, 177.9)			
						Gtf2i*:26	Gtf2i*: 140.05 ± 7.68	Gtf2i*: 137.00 (104.58, 172.6)			
						CD:18	CD: 114.74 ± 8.25	CD: 113.05 (88.85, 141.73)			
			minutes 31:40		adult	WT: 25	WT: 132.78 ± 11.22	WT: 129.1 (93.9, 156.2)			
						Gtf2i*:26	Gtf2i*: 138.82 ± 10.38	Gtf2i*: 136.45 (111.55, 172.4)			
						CD:18	CD: 122.76 ± 12.29	CD: 108.75 (90.23, 155.58)			
			minutes 41:50		adult	WT: 25	WT: 134.91 ± 10.24	WT: 122.4 (98.1, 158.00)			
						Gtf2i*:26	Gtf2i*: 144.05 ± 13.81	Gtf2i*: 145.65 (109.45, 187.8)			
						CD:18	CD: 119.71 ± 12.93	CD: 110.75 (78.48, 151.13)			
			minutes 51:60		adult	WT: 25	WT: 121.76 ± 10.82	WT: 119.2 (88.6, 154.6)			
						Gtf2i*:26	Gtf2i*: 115.33 ± 12.63	Gtf2i*: 93.9 (82.9, 161.8)			
						CD:18	CD: 105 ± 22.21	CD: 98 (83.58, 134.8)			
	b		Acquisition baseline Percent Freezing	baseline minute 1	adult	WT: 19	WT: 1.38 ± 0.64	WT: 0 (0, 1.33)	linear mixed model; Animal id random effect; Anova to test fixed effects; Tukey's HSD multiple comparison within minute	Genotype: F(2,53)=5.31, p=0.00794 Time: F(1,53)=7.2792, p=0.009335 Interaction: Time*genotype: F(2,53)=0.4134, p=0.6635	WT-Gtf2i*: p=1
						Gtf2i*:17	Gtf2i*: 1.35 ± 0.47	Gtf2i*: 0 (0, 2.65)			WT-CD: p=0.04*
						CD:20	CD: 4.29 ± 1.12	CD: 3.54 (0, 5.42)			Gtf2i*-CD: p=0.05
				baseline minute 2	adult	WT: 19	WT: 3.25 ± 0.72	WT: 2.68 (0.655, 5.11)			WT-Gtf2i*: p=1
						Gtf2i*:17	Gtf2i*: 3.29 ± 0.75	Gtf2i*: 2.67 (1.33, 4.42)			WT-CD: p=0.35
						CD:20	CD: 5.13 ± 0.73	CD: 5.11 (3, 5.79)			Gtf2i*-CD: p=0.404
			Acquisition Conditioned stimulus Percent Freezing	Conditioned Stimulus minute 3	adult	WT: 19	WT: 17.31 ± 4.10	WT: 10.22 (4.43, 25.775)	linear mixed model; Animal id random effect; Anova to test fixed effects; Tukey's HSD multiple comparison within minute	Genotype: F(2,53)=3.43, p=0.0397 Time: F(2,106)=100.307, p <2.2e-16 Interaction: Time*genotype: F(4,106)=3.97, p=0.0048	WT-Gtf2i*: p=0.98
						Gtf2i*:17	Gtf2i*: 12.76 ± 3.61	Gtf2i*: 7.56 (3.57, 15.18)			WT-CD: p=1
						CD:20	CD: 16.181 ± 2.35	CD: 14.63 (10.45, 20.12)			Gtf2i*-CD: p=1
				Conditioned Stimulus minute 4	adult	WT: 19	WT: 27.44 ± 3.81	WT: 26.22 (14.93, 35.55)			WT-Gtf2i*: p=1
						Gtf2i*:17	Gtf2i*: 26.83 ± 4.17	Gtf2i*: 24.89 (14.22, 39.56)			WT-CD: p=0.27
						CD:20	CD: 39.94 ± 3.87	CD: 42.67 (24.33, 50.84)			Gtf2i*-CD: p=0.25
				Conditioned Stimulus minute 5	adult	WT: 19	WT: 53.07 ± 5.63	WT: 47.32 (38.98, 70.54)			WT-Gtf2i*: p=0.065
						Gtf2i*:17	Gtf2i*: 36.25 ± 4.55	Gtf2i*: 39.11 (24.44, 45.54)			WT-CD: p=0.87
						CD:20	CD: 59.97 ± 5.86	CD: 61.16 (49.11, 81.59)			Gtf2i*-CD: p=0.002*
				Contextual Fear minute 1	adult	WT: 19	WT: 10.06 ± 2.46	WT: 5.33 (3.32, 14.38)			min1: WT-Gtf2i*: p=1
						Gtf2i*:17	Gtf2i*: 10.93 ± 3.01	Gtf2i*: 4.44 (1.33, 19.56)			min1: WT-CD: p=1
						CD:20	CD: 8.85 ± 1.51	CD: 9.11 (3.43, 11.61)			min1: Gtf2i* - CD: p=1
				Contextual Fear minute 2	adult	WT: 19	WT: 15.7 ± 4.36	WT: 9.33 (5.33, 18)			min2: WT-Gtf2i*: p=0.92
						Gtf2i*:17	Gtf2i*: 10.17 ± 2.76	Gtf2i*: 5.78 (3.11, 10.22)			min2: WT-CD: p=1
						CD:20	CD: 18.35 ± 3.27	CD: 13.81 (9.34, 22.22)			min2: Gtf2i* - CD: p=0.37

4	c	Conditioned Fear	Contextual Fear recall Percent Freezing	Contextual Fear minute 3	adult	WT: 19	WT: 8.47 ± 2.56	WT: 3.56 (1.33, 11.53)	linear mixed model; Animal id random effect; Anova to test fixed effects; Tukey's HSD multiple comparison within time	Genotype: F(2,53)=1.251, p=0.295 Time: F(7,371)=2.72,p=0.0093 Interaction: genotype*time:F(14,371)=2.499,p=0.0021	min3:WT-Gtf2i*:p=1
						Gtf2i*:17	Gtf2i*: 7.21 ± 1.51	Gtf2i*: 7.11 (1.34, 8.89)			min3:WT-CD:p=0.03*
						CD:20	CD: 19.62 ± 3.86	CD: 14.23 (6.8, 29.57)			min3:Gtf2i*- CD: p=0.01*
				Contextual Fear minute 4	adult	WT: 19	WT: 5.47 ± 0.76	WT: 5.78 (2.89, 7.34)			min4:WT-Gtf2i*:p=1
						Gtf2i*:17	Gtf2i*: 8.42 ± 2.15	Gtf2i*: 6.22 (4, 8)			min4:WT-CD:p=0.55
						CD:20	CD: 12.64 ± 2.00	CD: 10.45 (7.01, 19.00)			min4:Gtf2i*- CD: p=1
				Contextual Fear minute 5	adult	WT: 19	WT: 6.48 ± 1.87	WT: 4.42 (1.33, 6)			min5:WT-Gtf2i*:p=0.97
						Gtf2i*:17	Gtf2i*: 11.21 ± 3.03	Gtf2i*: 5.75 (3.56, 15.56)			min5:WT-CD:p=1
						CD:20	CD: 10.56 ± 2.91	CD: 7.09 (4, 14.89)			min5:Gtf2i*- CD: p=1
				Contextual Fear minute 6	adult	WT: 19	WT: 7.90 ± 1.38	WT: 5.78 (4.44, 9.31)			min6:WT-Gtf2i*:p=0.98
						Gtf2i*:17	Gtf2i*: 12.39 ± 2.30	Gtf2i*: 10.22 (6.67, 17.33)			min6:WT-CD:p=1
						CD:20	CD: 12.09 ± 2.93	CD: 7.34 (4.45, 16.61)			min6:Gtf2i*- CD: p=1
				Contextual Fear minute 7	adult	WT: 19	WT: 9.56 ± 2.40	WT: 6.22 (3.775, 9.78)			min7:WT-Gtf2i*:p=1
						Gtf2i*:17	Gtf2i*: 9.55 ± 1.46	Gtf2i*: 9.78 (4.87, 13.78)			min7:WT-CD:p=1
						CD:20	CD: 8.84 ± 1.11	CD: 7.34(6.21, 11.54)			min7:Gtf2i*- CD: p=1
				Contextual Fear minute 8	adult	WT: 19	WT: 13.08 ± 3.81	WT: 5.36 (2.24, 15.18)			min8:WT-Gtf2i*:p=0.88
						Gtf2i*:17	Gtf2i*: 7.22 ± 1.01	Gtf2i*: 6.7 (4.02, 8.93)			min8:WT-CD:p=1
						CD:20	CD: 9.26 ± 1.48	CD: 9.60 (4.12, 13.06)			min8:Gtf2i*- CD: p=1
	d		Cued Fear memory baseline Percent Freezing	baseline minute 1	adult	WT:19	WT: 1.12± 0.47	WT: 0 (0, 1.77)	linear mixed model; Animal id random effect; Anova to test fixed effects	Genotype:F(2,53)=1.061,p=0.353 Time: F(1,53)=9.037,p=0.004 Interaction: genotype*time:F(2,53)=0.2658,p=0.768	
						Gtf2i*:17	Gtf2i*: 1.12 ± 0.34	Gtf2i*: 0 (0, 1.78)			
						CD:20	CD: 1.91 ± 0.84	CD: 0 (0, 1.77)			
				baseline minute 2	adult	WT:19	WT: 1.99 ± 0.59	WT: 1.33 (0, 3.55)			
						Gtf2i*:17	Gtf2i*: 2.59 ± 0.77	Gtf2i*: 2.22 (0, 2.67)			
						CD:20	CD: 3.49 ± 0.91	CD: 2.44 (0, 5.55)			
			Cued Fear memory Percent Freezing	Cued Fear minute 3	adult	WT:19	WT: 61.87 ± 2.54	WT: 62.22 (54.22, 69.84)	linear mixed model; Animal id random effect; Anova to test fixed effects	Genotype:F(2,53)=0.3014, p=0.741 Time:F(7,371)=21.58,p < 2e-16 Interaction:genotype*time:F(14,371)=0.213,p=0.99	
						Gtf2i*:17	Gtf2i*: 59.01 ± 1.97	Gtf2i*: 59.11 (51.77, 64.44)			
						CD:20	CD: 61.89 ± 3.50	CD: 64.89 (50.95, 70.89)			
				Cued Fear minute 4	adult	WT:19	WT: 62.59 ± 3.80	WT: 63.27 (52, 72.89)			
						Gtf2i*:17	Gtf2i*: 59.04 ± 3.71	Gtf2i*: 60.89 (45.78, 68.89)			
						CD:20	CD: 61.01 ± 3.78	CD: 64.00 (51.11, 69.33)			
				Cued Fear minute 5	adult	WT:19	WT: 55.04 ± 5.58	WT: 53.78 (36.89, 74.67)			
						Gtf2i*:17	Gtf2i*: 53.23 ± 5.56	Gtf2i*: 59.11 (26.22, 70.67)			
						CD:20	CD: 55.76 ± 3.81	CD: 54.46 (48.56, 68.93)			
				Cued Fear minute 6	adult	WT:19	WT: 51.04 ± 6.51	WT: 48 (27.49, 80)			
						Gtf2i*:17	Gtf2i*: 43.97 ± 5.58	Gtf2i*: 38.67 (24, 65.93)			
						CD:20	CD: 50.74 ± 4.16	CD: 54.45 (40.33, 59.37)			
				Cued Fear minute 7	adult	WT:19	WT: 43.52 ± 5.96	WT: 41.41 (22.89, 59.89)			
						Gtf2i*:17	Gtf2i*: 42.16 ± 5.35	Gtf2i*: 38.22 (23.11, 58.85)			
						CD:20	CD: 46.47 ± 4.53	CD: 44.44 (37.68, 58.89)			
				Cued Fear minute 8	adult	WT:19	WT: 40.48 ± 6.50	WT: 32.89 (14.63, 61.56)			
						Gtf2i*:17	Gtf2i*: 35.37 ± 6.23	Gtf2i*: 32.44 (16.89, 57.33)			
						CD:20	CD: 36.79 ± 5.26	CD: 26.66 (21.25, 55.56)			
				Cued Fear minute 9	adult	WT:19	WT: 42.20 ± 5.12	WT: 38.22 (28.61, 54.89)			
						Gtf2i*:17	Gtf2i*: 33.62 ± 5.20	Gtf2i*: 40.44 (13.33, 47.11)			
						CD:20	CD: 41.42 ± 6.11	CD: 41.56 (22.53, 62.71)			
				Cued Fear minute 10	adult	WT:19	WT: 40.11 ± 7.02	WT: 29.46 (16.30, 67.25)			
						Gtf2i*:17	Gtf2i*: 38.18 ± 7.26	Gtf2i*: 26.67 (12.95, 61.61)			
						CD:20	CD: 40.96 ± 5.68	CD: 44.67 (17.74,, 54.24)			

Chapter 5: Conclusions and Future Directions

Nathan Kopp

5.1 Significance

In this thesis I have tested three extant hypotheses in the field of Williams syndrome biology, using both human and mouse genetics. First, I showed that variation on the remaining WSCR allele does not largely modify the social phenotype of individuals with WS as measured by the SRS. The study highlighted two SNPs in *BAZ1B* and *GTF2IRD1*, both of which have been implicated in the cognitive phenotypes of WS, providing further support for their importance in the pathogenesis of WS. I used the data to further describe the genetic variation within the exonic compartment of the WSCR, which can be queried to test for associations with other clinical phenotypes of WS. While 85 individuals is a small sample size to detect variants that have low effect sizes, this was the largest genetic dataset of WS analyzed, and will exist as a foundation to which other larger studies can build.

The second hypothesis I tested was how do the transcription factors *Gtf2i* and *Gtf2ird1* interact to affect behavior. These genes have been thought to contribute to the behavioral, cognitive, and craniofacial aspects of WS, but their affects on behavior have not been studied together. I leveraged the advantages of the mouse model system to study these genes. First, I generated a dataset that describes where these transcription factors bind in the developing brain and then tested the consequences of mutating just *Gtf2ird1* or both transcription factors together to examine how they interact to potentially affect transcription and behavior. Surprisingly, I showed that both transcription factors have little consequence on whole brain transcription, but mutating them still results in behavioral deficits mainly driven by homozygosity of *Gtf2ird1* mutations. The work I have done is some of the first *in vivo* biochemical analysis of Gtf2ird1, and I showed that *Gtf2ird1* is a difficult gene to knockout. These results help interpret the findings of other *Gtf2ird1* mouse models that still show some *Gtf2ird1* transcription and protein product (66, 101). The *Gtf2ird1* mouse model I characterized provided data that supports the functional role of the N-terminal end of the protein in behavior, although it did not result in decreased DNA binding genome wide. My data and methods will be useful to consider when designing future experiments around this gene. I also showed that knocking out *Gtf2i* along with *Gtf2ird1* did not result in more severe phenotypes in the heterozygous state. This suggests that *Gtf2ird1* is the main driver of the phenotypes tested in this study. Overall, I have created two new mouse lines to further model and study Williams syndrome and provided genomic datasets that can be used to generate future hypotheses concerning these two transcription factors.

Finally, I used another *Gtf2i/Gtf2ird1* double mutant mouse model and a mouse model that has the entire WSCR deleted (CD mouse) to test the current leading hypothesis that these two transcription factors are sufficient to replicate the phenotypes that are caused by deleting the whole region. My data suggests that these genes are not sufficient, which implicates the role of other genes in the region or an oligogenic contribution of several genes in the region. I also analyzed the adult hippocampal transcriptome of both mouse models and showed differences in synaptic genes in the CD compared to the double mutant, suggesting that synaptic functioning might be impaired in the CD animals that is not caused by *Gtf2i* or *Gtf2ird1*. These data should encourage studying the effects of other genes in the WSCR. Using the CRISPR/Cas9 technology will allow for the quick generation of mouse models with unique combinations of genes mutated so we can begin to dissect the interactions of the genes in the region, similarly to what has been done for other copy number disorders (102, 103, 204).

Overall, this thesis has generated human and mouse genomic datasets that can be used to design future studies to elucidate genetic influences on WS phenotypes. It also describes three new mouse models that can be used in to further understand how the general transcription factor

2i family contributes to WS phenotypes. Finally, it supports a role for genes outside of the general transcription factor 2i family, encouraging further characterization of single gene knock out mouse models as well as models with combinations of genes knocked out.

5.2 Future directions

5.2.1 Human studies

I described the analysis of the whole exome of 85 individuals with WS and tested for genetics associations with the social phenotype. The exome enriches for variants in the coding regions of genes, which aids in the interpretation of their effects. However, the exome covers only 1% of the genome and with the growing number of whole genome studies, the human genetics field is learning more about the consequences of non coding variation. Thus, it would be beneficial to use whole genome sequencing to analyze how the full spectrum of genetic variation could modify the phenotypes in WS. First, it would be interesting to catalogue the non-coding variation of the WSCR and couple that with the exonic data to look for modifiers within the locus.

Next, we could use the dense genotype data genome wide to calculate polygenic risk scores for different phenotypes of interest within the WS sample. I did this using the Psychiatric Genomics Consortium GWAS on ASD using the whole exome data, but this misses a lot of the common, noncoding variation that was genotyped. Using the whole genome data we could get a better understanding contribution of genomic variants to social behavior. Recently, it has been shown that high polygenic risk scores can convey similar risk to disease as monogenic causes (205). This information could be used to help explain the large variability of the social phenotype and other phenotypes of WS.

Finally, the genome data could be used to identify the breakpoints of the deletion in patients. This is important, because the current diagnostic method, clinical microarray, has difficulty accurately identifying the size of the deleted region due to the low copy repeats. I attempted to use the whole exome data to determine copy number of the *NCF1* alleles by using the ratio of the two base pair deletion that distinguishes the pseudogenes from the functional copy. This gave promising results, but this strategy cannot distinguish the exact break point of the deletion. Since whole genome sequencing provides even coverage it could be used to detect the size of the deletion, which has been shown to affect cognitive and behavioral phenotypes (41). I have preliminarily tried to call the deletion size using the coverage from the whole genome sequencing data. I was able to identify atypical deletions, but the typical deletions all had similar profiles with drops in coverage in the area in the low copy repeats. We could potentially use the polymorphisms that distinguish between the functional and pseudogenes of the regions, but short read whole genome sequencing data may not be able to overcome the challenges of the repetitive regions. Long read technology could be used to try and surmount the difficulty of mapping to the region to better detect the breakpoints.

5.2.2 *Gtf2i* and *Gtf2ird1* mouse studies

I have generated several new mouse models that can be used as tools to understand *Gtf2i* and *Gtf2ird1* biology. One of the more interesting findings from the mutations in *Gtf2ird1* was not expected, and that is that this is a difficult protein to knock out. Two separate frameshift mutations that create premature stop codons with exon three, and a large deletion removing all but 14 base pairs of exon three of *Gtf2ird1* resulted in more *Gtf2ird1* transcript and slightly lower levels of a N-truncated protein. The tight regulation of the transcript and protein levels of *Gtf2ird1* hints at a conserved important function. It would be interesting to further understand how this transcript and protein are regulated. We have data to suggest that *Gtf2ird1* transcripts that contain the mutant alleles are more stable than the WT transcript. Studying the mRNA dynamics and stability of *Gtf2ird1* could provide insight on how it is regulated and lead to further investigation about why it is regulated so tightly. We have looked into using click-it technology to determine the half-life of the WT and mutant mRNA and have the potential to clone the mutated alleles into a plasmid vector, which could then be manipulated. This would also allow us to directly test if the N-truncated mutation is caused by the hypothesized translation re-initiation event at a downstream methionine using pharmaceutical manipulations in cell culture.

I was initially interested in phenotyping these mouse models for social behaviors, as other models knocking out *Gtf2ird1* and *Gtf2i* have shown social phenotypes. In the social tasks we have done which include the three chamber social approach, tube test, and resident intruder, we have seen either no difference between genotypes, strain dependent effects, social effects in the opposite direction, or non replicable phenotypes. The CD model on the C57Bl/6J background should have the largest social phenotype as described in the mouse literature. The Dougherty lab has a new social operant paradigm that would allow us to test the social motivation of the CD animals, which we would predict to have increased motivation. We could then run our other WS models through this paradigm as well to test specifically for social motivation deficits.

We also see a conditioned fear response in the CD animals and the *Gtf2i*$^{+/-}$/*Gtf2ird1*$^{-/-}$ genotype. The oxytocin system has been largely speculated to contribute to the phenotypes in WS (118, 206) and it has been shown to affect conditioned fear in mice (207). I have generated preliminary data that suggests oxytocin is slightly upregulated in the hypothalamus of CD animals, however, I have not noticed an increased in oxytocin positive neurons in the hypothalamus. We can manipulate the oxytocin system, genetically or pharmaceutically, in CD animals to test if this rescues or exacerbates the behavioral phenotypes. I have also generated RNA-seq data from the adult hypothalamus of the CD and the *Gtf2i** animals. This can further inform the involvement of oxytocin and vasopressin in behavioral phenotypes and be used to design downstream experiments regarding these neuropeptides.

I am also interested in understanding how other genes in the WSCR could possibly modify the effects of the general transcription factor 2i family. I have generated a new mouse model that has a frameshift mutation in just *Gtf2i*. Characterizing this model will let us understand the effects of *Gtf2i* on behavior without a *Gtf2ird1* mutation. In collaboration with Dr. Kozel, we could cross our single mutants, double mutants, and the CD animals to a *Baz1b* knock out line to test how this chromatin modifier affects behavior.

5.3 Summary

This thesis used both human and mouse genetics to further understand genetic contributions of the WSCR to behavior. I have analyzed the largest genetic dataset of humans with WS and showed that variants on the remaining WSCR allele do not largely affect the social phenotype, but there is suggesting evidence for the role of variants in the *BAZ1B* and *GTF2IRD1* genes. This dataset can be used to query other clinically relevant phenotypes of WS. Further, I have generated and characterized new mouse models of *Gtf2i* and *Gtf2ird1* and showed that other genes in the WSCR are critical for causing the phenotypes seen when the whole region is deleted. The data produced here can be used to appreciate the genetic complexity of the WSCR and encourage research that looks at the interaction of the genes in the region.

References

1. Colleen A. Morris,M.D. and Carolyn B. Mervis,Ph.D. (2000) Williams Syndrome and Related Disorders. *Annual Review of Genomics and Human Genetics*, **1**, 461–484.

2. Korenberg,J.R., Chen,X.-N., Hirota,H., Lai,Z., Bellugi,U., Burian,D., Roe,B. and Matsuoka,R. (2000) VI. Genome Structure and Cognitive Map of Williams Syndrome. *Journal of Cognitive Neuroscience*, **12**, 89–107.

3. Ewart,A.K., Morris,C.A., Atkinson,D., Jin,W., Sternes,K., Spallone,P., Stock,A.D., Leppert,M. and Keating,M.T. (1993) Hemizygosity at the elastin locus in a developmental disorder, Williams syndrome. *Nature Genetics*, **5**, 11.

4. Williams,J.C.P., Barratt-Boyes,B.G. and Lowe,J.B. (1961) Supravalvular Aortic Stenosis. *Circulation*, **24**, 1311–1318.

5. Supravalvular Aortic Stenosis in Association with Mental Retardation and a Certain Facial Appearance.

6. Ewart,A.K., Morris,C.A., Ensing,G.J., Loker,J., Moore,C., Leppert,M. and Keating,M. (1993) A human vascular disorder, supravalvular aortic stenosis, maps to chromosome 7. *PNAS*, **90**, 3226–3230.

7. Lowery,M.C., Morris,C.A., Ewart,A., Brothman,L.J., Zhu,X.L., Leonard,C.O., Carey,J.C., Keating,M. and Brothman,A.R. (1995) Strong correlation of elastin deletions, detected by FISH, with Williams syndrome: evaluation of 235 patients. *Am J Hum Genet*, **57**, 49–53.

8. Meng,X., Lu,X., Li,Z., Green,E.D., Massa,H., Trask,B.J., Morris,C.A. and Keating,M.T. (1998) Complete physical map of the common deletion region in Williams syndrome and identification and characterization of three novel genes. *Hum Genet*, **103**, 590–599.

9. Peoples,R., Franke,Y., Wang,Y.-K., Pérez-Jurado,L., Paperna,T., Cisco,M. and Francke,U. (2000) A Physical Map, Including a BAC/PAC Clone Contig, of the Williams-Beuren Syndrome–Deletion Region at 7q11.23. *The American Journal of Human Genetics*, **66**, 47–68.

10. Scherer,S.W., Cheung,J., MacDonald,J.R., Osborne,L.R., Nakabayashi,K., Herbrick,J.-A., Carson,A.R., Parker-Katiraee,L., Skaug,J., Khaja,R., *et al.* (2003) Human Chromosome 7: DNA Sequence and Biology. *Science*, **300**, 767–772.

11. Antonell,A., Luis,O. de, Domingo-Roura,X. and Pérez-Jurado,L.A. (2005) Evolutionary mechanisms shaping the genomic structure of the Williams-Beuren syndrome chromosomal region at human 7q11.23. *Genome Res.*, **15**, 1179–1188.

12. Urbán,Z., Helms,C., Fekete,G., Csiszár,K., Bonnet,D., Munnich,A., Donis-Keller,H. and Boyd,C.D. (1996) 7q11.23 deletions in Williams syndrome arise as a consequence of unequal meiotic crossover. *Am J Hum Genet*, **59**, 958–962.

13. Merla,G., Brunetti-Pierri,N., Micale,L. and Fusco,C. (2010) Copy number variants at Williams–Beuren syndrome 7q11.23 region. *Hum Genet*, **128**, 3–26.

14. Järvinen,A., Korenberg,J.R. and Bellugi,U. (2013) The social phenotype of Williams syndrome. *Current Opinion in Neurobiology*, **23**, 414–422.

15. Meyer-Lindenberg,A., Mervis,C.B. and Faith Berman,K. (2006) Neural mechanisms in Williams syndrome: a unique window to genetic influences on cognition and behaviour. *Nat Rev Neurosci*, **7**, 380–393.

16. Schubert,C. (2009) The genomic basis of the Williams – Beuren syndrome. *Cell. Mol. Life Sci.*, **66**, 1178–1197.

17. Mervis,C.B., Robinson,B.F., Bertrand,J., Morris,C.A., Klein-Tasman,B.P. and Armstrong,S.C. (2000) The Williams Syndrome Cognitive Profile. *Brain and Cognition*, **44**, 604–628.

18. Riby,D. and Hancock,P.J.B. (2009) Looking at movies and cartoons: eye-tracking evidence from Williams syndrome and autism. *Journal of Intellectual Disability Research*, **53**, 169–181.

19. Doyle,T.F., Bellugi,U., Korenberg,J.R. and Graham,J. (2004) "Everybody in the world is my friend" hypersociability in young children with Williams syndrome. *Am. J. Med. Genet.*, **124A**, 263–273.

20. Klein-Tasman,B.P., Li-Barber,K.T. and Magargee,E.T. (2010) Honing in on the Social Phenotype in Williams Syndrome Using Multiple Measures and Multiple Raters. *J Autism Dev Disord*, **41**, 341–351.

21. Dykens,E.M. (2003) Anxiety, Fears, and Phobias in Persons With Williams Syndrome. *Developmental Neuropsychology*, **23**, 291–316.

22. Leyfer,O.T., Woodruff-Borden,J., Klein-Tasman,B.P., Fricke,J.S. and Mervis,C.B. (2006) Prevalence of Psychiatric Disorders in 4 - 16-Year-Olds with Williams Syndrome. *Am J Med Genet B Neuropsychiatr Genet*, **141B**, 615–622.

23. Somerville,M.J., Mervis,C.B., Young,E.J., Seo,E.-J., del Campo,M., Bamforth,S., Peregrine,E., Loo,W., Lilley,M., Pérez-Jurado,L.A., *et al.* (2005) Severe Expressive-Language Delay Related to Duplication of the Williams–Beuren Locus. *New England Journal of Medicine*, **353**, 1694–1701.

24. Morris,C.A., Mervis,C.B., Paciorkowski,A.P., Abdul-Rahman,O., Dugan,S.L., Rope,A.F., Bader,P., Hendon,L.G., Velleman,S.L., Klein-Tasman,B.P., *et al.* (2015) 7q11.23 Duplication syndrome: Physical characteristics and natural history. *American Journal of Medical Genetics Part A*, **167**, 2916–2935.

25. Mervis,C.B., Klein-Tasman,B.P., Huffman,M.J., Velleman,S.L., Pitts,C.H., Henderson,D.R., Woodruff-Borden,J., Morris,C.A. and Osborne,L.R. (2015) Children with 7q11.23 duplication syndrome: Psychological characteristics. *American Journal of Medical Genetics Part A*, **167**, 1436–1450.

26. Sanders,S.J., Ercan-Sencicek,A.G., Hus,V., Luo,R., Murtha,M.T., Moreno-De-Luca,D., Chu,S.H., Moreau,M.P., Gupta,A.R., Thomson,S.A., *et al.* (2011) Multiple Recurrent De Novo CNVs, Including Duplications of the 7q11.23 Williams Syndrome Region, Are Strongly Associated with Autism. *Neuron*, **70**, 863–885.

27. Klein-Tasman,B.P. and Mervis,C.B. (2018) Autism Spectrum Symptomatology Among Children with Duplication 7q11.23 Syndrome. *J Autism Dev Disord*, 10.1007/s10803-017-3439-z.

28. Klein-Tasman,B.P., Fluit,F. van der and Mervis,C.B. (2018) Autism Spectrum Symptomatology in Children with Williams Syndrome Who Have Phrase Speech or Fluent Language. *J Autism Dev Disord*, 10.1007/s10803-018-3555-4.

29. Mervis,C.B., Dida,J., Lam,E., Crawford-Zelli,N.A., Young,E.J., Henderson,D.R., Onay,T., Morris,C.A., Woodruff-Borden,J., Yeomans,J., *et al.* (2012) Duplication of GTF2I Results in Separation Anxiety in Mice and Humans. *The American Journal of Human Genetics*, **90**, 1064–1070.

30. Frangiskakis,J.M., Ewart,A.K., Morris,C.A., Mervis,C.B., Bertrand,J., Robinson,B.F., Klein,B.P., Ensing,G.J., Everett,L.A., Green,E.D., *et al.* (1996) LIM-kinase1 Hemizygosity Implicated in Impaired Visuospatial Constructive Cognition. *Cell*, **86**, 59–69.

31. Morris,C.A., Mervis,C.B., Hobart,H.H., Gregg,R.G., Bertrand,J., Ensing,G.J., Sommer,A., Moore,C.A., Hopkin,R.J., Spallone,P.A., *et al.* (2003) GTF2I hemizygosity implicated in mental retardation in Williams syndrome: Genotype–phenotype analysis of five families with deletions in the Williams syndrome region. *Am. J. Med. Genet.*, **123A**, 45–59.

32. Tassabehji,M., Metcalfe,K., Karmiloff-Smith,A., Carette,M.J., Grant,J., Dennis,N., Reardon,W., Splitt,M., Read,A.P. and Donnai,D. (1999) Williams Syndrome: Use of Chromosomal Microdeletions as a Tool to Dissect Cognitive and Physical Phenotypes. *The American Journal of Human Genetics*, **64**, 118–125.

33. Karmiloff-Smith,A., Grant,J., Ewing,S., Carette,M.J., Metcalfe,K., Donnai,D., Read,A.P. and Tassabehji,M. (2003) Using case study comparisons to explore genotype-phenotype correlations in Williams-Beuren syndrome. *Journal of Medical Genetics*, **40**, 136–140.

34. Botta,A., Novelli,G., Mari,A., Novelli,A., Sabani,M., Korenberg,J., Osborne,L.R., Digilio,M.C., Giannotti,A. and Dallapiccola,B. (1999) Detection of an atypical 7q11.23 deletion in Williams syndrome patients which does not include the STX1A and FZD3 genes. *Journal of Medical Genetics*, **36**, 478–480.

35. Heller,R., Rauch,A., Lüttgen,S., Schröder,B. and Winterpacht,A. (2003) Partial deletion of the critical 1.5 Mb interval in Williams-Beuren syndrome. *Journal of Medical Genetics*, **40**, e99–e99.

36. Hirota,H., Matsuoka,R., Chen,X.-N., Salandanan,L.S., Lincoln,A., Rose,F.E., Sunahara,M., Osawa,M., Bellugi,U. and Korenberg,J.R. (2003) Williams syndrome deficits in visual spatial processing linked to GTF2IRD1 and GTF2I on Chromosome 7q11.23. *Genet Med*, **5**, 311–321.

37. Gagliardi,C., Bonaglia,M.C., Selicorni,A., Borgatti,R. and Giorda,R. (2003) Unusual cognitive and behavioural profile in a Williams syndrome patient with atypical 7q11.23 deletion. *Journal of Medical Genetics*, **40**, 526–530.

38. Dai,L., Bellugi,U., Chen,X.-N., Pulst-Korenberg,A. m., Järvinen-Pasley,A., Tirosh-Wagner,T., Eis,P. s., Graham,J., Mills,D., Searcy,Y., *et al.* (2009) Is it Williams syndrome? GTF2IRD1 implicated in visual–spatial construction and GTF2I in sociability revealed by high resolution arrays. *Am. J. Med. Genet.*, **149A**, 302–314.

39. Tassabehji,M., Hammond,P., Karmiloff-Smith,A., Thompson,P., Thorgeirsson,S.S., Durkin,M.E., Popescu,N.C., Hutton,T., Metcalfe,K., Rucka,A., *et al.* (2005) GTF2IRD1 in Craniofacial Development of Humans and Mice. *Science*, **310**, 1184–1187.

40. Antonell,A., Campo,M.D., Magano,L.F., Kaufmann,L., Iglesia,J.M. de la, Gallastegui,F., Flores,R., Schweigmann,U., Fauth,C., Kotzot,D., *et al.* (2010) Partial 7q11.23 deletions further implicate GTF2I and GTF2IRD1 as the main genes responsible for the Williams–Beuren syndrome neurocognitive profile. *Journal of Medical Genetics*, **47**, 312–320.

41. Porter,M.A., Dobson-Stone,C., Kwok,J.B.J., Schofield,P.R., Beckett,W. and Tassabehji,M. (2012) A Role for Transcription Factor GTF2IRD2 in Executive Function in Williams-Beuren Syndrome. *PLOS ONE*, **7**, e47457.

42. Serrano-Juárez,C.A., Venegas-Vega,C.A., Yáñez-Téllez,M.G., Rodríguez-Camacho,M., Silva-Pereyra,J., Salgado-Ceballos,H. and Prieto-Corona,B. (2018) Cognitive, Behavioral, and Adaptive Profiles in Williams Syndrome With and Without Loss of GTF2IRD2. *Journal of the International Neuropsychological Society*, **24**, 896–904.

43. Kozel Beth A., Danback Joshua R., Waxler Jessica L., Knutsen Russell H., de las Fuentes Lisa, Reusz Gyorgy S., Kis Eva, Bhatt Ami B. and Pober Barbara R. (2014) Williams Syndrome Predisposes to Vascular Stiffness Modified by Antihypertensive Use and Copy Number Changes in NCF1. *Hypertension*, **63**, 74–79.

44. Adamo,A., Atashpaz,S., Germain,P.-L., Zanella,M., D'Agostino,G., Albertin,V., Chenoweth,J., Micale,L., Fusco,C., Unger,C., *et al.* (2015) 7q11.23 dosage-dependent dysregulation in human pluripotent stem cells affects transcriptional programs in disease-relevant lineages. *Nat Genet*, **47**, 132–141.

45. Chailangkarn,T. and Muotri,A.R. (2017) Modeling Williams syndrome with induced pluripotent stem cells. *Neurogenesis*, **4**, e1283187.

46. Chailangkarn,T., Trujillo,C.A., Freitas,B.C., Hrvoj-Mihic,B., Herai,R.H., Yu,D.X., Brown,T.T., Marchetto,M.C., Bardy,C., McHenry,L., *et al.* (2016) A human neurodevelopmental model for Williams syndrome. *Nature*, **advance online publication**.

47. Kinnear,C., Chang,W.Y., Khattak,S., Hinek,A., Thompson,T., de Carvalho Rodrigues,D., Kennedy,K., Mahmut,N., Pasceri,P., Stanford,W.L., *et al.* (2013) Modeling and Rescue of the Vascular Phenotype of Williams-Beuren Syndrome in Patient Induced Pluripotent Stem Cells. *STEM CELLS Translational Medicine*, **2**, 2–15.

48. Khattak,S., Brimble,E., Zhang,W., Zaslavsky,K., Strong,E., Ross,P.J., Hendry,J., Mital,S., Salter,M.W., Osborne,L.R., *et al.* (2015) Human induced pluripotent stem cell derived neurons as a model for Williams-Beuren syndrome. *Molecular Brain*, **8**.

49. Lalli,M.A., Jang,J., Park,J.-H.C., Wang,Y., Guzman,E., Zhou,H., Audouard,M., Bridges,D., Tovar,K.R., Papuc,S.M., *et al.* (2016) Haploinsufficiency of BAZ1B contributes to Williams syndrome through transcriptional dysregulation of neurodevelopmental pathways. *Hum Mol Genet*, **25**, 1294–1306.

50. Malenfant,P., Liu,X., Hudson,M.L., Qiao,Y., Hrynchak,M., Riendeau,N., Hildebrand,M.J., Cohen,I.L., Chudley,A.E., Forster-Gibson,C., *et al.* (2011) Association of GTF2i in the Williams-Beuren Syndrome Critical Region with Autism Spectrum Disorders. *J Autism Dev Disord*, **42**, 1459–1469.

51. Swartz,J.R., Waller,R., Bogdan,R., Knodt,A.R., Sabhlok,A., Hyde,L.W. and Hariri,A.R. A common polymorphism in a Williams syndrome gene predicts amygdala reactivity and extraversion in healthy adults. *Biological Psychiatry*, 10.1016/j.biopsych.2015.12.007.

52. Fan,C.C., Schork,A.J., Brown,T.T., Spencer,B.E., Akshoomoff,N., Chen,C.-H., Kuperman,J.M., Hagler,D.J., Steen,V.M., Hellard,S., *et al.* (2018) Williams Syndrome neuroanatomical score associates with GTF2IRD1 in large-scale magnetic resonance imaging cohorts: a proof of concept for multivariate endophenotypes. *Translational Psychiatry*, **8**, 114.

53. Crespi,B.J. and Hurd,P.L. (2014) Cognitive-behavioral phenotypes of Williams syndrome are associated with genetic variation in the GTF2I gene, in a healthy population. *BMC Neuroscience*, **15**, 127.

54. Popp,B., Trollmann,R., Büttner,C., Caliebe,A., Thiel,C.T., Hüffmeier,U., Reis,A. and Zweier,C. Do the exome: A case of Williams-Beuren syndrome with severe epilepsy due to a truncating de novo variant in GABRA1. *European Journal of Medical Genetics*, 10.1016/j.ejmg.2016.09.002.

55. Egloff,M., Nguyen,L.-S., Siquier-Pernet,K., Cormier-Daire,V., Baujat,G., Attié-Bitach,T., Bole-Feysot,C., Nitschke,P., Vekemans,M., Colleaux,L., *et al.* (2018) Whole-exome

sequence analysis highlights the role of unmasked recessive mutations in copy number variants with incomplete penetrance. *European Journal of Human Genetics*, **26**, 912–918.

56. Delio,M., Pope,K., Wang,T., Samanich,J., Haldeman-Englert,C.R., Kaplan,P., Shaikh,T.H., Cai,J., Marion,R.W., Morrow,B.E., *et al.* (2013) Spectrum of elastin sequence variants and cardiovascular phenotypes in 49 patients with Williams–Beuren syndrome. *Am. J. Med. Genet.*, **161**, 527–533.

57. Yang,W. and Desiderio,S. (1997) BAP-135, a target for Bruton's tyrosine kinase in response to B cell receptor engagement. *PNAS*, **94**, 604–609.

58. Grueneberg,D.A., Henry,R.W., Brauer,A., Novina,C.D., Cheriyath,V., Roy,A.L. and Gilman,M. (1997) A multifunctional DNA-binding protein that promotes the formation of serum response factor/homeodomain complexes: identity to TFII-I. *Genes Dev.*, **11**, 2482–2493.

59. Roy,A.L., Meisterernst,M., Pognonec,P. and Roeder,R.G. (1991) Cooperative interaction of an initiator-binding transcription initiation factor and the helix–loop–helix activator USF. *Nature*, **354**, 245–248.

60. O'Mahoney,J.V., Guven,K.L., Lin,J., Joya,J.E., Robinson,C.S., Wade,R.P. and Hardeman,E.C. (1998) Identification of a Novel Slow-Muscle-Fiber Enhancer Binding Protein, MusTRD1. *Molecular and Cellular Biology*, **18**, 6641–6652.

61. Franke,Y., Peoples,R.J. and Francke,U. (1999) Identification of GTF2IRD1, a putative transcription factor within the Williams-Beuren syndrome deletion at 7q11.23. *CGR*, **86**, 296–304.

62. Osborne,L.R., Campbell,T., Daradich,A., Scherer,S.W. and Tsui,L.-C. (1999) Identification of a Putative Transcription Factor Gene (WBSCR11) That Is Commonly Deleted in Williams–Beuren Syndrome. *Genomics*, **57**, 279–284.

63. Makeyev,A.V. and Bayarsaihan,D. (2009) Alternative splicing and promoter use in TFII-I genes. *Gene*, **433**, 16–25.

64. Tipney,H.J., Hinsley,T.A., Brass,A., Metcalfe,K., Donnai,D. and Tassabehji,M. (2004) Isolation and characterisation of GTF2IRD2, a novel fusion gene and member of the TFII-I family of transcription factors, deleted in Williams–Beuren syndrome. *Eur J Hum Genet*, **12**, 551–560.

65. Lazebnik,M.B., Tussie-Luna,M.I. and Roy,A.L. (2008) Determination and Functional Analysis of the Consensus Binding Site for TFII-I Family Member BEN, Implicated in Williams-Beuren Syndrome. *J. Biol. Chem.*, **283**, 11078–11082.

66. Palmer,S.J., Santucci,N., Widagdo,J., Bontempo,S.J., Taylor,K.M., Tay,E.S.E., Hook,J., Lemckert,F., Gunning,P.W. and Hardeman,E.C. (2010) Negative Autoregulation of

GTF2IRD1 in Williams-Beuren Syndrome via a Novel DNA Binding Mechanism. *J. Biol. Chem.*, **285**, 4715–4724.

67. Lucena,J., Pezzi,S., Aso,E., Valero,M.C., Carreiro,C., Dubus,P., Sampaio,A., Segura,M., Barthelemy,I., Zindel,M.Y., *et al.* (2010) Essential role of the N-terminal region of TFII-I in viability and behavior. *BMC Medical Genetics*, **11**, 61.

68. Gunbin,K.V. and Ruvinsky,A. (2012) Evolution of General Transcription Factors. *J Mol Evol*, **76**, 28–47.

69. Bayarsaihan,D., Dunai,J., Greally,J.M., Kawasaki,K., Sumiyama,K., Enkhmandakh,B., Shimizu,N. and Ruddle,F.H. (2002) Genomic Organization of the Genes Gtf2ird1, Gtf2i, and Ncf1 at the Mouse Chromosome 5 Region Syntenic to the Human Chromosome 7q11.23 Williams Syndrome Critical Region. *Genomics*, **79**, 137–143.

70. Enkhmandakh,B., Bitchevaia,N., Ruddle,F. and Bayarsaihan,D. (2004) The early embryonic expression of TFII-I during mouse preimplantation development. *Gene Expression Patterns*, **4**, 25–28.

71. Fijalkowska,I., Sharma,D., Bult,C.J. and Danoff,S.K. (2010) Expression of the transcription factor, TFII-I, during post-implantation mouse embryonic development. *BMC Research Notes*, **3**, 203.

72. Danoff,S.K., Taylor,H.E., Blackshaw,S. and Desiderio,S. (2004) TFII-I, a candidate gene for Williams syndrome cognitive profile: parallels between regional expression in mouse brain and human phenotype. *Neuroscience*, **123**, 931–938.

73. Roy,A.L. (2001) Biochemistry and biology of the inducible multifunctional transcription factor TFII-I. *Gene*, **274**, 1–13.

74. Cheriyath,V., Desgranges,Z.P. and Roy,A.L. (2002) c-Src-dependent Transcriptional Activation of TFII-I. *J. Biol. Chem.*, **277**, 22798–22805.

75. Sinai,L., Ivakine,E.A., Lam,E., Deurloo,M., Dida,J., Zirngibl,R.A., Jung,C., Aubin,J.E., Feng,Z.-P., Yeomans,J., *et al.* (2015) Disruption of Src Is Associated with Phenotypes Related to Williams-Beuren Syndrome and Altered Cellular Localization of TFII-I. *eNeuro*, **2**, ENEURO.0016-14.2015.

76. Caraveo,G., Rossum,D.B. van, Patterson,R.L., Snyder,S.H. and Desiderio,S. (2006) Action of TFII-I Outside the Nucleus as an Inhibitor of Agonist-Induced Calcium Entry. *Science*, **314**, 122–125.

77. Deurloo,M.H.S., Turlova,E., Chen,W.-L., Lin,Y.W., Tam,E., Tassew,N.G., Wu,M., Huang,Y.-C., Crawley,J.N., Monnier,P.P., *et al.* (2018) Transcription Factor 2I Regulates Neuronal Development via TRPC3 in 7q11.23 Disorder Models. *Mol Neurobiol*, 10.1007/s12035-018-1290-7.

78. Palmer,S.J., Tay,E.S.E., Santucci,N., Cuc Bach,T.T., Hook,J., Lemckert,F.A., Jamieson,R.V., Gunnning,P.W. and Hardeman,E.C. (2007) Expression of Gtf2ird1, the Williams syndrome-associated gene, during mouse development. *Gene Expression Patterns*, **7**, 396–404.

79. Carmona-Mora,P., Widagdo,J., Tomasetig,F., Canales,C.P., Cha,Y., Lee,W., Alshawaf,A., Dottori,M., Whan,R.M., Hardeman,E.C., *et al.* (2015) The nuclear localization pattern and interaction partners of GTF2IRD1 demonstrate a role in chromatin regulation. *Hum Genet*, **134**, 1099–1115.

80. Masuda,T., Zhang,X., Berlinicke,C., Wan,J., Yerrabelli,A., Conner,E.A., Kjellstrom,S., Bush,R., Thorgeirsson,S.S., Swaroop,A., *et al.* (2014) The Transcription Factor GTF2IRD1 Regulates the Topology and Function of Photoreceptors by Modulating Photoreceptor Gene Expression across the Retina. *J. Neurosci.*, **34**, 15356–15368.

81. Hasegawa,Y., Ikeda,K., Chen,Y., Alba,D.L., Stifler,D., Shinoda,K., Hosono,T., Maretich,P., Yang,Y., Ishigaki,Y., *et al.* (2018) Repression of Adipose Tissue Fibrosis through a PRDM16-GTF2IRD1 Complex Improves Systemic Glucose Homeostasis. *Cell Metabolism*, **27**, 180-194.e6.

82. Widagdo,J., Taylor,K.M., Gunning,P.W., Hardeman,E.C. and Palmer,S.J. (2012) SUMOylation of GTF2IRD1 Regulates Protein Partner Interactions and Ubiquitin-Mediated Degradation. *PLoS ONE*, **7**, e49283.

83. Vullhorst,D. and Buonanno,A. (2005) Multiple GTF2I-like Repeats of General Transcription Factor 3 Exhibit DNA Binding Properties EVIDENCE FOR A COMMON ORIGIN AS A SEQUENCE-SPECIFIC DNA INTERACTION MODULE. *J. Biol. Chem.*, **280**, 31722–31731.

84. Makeyev,A.V., Enkhmandakh,B., Hong,S.-H., Joshi,P., Shin,D.-G. and Bayarsaihan,D. (2012) Diversity and Complexity in Chromatin Recognition by TFII-I Transcription Factors in Pluripotent Embryonic Stem Cells and Embryonic Tissues. *PLOS ONE*, **7**, e44443.

85. Chimge,N.-O., Mungunsukh,O., Ruddle,F. and Bayarsaihan,D. (2007) Expression profiling of BEN regulated genes in mouse embryonic fibroblasts. *J. Exp. Zool.*, **308B**, 209–224.

86. Chimge,N.-O., Mungunsukh,O., Ruddle,F. and Bayarsaihan,D. (2007) Gene expression analysis of TFII-I modulated genes in mouse embryonic fibroblasts. *J. Exp. Zool.*, **308B**, 225–235.

87. Enkhmandakh,B., Makeyev,A.V., Erdenechimeg,L., Ruddle,F.H., Chimge,N.-O., Tussie-Luna,M.I., Roy,A.L. and Bayarsaihan,D. (2009) Essential functions of the Williams-Beuren syndrome-associated TFII-I genes in embryonic development. *PNAS*, **106**, 181–186.

88. O'Leary,J. and Osborne,L.R. (2011) Global Analysis of Gene Expression in the Developing Brain of Gtf2ird1 Knockout Mice. *PLoS ONE*, **6**, e23868.

89. Corley,S.M., Canales,C.P., Carmona-Mora,P., Mendoza-Reinosa,V., Beverdam,A., Hardeman,E.C., Wilkins,M.R. and Palmer,S.J. (2016) RNA-Seq analysis of Gtf2ird1 knockout epidermal tissue provides potential insights into molecular mechanisms underpinning Williams-Beuren syndrome. *BMC Genomics*, **17**, 450.

90. Borralleras,C., Sahun,I., Pérez-Jurado,L.A. and Campuzano,V. (2015) Intracisternal Gtf2i Gene Therapy Ameliorates Deficits in Cognition and Synaptic Plasticity of a Mouse Model of Williams–Beuren Syndrome. *Mol Ther*, 10.1038/mt.2015.130.

91. Ortiz-Romero,P., Borralleras,C., Bosch-Morató,M., Guivernau,B., Albericio,G., Muñoz,F.J., Pérez-Jurado,L.A. and Campuzano,V. (2018) Epigallocatechin-3-gallate improves cardiac hypertrophy and short-term memory deficits in a Williams-Beuren syndrome mouse model. *PLOS ONE*, **13**, e0194476.

92. Howard,M.L., Palmer,S.J., Taylor,K.M., Arthurson,G.J., Spitzer,M.W., Du,X., Pang,T.Y.C., Renoir,T., Hardeman,E.C. and Hannan,A.J. (2012) Mutation of Gtf2ird1 from the Williams–Beuren syndrome critical region results in facial dysplasia, motor dysfunction, and altered vocalisations. *Neurobiology of Disease*, **45**, 913–922.

93. Segura-Puimedon,M., Sahún,I., Velot,E., Dubus,P., Borralleras,C., Rodrigues,A.J., Valero,M.C., Valverde,O., Sousa,N., Herault,Y., *et al.* (2014) Heterozygous deletion of the Williams–Beuren syndrome critical interval in mice recapitulates most features of the human disorder. *Hum. Mol. Genet.*, **23**, 6481–6494.

94. Li,H.H., Roy,M., Kuscuoglu,U., Spencer,C.M., Halm,B., Harrison,K.C., Bayle,J.H., Splendore,A., Ding,F., Meltzer,L.A., *et al.* (2009) Induced chromosome deletions cause hypersociability and other features of Williams–Beuren syndrome in mice. *EMBO Molecular Medicine*, **1**, 50–65.

95. Osborne,L.R. (2010) Animal models of Williams syndrome. *Am. J. Med. Genet.*, **154C**, 209–219.

96. Sakurai,T., Dorr,N.P., Takahashi,N., McInnes,L.A., Elder,G.A. and Buxbaum,J.D. (2011) Haploinsufficiency of Gtf2i, a gene deleted in Williams Syndrome, leads to increases in social interactions. *Autism Res*, **4**, 28–39.

97. Martin,L.A., Iceberg,E. and Allaf,G. Consistent hypersocial behavior in mice carrying a deletion of Gtf2i but no evidence of hyposocial behavior with Gtf2i duplication: Implications for Williams–Beuren syndrome and autism spectrum disorder. *Brain Behav*, 10.1002/brb3.895.

98. Durkin,M.E., Keck-Waggoner,C.L., Popescu,N.C. and Thorgeirsson,S.S. (2001) Integration of a c-myc Transgene Results in Disruption of the Mouse Gtf2ird1 Gene, the Homologue

of the Human GTF2IRD1 Gene Hemizygously Deleted in Williams–Beuren Syndrome. *Genomics*, **73**, 20–27.

99. van Hagen,J.M., van der Geest,J.N., van der Giessen,R.S., Lagers-van Haselen,G.C., Eussen,H.J.F.M.M., Gille,J.J.P., Govaerts,L.C.P., Wouters,C.H., de Coo,I.F.M., Hoogenraad,C.C., *et al.* (2007) Contribution of CYLN2 and GTF2IRD1 to neurological and cognitive symptoms in Williams Syndrome. *Neurobiology of Disease*, **26**, 112–124.

100. Schneider,T., Skitt,Z., Liu,Y., Deacon,R.M.J., Flint,J., Karmiloff-Smith,A., Rawlins,J.N.P. and Tassabehji,M. (2012) Anxious, hypoactive phenotype combined with motor deficits in Gtf2ird1 null mouse model relevant to Williams syndrome. *Behavioural Brain Research*, **233**, 458–473.

101. Young,E.J., Lipina,T., Tam,E., Mandel,A., Clapcote,S.J., Bechard,A.R., Chambers,J., Mount,H.T.J., Fletcher,P.J., Roder,J.C., *et al.* (2008) Reduced fear and aggression and altered serotonin metabolism in Gtf2ird1-targeted mice. *Genes, Brain and Behavior*, **7**, 224–234.

102. Qiu,Y., Arbogast,T., Lorenzo,S.M., Li,H., Shih,T., Ellen,R., Hong,O., Cho,S., Shanta,O., Timothy,P., *et al.* (2019) Oligogenic effects of 16p11.2 copy number variation on craniofacial development. *bioRxiv*, 10.1101/540732.

103. Iyer,J., Singh,M.D., Jensen,M., Patel,P., Pizzo,L., Huber,E., Koerselman,H., Weiner,A.T., Lepanto,P., Vadodaria,K., *et al.* (2018) Pervasive genetic interactions modulate neurodevelopmental defects of the autism-associated 16p11.2 deletion in Drosophila melanogaster. *Nature Communications*, **9**, 2548.

104. Li,D.Y., Toland,A.E., Boak,B.B., Atkinson,D.L., Ensing,G.J., Morris,C.A. and Keating,M.T. (1997) Elastin Point Mutations Cause an Obstructive Vascular Disease, Supravalvular Aortic Stenosis. *Hum Mol Genet*, **6**, 1021–1028.

105. Ashe,A., Morgan,D.K., Whitelaw,N.C., Bruxner,T.J., Vickaryous,N.K., Cox,L.L., Butterfield,N.C., Wicking,C., Blewitt,M.E., Wilkins,S.J., *et al.* (2008) A genome-wide screen for modifiers of transgene variegation identifies genes with critical roles in development. *Genome Biology*, **9**, R182.

106. Fujiwara,T., Sanada,M., Kofuji,T. and Akagawa,K. (2016) Unusual social behavior in HPC-1/syntaxin1A knockout mice is caused by disruption of the oxytocinergic neural system. *J. Neurochem.*, 10.1111/jnc.13634.

107. Gao,M.C., Bellugi,U., Dai,L., Mills,D.L., Sobel,E.M., Lange,K. and Korenberg,J.R. (2010) Intelligence in Williams Syndrome Is Related to STX1A, Which Encodes a Component of the Presynaptic SNARE Complex. *PLOS ONE*, **5**, e10292.

108. Hoogenraad,C.C., Koekkoek,B., Akhmanova,A., Krugers,H., Dortland,B., Miedema,M., van Alphen,A., Kistler,W.M., Jaegle,M., Koutsourakis,M., *et al.* (2002) Targeted mutation of Cyln2 in the Williams syndrome critical region links CLIP-115

haploinsufficiency to neurodevelopmental abnormalities in mice. *Nat Genet*, **32**, 116–127.

109. Meng,Y., Zhang,Y., Tregoubov,V., Janus,C., Cruz,L., Jackson,M., Lu,W.-Y., MacDonald,J.F., Wang,J.Y., Falls,D.L., *et al.* (2002) Abnormal Spine Morphology and Enhanced LTP in LIMK-1 Knockout Mice. *Neuron*, **35**, 121–133.

110. Gosch,A. and Pankau,R. (1997) Personality characteristics and behaviour problems in individuals of different ages with Williams syndrome. *Developmental Medicine & Child Neurology*, **39**, 527–533.

111. Reilly,J., Klima,E.S. and Bellugi,U. (1990) Once more with feeling: Affect and language in atypical populations. *Development and Psychopathology*, **2**, 367–391.

112. Tager-Flusberg,H. and Sullivan,K. (2000) A componential view of theory of mind: evidence from Williams syndrome. *Cognition*, **76**, 59–90.

113. Klein-Tasman,B.P., Mervis,C.B., Lord,C. and Phillips,K.D. (2007) Socio-Communicative Deficits in Young Children with Williams Syndrome: Performance on the Autism Diagnostic Observation Schedule. *Child Neuropsychology*, **13**, 444–467.

114. Klein-Tasman,B.P., Phillips,K.D., Lord,C.E., Mervis,C.B. and Gallo,F. (2009) Overlap with the Autism Spectrum in Young Children with Williams Syndrome. *J Dev Behav Pediatr*, **30**, 289–299.

115. Richards,C., Jones,C., Groves,L., Moss,J. and Oliver,C. (2015) Prevalence of autism spectrum disorder phenomenology in genetic disorders: a systematic review and meta-analysis. *The Lancet Psychiatry*, **2**, 909–916.

116. Brawn,G. and Porter,M. (2014) Adaptive functioning in Williams syndrome and its relation to demographic variables and family environment. *Research in Developmental Disabilities*, **35**, 3606–3623.

117. Jabbi,M., Chen,Q., Turner,N., Kohn,P., White,M., Kippenhan,J.S., Dickinson,D., Kolachana,B., Mattay,V., Weinberger,D.R., *et al.* (2015) Variation in the Williams syndrome GTF2I gene and anxiety proneness interactively affect prefrontal cortical response to aversive stimuli. *Transl Psychiatry*, **5**, e622.

118. Procyshyn,T.L., Spence,J., Read,S., Watson,N.V. and Crespi,B.J. (2017) The Williams syndrome prosociality gene GTF2I mediates oxytocin reactivity and social anxiety in a healthy population. *Biology Letters*, **13**, 20170051.

119. Constantino JN and Todd RD (2003) Autistic traits in the general population: A twin study. *Arch Gen Psychiatry*, **60**, 524–530.

120. Moreno-De-Luca,A., Evans,D.W., Boomer,K.B., Hanson,E., Bernier,R., Goin-Kochel,R.P., Myers,S.M., Challman,T.D., Moreno-De-Luca,D., Slane,M.M., *et al.* (2015) The Role of

Parental Cognitive, Behavioral, and Motor Profiles in Clinical Variability in Individuals With Chromosome 16p11.2 Deletions. *JAMA Psychiatry*, **72**, 119–126.

121. Sanders,S.J., He,X., Willsey,A.J., Ercan-Sencicek,A.G., Samocha,K.E., Cicek,A.E., Murtha,M.T., Bal,V.H., Bishop,S.L., Dong,S., *et al.* (2015) Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron*, **87**, 1215–1233.

122. Klein-Tasman,B.P. and Mervis,C.B. (2003) Distinctive Personality Characteristics of 8-, 9-, and 10-Year-Olds With Williams Syndrome. *Developmental Neuropsychology*, **23**, 269–290.

123. Einfeld,S.L., Tonge,B.J. and Florio,T. (1997) Behavioral and Emotional Disturbance in Individuals With Williams Syndrome. *American Journal on Mental Retardation*, **102**, 45–53.

124. Einfeld,S.L., Tonge,B.J. and Rees,V.W. (2001) Longitudinal Course of Behavioral and Emotional Problems in Williams Syndrome. *American Journal on Mental Retardation*, **106**, 73–81.

125. Switaj,D.M. (2001) Identification and measurement of anxiety and obsessive-compulsive tendencies in the Williams Syndrome behavioral phenotype.

126. Dong,C., Wei,P., Jian,X., Gibbs,R., Boerwinkle,E., Wang,K. and Liu,X. (2015) Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum Mol Genet*, **24**, 2125–2137.

127. Li,D.Y., Brooke,B., Davis,E.C., Mecham,R.P., Sorensen,L.K., Boak,B.B., Eichwald,E. and Keating,M.T. (1998) Elastin is an essential determinant of arterial morphogenesis. *Nature*, **393**, 276–280.

128. Manolio,T.A., Collins,F.S., Cox,N.J., Goldstein,D.B., Hindorff,L.A., Hunter,D.J., McCarthy,M.I., Ramos,E.M., Cardon,L.R., Chakravarti,A., *et al.* (2009) Finding the missing heritability of complex diseases. *Nature*, **461**, 747–753.

129. Brook,J.D., McCurrach,M.E., Harley,H.G., Buckler,A.J., Church,D., Aburatani,H., Hunter,K., Stanton,V.P., Thirion,J.-P., Hudson,T., *et al.* (1992) Molecular basis of myotonic dystrophy: Expansion of a trinucleotide (CTG) repeat at the 3′ end of a transcript encoding a protein kinase family member. *Cell*, **68**, 799–808.

130. Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia (2017) *Mol Autism*, **8**.

131. Euesden,J., Lewis,C.M. and O'Reilly,P.F. (2015) PRSice: Polygenic Risk Score software. *Bioinformatics*, **31**, 1466–1468.

132. Anney,R., Klei,L., Pinto,D., Almeida,J., Bacchelli,E., Baird,G., Bolshakova,N., Bölte,S., Bolton,P.F., Bourgeron,T., *et al.* (2012) Individual common variants exert weak effects on the risk for autism spectrum disorders. *Hum. Mol. Genet.*, **21**, 4781–4792.

133. Joyce,C.A., Zorich,B., Pike,S.J., Barber,J.C. and Dennis,N.R. (1996) Williams-Beuren syndrome: phenotypic variability and deletions of chromosomes 7, 11, and 22 in a series of 52 patients. *J Med Genet*, **33**, 986–992.

134. Porter,M.A. and Coltheart,M. (2005) Cognitive Heterogeneity in Williams Syndrome. *Developmental Neuropsychology*, **27**, 275–306.

135. Lek,M., Karczewski,K.J., Minikel,E.V., Samocha,K.E., Banks,E., Fennell,T., O'Donnell-Luria,A.H., Ware,J.S., Hill,A.J., Cummings,B.B., *et al.* (2016) Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, **536**, 285–291.

136. Crespi,B.J. and Procyshyn,T.L. (2017) Williams syndrome deletions and duplications: Genetic windows to understanding anxiety, sociality, autism, and schizophrenia. *Neuroscience & Biobehavioral Reviews*, **79**, 14–26.

137. Ekström,A.-B., Hakenäs-Plate,L., Samuelsson,L., Tulinius,M. and Wentz,E. (2008) Autism spectrum conditons in myotonic dystrophy type 1: A study on 57 individuals with congenital and childhood forms. *Am. J. Med. Genet.*, **147B**, 918–926.

138. Constantino,J.N., Davis,S.A., Todd,R.D., Schindler,M.K., Gross,M.M., Brophy,S.L., Metzger,L.M., Shoushtari,C.S., Splinter,R. and Reich,W. (2003) Validation of a Brief Quantitative Measure of Autistic Traits: Comparison of the Social Responsiveness Scale with the Autism Diagnostic Interview-Revised. *J Autism Dev Disord*, **33**, 427–433.

139. Sturm,A., Kuhfeld,M., Kasari,C. and McCracken,J.T. (2017) Development and validation of an item response theory-based Social Responsiveness Scale short form. *Journal of Child Psychology and Psychiatry*, **58**, 1053–1061.

140. Li,H. and Durbin,R. (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, **25**, 1754–1760.

141. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G. and Durbin,R. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

142. DePristo,M.A., Banks,E., Poplin,R., Garimella,K.V., Maguire,J.R., Hartl,C., Philippakis,A.A., del Angel,G., Rivas,M.A., Hanna,M., *et al.* (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*, **43**, 491–498.

143. Danecek,P., Auton,A., Abecasis,G., Albers,C.A., Banks,E., DePristo,M.A., Handsaker,R.E., Lunter,G., Marth,G.T., Sherry,S.T., *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.

144. Wang,K., Li,M. and Hakonarson,H. (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*, **38**, e164.

145. Landrum,M.J., Lee,J.M., Benson,M., Brown,G., Chao,C., Chitipiralla,S., Gu,B., Hart,J., Hoffman,D., Hoover,J., *et al.* (2016) ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res*, **44**, D862–D868.

146. Liu,X., Wu,C., Li,C. and Boerwinkle,E. (2016) dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs. *Human Mutation*, **37**, 235–241.

147. Kircher,M., Witten,D.M., Jain,P., O'Roak,B.J., Cooper,G.M. and Shendure,J. (2014) A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*, **46**, 310–315.

148. Purcell,S., Cherny,S.S. and Sham,P.C. (2003) Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics*, **19**, 149–150.

149. Purcell,S., Neale,B., Todd-Brown,K., Thomas,L., Ferreira,M.A.R., Bender,D., Maller,J., Sklar,P., de Bakker,P.I.W., Daly,M.J., *et al.* (2007) PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*, **81**, 559–575.

150. Lee,S., Wu,M.C. and Lin,X. (2012) Optimal tests for rare variant effects in sequencing association studies. *Biostat*, **13**, 762–775.

151. Turner,S.D. (2014) qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *bioRxiv*, 10.1101/005165.

152. Schaefer,M.L., Wong,S.T., Wozniak,D.F., Muglia,L.M., Liauw,J.A., Zhuo,M., Nardi,A., Hartman,R.E., Vogt,S.K., Luedke,C.E., *et al.* (2000) Altered stress-induced anxiety in adenylyl cyclase type VIII-deficient mice. *J. Neurosci.*, **20**, 4809–4820.

153. Barak,B., Zhang,Z., Liu,Y., Nir,A., Trangle,S.S., Ennis,M., Levandowski,K.M., Wang,D., Quast,K., Boulting,G.L., *et al.* (2019) Neuronal deletion of Gtf2i , associated with Williams syndrome, causes behavioral and myelin alterations rescuable by a remyelinating drug. *Nature Neuroscience*, **22**, 700.

154. Bayarsaihan,D. and Ruddle,F.H. (2000) Isolation and characterization of BEN, a member of the TFII-I family of DNA-binding proteins containing distinct helix–loop–helix domains. *PNAS*, **97**, 7342–7347.

155. Bayarsaihan,D., Makeyev,A.V. and Enkhmandakh,B. (2012) Epigenetic modulation by TFII-I during embryonic stem cell differentiation. *Journal of Cellular Biochemistry*, **113**, 3056–3060.

156. Peña-Hernández,R., Marques,M., Hilmi,K., Zhao,T., Saad,A., Alaoui-Jamali,M.A., Rincon,S.V. del, Ashworth,T., Roy,A.L., Emerson,B.M., *et al.* (2015) Genome-wide

targeting of the epigenetic regulatory protein CTCF to gene promoters by the transcription factor TFII-I. *PNAS*, **112**, E677–E686.

157. Oya,H., Yokoyama,A., Yamaoka,I., Fujiki,R., Yonezawa,M., Youn,M.-Y., Takada,I., Kato,S. and Kitagawa,H. (2009) Phosphorylation of Williams Syndrome Transcription Factor by MAPK Induces a Switching between Two Distinct Chromatin Remodeling Complexes. *J. Biol. Chem.*, **284**, 32472–32482.

158. Hinsley,T.A., Cunliffe,P., Tipney,H.J., Brass,A. and Tassabehji,M. (2004) Comparison of TFII-I gene family members deleted in Williams-Beuren syndrome. *Protein Science*, **13**, 2588–2599.

159. Deacon,R.M.J., Croucher,A. and Rawlins,J.N.P. (2002) Hippocampal cytotoxic lesion effects on species-typical behaviours in mice. *Behavioural Brain Research*, **132**, 203–213.

160. Kopp,N.D., McCullough,K., Maloney,S.E. and Dougherty,J.D. (2019) Gtf2i and Gtf2ird1 mutation are not sufficient to reproduce mouse phenotypes caused by the Williams Syndrome critical region. *bioRxiv*, 10.1101/558544.

161. Bolger,A.M., Lohse,M. and Usadel,B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.

162. Langmead,B., Wilks,C., Antonescu,V. and Charles,R. (2019) Scaling read aligners to hundreds of threads on general-purpose processors. *Bioinformatics*, **35**, 421–432.

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

164. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.

165. Robinson,M.D., McCarthy,D.J. and Smyth,G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.

166. Heinz,S., Benner,C., Spann,N., Bertolino,E., Lin,Y.C., Laslo,P., Cheng,J.X., Murre,C., Singh,H. and Glass,C.K. (2010) Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell*, **38**, 576–589.

167. Stamatoyannopoulos,J.A., Snyder,M., Hardison,R., Ren,B., Gingeras,T., Gilbert,D.M., Groudine,M., Bender,M., Kaul,R., Canfield,T., *et al.* (2012) An encyclopedia of mouse DNA elements (Mouse ENCODE). *Genome Biology*, **13**, 418.

168. Ramírez,F., Ryan,D.P., Grüning,B., Bhardwaj,V., Kilpert,F., Richter,A.S., Heyne,S., Dündar,F. and Manke,T. (2016) deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res*, **44**, W160–W165.

169. Dobin,A., Davis,C.A., Schlesinger,F., Drenkow,J., Zaleski,C., Jha,S., Batut,P., Chaisson,M. and Gingeras,T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.

170. Anders,S., Pyl,P.T. and Huber,W. (2015) HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*, **31**, 166–169.

171. Maloney,S.E., Akula,S., Rieger,M.A., McCullough,K.B., Chandler,K., Corbett,A.M., McGowin,A.E. and Dougherty,J.D. (2018) Examining the Reversibility of Long-Term Behavioral Disruptions in Progeny of Maternal SSRI Exposure. *eNeuro*, **5**.

172. Maloney,S., Yuede,C.M., Creeley,C.E., Huffman,J., Taylor,G., Noguchi,K. and Wozniak,D.F. (in press) Repeated neonatal isoflurane exposures in the mouse induce apoptotic degenerative changes in the brain with relatively mild long-term behavioral deficits. *Scientific Reports*.

173. Hothorn,T., Bretz,F. and Westfall,P. (2008) Simultaneous Inference in General Parametric Models. *Biometrical Journal*, **50**, 346–363.

174. Kopp,N.D., Parrish,P.C.R., Lugo,M., Dougherty,J.D. and Kozel,B.A. Exome sequencing of 85 Williams–Beuren syndrome cases rules out coding variation as a major contributor to remaining variance in social behavior. *Molecular Genetics & Genomic Medicine*, **0**.

175. Ghaffari,M., Birgani,M.T., Kariminejad,R. and Saberi,A. (2018) Genotype–phenotype correlation and the size of microdeletion or microduplication of 7q11.23 region in patients with Williams-Beuren syndrome. *Annals of Human Genetics*, **82**, 469–476.

176. Howald,C., Merla,G., Digilio,M.C., Amenta,S., Lyle,R., Deutsch,S., Choudhury,U., Bottani,A., Antonarakis,S.E., Fryssira,H., *et al.* (2006) Two high throughput technologies to detect segmental aneuploidies identify new Williams-Beuren syndrome patients with atypical deletions. *Journal of Medical Genetics*, **43**, 266–273.

177. Rieger,M.A. and Dougherty,J.D. (2016) Analysis of within Subjects Variability in Mouse Ultrasonic Vocalization: Pups Exhibit Inconsistent, State-Like Patterns of Call Production. *Front. Behav. Neurosci.*, **10**.

178. Lindzey,G., Manosevitz,M. and Winston,H. (1966) Social dominance in the mouse. *Psychon Sci*, **5**, 451–452.

179. Thomas,A., Burant,A., Bui,N., Graham,D., Yuva-Paylor,L.A. and Paylor,R. (2009) Marble burying reflects a repetitive and perseverative behavior more than novelty-induced anxiety. *Psychopharmacology (Berl)*, **204**, 361–373.

180. Mervis,C.B. and John,A.E. (2010) Cognitive and behavioral characteristics of children with Williams syndrome: Implications for intervention approaches. *Am. J. Med. Genet.*, **154C**, 229–248.

181. Strekalova,T., Zörner,B., Zacher,C., Sadovska,G., Herdegen,T. and Gass,P. (2003) Memory retrieval after contextual fear conditioning induces c-Fos and JunB expression in CA1 hippocampus. *Genes, Brain and Behavior*, **2**, 3–10.

182. Borralleras,C., Mato,S., Amédée,T., Matute,C., Mulle,C., Pérez-Jurado,L.A. and Campuzano,V. (2016) Synaptic plasticity and spatial working memory are impaired in the CD mouse model of Williams-Beuren syndrome. *Molecular Brain*, **9**, 76.

183. Nygaard,K.R., Maloney,S.E. and Dougherty,J.D. (2019) Erroneous inference based on a lack of preference within one group: autism, mice, and the Social Approach Task. *bioRxiv*, 10.1101/530279.

184. Knoll,A.T., Jiang,K. and Levitt,P. Quantitative trait locus mapping and analysis of heritable variation in affiliative social behavior and co-occurring traits. *Genes, Brain and Behavior*, 10.1111/gbb.12431.

185. Owen,E.H., Logue,S.F., Rasmussen,D.L. and J. M. Wehner (1997) Assessment of learning by the Morris water task and fear conditioning in inbred mouse strains and F1 hybrids: implications of genetic background for single gene mutations and quantitative trait loci analyses. *Neuroscience*, **80**, 1087–1099.

186. Brill,S., Li,S., Lyman,C.W., Church,D.M., Wasmuth,J.J., Weissbach,L., Bernards,A. and Snijders,A.J. (1996) The Ras GTPase-activating-protein-related human protein IQGAP2 harbors a potential actin binding domain and interacts with calmodulin and Rho family GTPases. *Molecular and Cellular Biology*, **16**, 4869–4878.

187. Schindelin,J., Arganda-Carreras,I., Frise,E., Kaynig,V., Longair,M., Pietzsch,T., Preibisch,S., Rueden,C., Saalfeld,S., Schmid,B., *et al.* (2012) Fiji: an open-source platform for biological-image analysis. *Nature Methods*, **9**, 676–682.

188. Young,M.D., Wakefield,M.J., Smyth,G.K. and Oshlack,A. (2010) Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biology*, **11**, R14.

189. Branchi,I., Santucci,D. and Alleva,E. (2001) Ultrasonic vocalisation emitted by infant rodents: a tool for assessment of neurobehavioural development. *Behavioural Brain Research*, **125**, 49–56.

190. Maloney,S.E., Chandler,K.C., Anastasaki,C., Rieger,M.A., Gutmann,D.H. and Dougherty,J.D. (2018) Characterization of early communicative behavior in mouse models of neurofibromatosis type 1. *Autism Research*, **11**, 44–58.

191. Holy,T.E. and Guo,Z. (2005) Ultrasonic Songs of Male Mice. *PLOS Biology*, **3**, e386.

192. Wozniak,D.F., Hartman,R.E., Boyle,M.P., Vogt,S.K., Brooks,A.R., Tenkova,T., Young,C., Olney,J.W. and Muglia,L.J. (2004) Apoptotic neurodegeneration induced by ethanol in neonatal mice is associated with profound learning/memory deficits in juveniles followed by progressive functional recovery in adults. *Neurobiology of Disease*, **17**, 403–414.

193. Moy,S.S., Nadler,J.J., Young,N.B., Nonneman,R.J., Segall,S.K., Andrade,G.M., Crawley,J.N. and Magnuson,T.R. (2008) Social approach and repetitive behavior in eleven inbred mouse strains. *Behavioural Brain Research*, **191**, 118–129.

194. Dougherty,J.D., Maloney,S.E., Wozniak,D.F., Rieger,M.A., Sonnenblick,L., Coppola,G., Mahieu,N.G., Zhang,J., Cai,J., Patti,G.J., *et al.* (2013) The Disruption of Celf6, a Gene Identified by Translational Profiling of Serotonergic Neurons, Results in Autism-Related Behaviors. *J. Neurosci.*, **33**, 2732–2753.

195. Nadler,J.J., Moy,S.S., Dold,G., Simmons,N., Perez,A., Young,N.B., Barbaro,R.P., Piven,J., Magnuson,T.R. and Crawley,J.N. (2004) Automated apparatus for quantitation of social approach behaviors in mice. *Genes, Brain and Behavior*, **3**, 303–314.

196. Wang,F., Zhu,J., Zhu,H., Zhang,Q., Lin,Z. and Hu,H. (2011) Bidirectional Control of Social Hierarchy by Synaptic Efficacy in Medial Prefrontal Cortex. *Science*, **334**, 693–697.

197. Grady,R.M., Wozniak,D.F., Ohlemiller,K.K. and Sanes,J.R. (2006) Cerebellar Synaptic Defects and Abnormal Motor Behavior in Mice Lacking α- and β-Dystrobrevin. *J. Neurosci.*, **26**, 2841–2851.

198. Brown,J.A., Emnett,R.J., White,C.R., Yuede,C.M., Conyers,S.B., O’Malley,K.L., Wozniak,D.F. and Gutmann,D.H. (2010) Reduced striatal dopamine underlies the attention system dysfunction in neurofibromatosis-1 mutant mice. *Hum Mol Genet*, **19**, 4515–4528.

199. Boyle,M.P., Kolber,B.J., Vogt,S.K., Wozniak,D.F. and Muglia,L.J. (2006) Forebrain Glucocorticoid Receptors Modulate Anxiety-Associated Locomotor Activation and Adrenal Responsiveness. *J. Neurosci.*, **26**, 1971–1978.

200. Brody,S.A., Nakanishi,N., Tu,S., Lipton,S.A. and Geyer,M.A. (2005) A Developmental Influence of the N-Methyl-D-Aspartate Receptor NR3A Subunit on Prepulse Inhibition of Startle. *Biological Psychiatry*, **57**, 1147–1152.

201. Wozniak,D.F., Xiao,M., Xu,L., Yamada,K.A. and Ornitz,D.M. (2007) Impaired spatial learning and defective theta burst induced LTP in mice lacking fibroblast growth factor 14. *Neurobiology of Disease*, **26**, 14–26.

202. Yuede,C.M., Olney,J.W. and Creeley,C.E. (2013) Developmental Neurotoxicity of Alcohol and Anesthetic Drugs Is Augmented by Co-Exposure to Caffeine. *Brain Sciences*, **3**, 1128–1152.

203. Fox,J. and Weisberg,S. (2011) An R Companion to Applied Regression Second. Sage.

204. Jensen,M. and Girirajan,S. (2019) An interaction-based model for neuropsychiatric features of copy-number variants. *PLOS Genetics*, **15**, e1007879.

205. Khera,A.V., Chaffin,M., Aragam,K.G., Haas,M.E., Roselli,C., Choi,S.H., Natarajan,P., Lander,E.S., Lubitz,S.A., Ellinor,P.T., *et al.* (2018) Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature Genetics*, **50**, 1219–1224.

206. Dai,L., Carter,C.S., Ying,J., Bellugi,U., Pournajafi-Nazarloo,H. and Korenberg,J.R. (2012) Oxytocin and Vasopressin Are Dysregulated in Williams Syndrome, a Genetic Disorder Affecting Social Behavior. *PLoS ONE*, **7**, e38513.

207. Guzmán,Y.F., Tronson,N.C., Jovasevic,V., Sato,K., Guedea,A.L., Mizukami,H., Nishimori,K. and Radulovic,J. (2013) Fear-enhancing effects of septal oxytocin receptors. *Nature Neuroscience*, **16**, 1185.