Winter 12-15-2022

Extensive Behavioral Phenotyping of Williams Syndrome Locus Relevant Mouse Models to Assess Contributions of Oxytocin and Gtf2ird1

Kayla Rose Nygaard
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology and Biomedical Sciences
Molecular Genetics and Genomics

Dissertation Examination Committee:
Joseph Dougherty, Chair
Ream Al-Hasani
John Edwards
Heather Lawson
Jose Moron-Concepcion

Extensive Behavioral Phenotyping of Williams Syndrome Locus Relevant Mouse Models to Assess Contributions of Oxytocin and Gtf2ird1

by
Kayla R. Nygaard

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

August 2022
St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

CHAPTER 2

CHAPTER 3

APPENDIX

Acknowledgments

Writing and organizing a dissertation is difficult –not to mention the effort to produce the work that fills its pages– and though few may read it, this document feels life-changing, especially from the perspective of a neurodivergent, first-generation college graduate who dreams big, yet still expects relatively little from her life. There were more moments of doubt than I let anyone know, and as I constantly struggle to appreciate myself and acknowledge my worth, I would be remiss if I did not acknowledge my own contributions to this slow, painstaking process because without my perseverance, these pages would be blank. That said, I must also recognize the people who made the road a little smoother.

Indeed, my load was much lighter because of those of you who moved metaphorical stones, supported me across the desolate valleys, and celebrated at the triumphant, yet always temporary, peaks. My journey at WashU started in a basement watching other people play ping pong – Jim Skeath tipped the scales in favor of St. Louis, and I am forever grateful. It was a gift to be part of the MGG program and the larger umbrella of DBBS, which were both filled with great people at all levels, including those who fill my thesis committee and provided marvelous insight and exciting ideas, even if I never got to see them through. The optional academic and professional support from DBBS was just as useful as the required courses, helping me achieve the NSF Fellowship that funded three of my years and providing space to explore various careers. The pre-Covid social opportunities connected me with my "Nerd Herd" friends who successfully ruined my former-teacher adult sleep schedule, while supplying endless good times. Thanks especially to Brian and Jennie, who feel like family.

Finding a space in the Dougherty lab may have been the ultimate key to my success and made earning my doctoral degree decidedly more enjoyable. From my original mentor, Nathan, to

my cohort buddy, Tomás, every single person in the lab has brightened each day and I thank you all for making this journey, dare I say, fun! Thanks especially to Katie, Raylynn, Mike, and Rachel who were incredibly generous with their time, and helped take care of mice, dissect brains, run behavior, and analyze data – I couldn't have done it all without you. Susan Maloney, you deserve more than a sentence for the incredible support and added mentorship you have given me. Overcoming disappointing revelations and barreling through mountains of data was only possible with your unwavering and kind support. Finally, thank you to Joe, a superb mentor, who is always cheering us on, whether we are excited by the data, struggling to write about it, or setting it to the side to battle in one of the various Doc Lab Tournaments. I am in awe of the incredible team you continue to build and nurture, and I appreciate all the ways you mindfully capitalize on learning opportunities for us all. Thank you so much for never giving up on me, and for showing me how being a productive scientist and a caring human can easily go hand in hand.

A huge thank you to my family for letting me wander through most of my childhood with my nose in a book, whether I was supposed to be sleeping or watching a race or enjoying family Christmas. Special thanks to my grandpa Don who fueled my love of learning with his own penchant for a large library, to my dad for building Komodo dragons out of chicken wire with me, and to my mom who drove me to Carleton to deliver my college enrollment check on the last possible day (please note, I was accepted after submitting an essay about how much I loved books). Perhaps I should also be thanking all the authors of those books... I also want to thank my acquired family who provided support in kind words, delicious dinners, and beautiful flowers. I appreciate all the support from the Conrad and Roach families I am lucky enough to be part of now!

Jimmy, I owe you a lifetime of thanks. I am so glad I accosted you almost exactly 10 years ago in a Super 8 lobby in Hazard, KY. Since then, your support has never wavered, despite how

hard I try to make it. Thank you for feeding me and watering me and generally keeping me alive. Thank you for letting me cry and appreciating the compassion I cannot control that interferes with life more than a little. Thank you for trying to get me to go to bed and don't feel bad; even my own parents failed at that, as I stayed awake long after lights out, reading by the light of the moon. Thank you for giving me 6 years in St. Louis and thank you for encouraging me to even get this far in my academic journey. Finally, thank you for being an amazing dad and for taking a disproportionately large share of the parenting load recently. You make my dreams possible, and I love you.

Rory, Rory, mommy and daddy love you! Thank you for helping me see the joy in the little things again and for being the silliest little human. Though you drain my energy like no other, you also bring me to life. Thank you so much for your smiles and your giggles and your incessant need to be near me. I can't imagine my life without you.

Kayla R. Nygaard

Washington University in St. Louis

August 2022

Dedicated to my grandfather, Don Estrem, who instilled in me a love of learning and gifted a book every Christmas

to Jimmy, who made it all possible

and to Rory, whose birth was debatably less difficult, though decidedly more dangerous, than delivering this dissertation

ABSTRACT OF THE DISSERTATION

Extensive Behavioral Phenotyping of Williams Syndrome Locus Relevant Mouse Models to Assess Contributions of Oxytocin and Gtf2ird1

by

Kayla R. Nygaard

Doctor of Philosophy in Biology and Biomedical Sciences

Molecular Genetics and Genomics

Washington University in St. Louis, 2022

Professor Joseph Dougherty, PhD, Chair

The Williams Syndrome Critical Region (WSCR) at chromosome 7q11.23 provides a unique opportunity to untangle the relationship between genotype and phenotype in complex behaviors, from fear and anxiety to sociability and sensorimotor processing. Copy number variations (CNVs) in this region result in two syndromes, Williams Syndrome (WS) and Duplication 7q11.23 Syndrome (Dup7), which display phenotypes that may align, indicating a common disruption of a system, or diverge, reflecting an underlying gene dosage-dependent effect. While case studies of atypical deletions resulting in WS have implicated telomeric genes *Gtf2ird1* and *Gtf2i* in the cognitive and behavioral profiles of WS, proving causation requires utilizing mouse models. Thus, I leveraged the construct validity of a mouse line modeling the most common deletion in WS to assess pharmacological and genetic interventions in an attempt to ameliorate deficits caused by CNVs in the WSCR. I assessed the role of oxytocin in fear conditioning deficits observed in the Complete Deletion (CD) mice and show that an oxytocin antagonist delivered to the central nervous system does not rescue the contextual and cued recall impairments, suggesting no direct

role for oxytocin dysregulation in these features of the CD model. No significant differences in oxytocin receptor density or distribution were found either. I then present a novel transgenic model designed to overexpress *Gtf2ird1*, one of the genes implicated in the hallmark cognitive and behavioral features of WS and characterize the effect of its overexpression on a C57BL/6J wild type background and its molecular rescue of *Gtf2ird1* expression on the CD background in a comprehensive assessment of sensorimotor, anxiety, fear, and social behaviors. Deficits in the CD model are shown in all of these domains to various degrees and while *Gtf2ird1* did not play a role in the enhanced social approach or motivation observed in the CD line, it did ameliorate deficits in three tasks (Platform, Rotarod and Light/Dark Box). These results may support the idea that *Gtf2ird1* is involved in sensory processing, which has been suggested particularly regarding the visuospatial deficits seen in WS.

Chapter 1: Introduction

The field of molecular genetics aims to understand how the genetic code influences the variable traits expressed by organisms. This task of linking genotype to phenotype can be quite daunting given the vast expanse of the human genome. Exploring the natural variation within the genome is a useful starting point for studying the genetic basis of behavior, especially in the context of neurodevelopmental disorders. With various etiologies resulting in collections of behavioral traits that can affect sociability, anxiety, fear, and movement to varying degrees, understanding the genetic basis and mechanistic underpinnings of the symptoms characterizing these disorders can be difficult. Disorders with a defined genetic cause supply a concrete starting point. For example, while both Autism and Williams Syndrome result in motor coordination difficulties, social differences, and increased anxiety, only Williams Syndrome is caused by a deletion of a single region of the genome with no evidence of external risk factors, thus limiting the initial search to the genes within that locus and eliminating the need to consider environmental effects.[1]

Of the phenotypes described above, social behavior presents a particular challenge. Despite the obvious inheritability evident in numerous neurodevelopmental disorders, the mechanisms influencing abnormal social responses in these disorders have been woefully underexplained. The oxytocin system is a known player in social behaviors and is a favored potential pharmacological target for social disorders, despite the challenges to measuring the neuropeptide and limited evidence for any genetic causes potentially disrupting the oxytocin system in these disorders. Identifying clear mechanistic avenues underlying oxytocin disruption will be imperative in developing the most effective treatments for neurodevelopmental disorders with varied causes.

In this thesis, I focused on the genetic locus responsible for a set of neurodevelopmental disorders, Williams Syndrome and 7q11.23 Duplication Syndrome, to investigate genetic contributions to motor, social, anxiety and fear behaviors. Using a mouse modeling the most common deletion of the locus, I assessed the impact of oxytocin on fear learning by delivering an oxytocin antagonist to the brain prior to the fear conditioning and recall. To better characterize this same mouse model for use in future studies, I broadly evaluated behaviors using tasks to address social, anxiety, fear, and motor domains. Simultaneously, I validated a novel mouse line and used it to investigate the impact of a single gene within the region, *Gtf2ird1*, assessing both the effect of its overexpression alone and its rescue on the complete deletion background.

1.1 The Williams Syndrome Locus

The Williams Syndrome locus, also known as the Williams Syndrome Critical Region (WSCR), is found on chromosome band 7q11.23 where three large low-copy repeats facilitate nonhomologous allelic recombination during meiosis, leading to a pair of neurodevelopmental disorders caused by the deletion or duplication of 25-27 genes.[2,3] While these opposite copy number variations (CNVs) in the WSCR presumably occur at similar rates in the population, more is known about the deletion, which causes Williams Syndrome (WS), as it was discovered decades prior to the duplication resulting in 7q11.23 Duplication Syndrome (Dup7).[4,5]

1.1.1 Deletion: Williams Syndrome

Williams Syndrome (a.k.a. Williams-Beuren Syndrome) was first characterized in the 1960s separately by Dr. Williams and Dr. Beuren, who both noted distinctive facial features and intellectual disability in patients with supravalvular aortic stenosis (SVAS).[4,6] The genetic basis

for WS, a deletion of ~2Mb on chromosome 7, was determined in the 90s.[2,7] Most individuals with WS lose one copy of the genes between *GTF2I* and *FKBP6*, which results in diverse symptoms spanning cardiovascular, endocrine, gastrointestinal, and neurological systems, among others. It is a rare disorder, estimated to be found in 1 out of every 7,500 individuals, though it is likely underdiagnosed due to variable expression of the core features.[8] Initial diagnosis usually follows discovery of abnormal cardiac features, developmental delay, or facial dysmorphism, which are prevalent characteristics in WS.[9,10]

While not every individual with WS has SVAS, or aortic narrowing, it is the most common cardiac feature, and its presence typically results in an earlier diagnosis due to the interventions required to manage it. The distinctive facial features observed often include bitemporal narrowing, a short nose with a flat bridge, a long, smooth upper lip, and fullness in the eyes, cheeks and lips (**Figure 1A**).[1,2] Other common features initially noted to affect more than 75% of the patients in a 1996 study by Pérez Jurado *et al.* include irritability in infancy, mild to moderate intellectual disability, an outgoing personality with relatively strong language capabilities, a hoarse voice, and sensitivity to sound. Despite sensitivity to sound, individuals with WS also tend to enjoy music and have good auditory short-term memory relative to their broad cognitive abilities, which may play into their unique cognitive profile.[11]

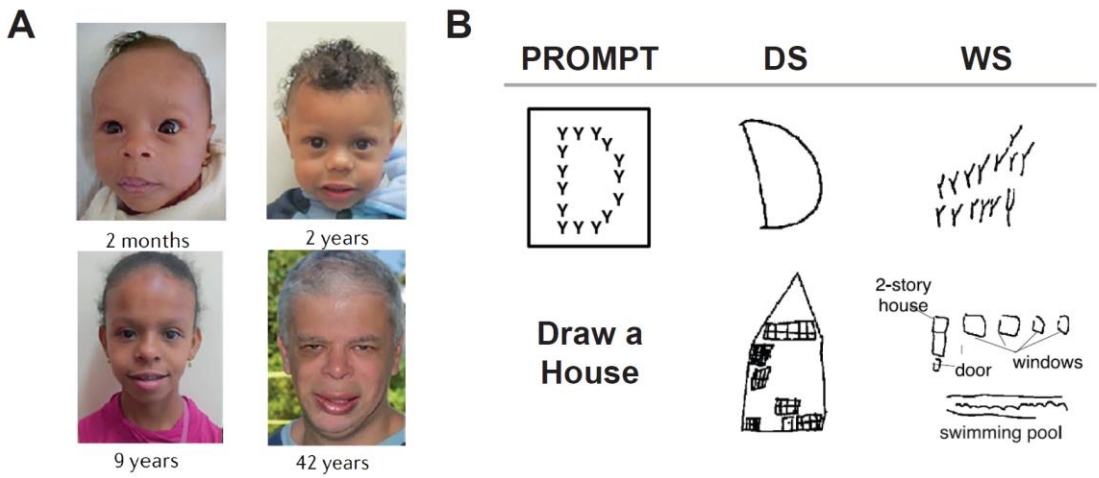

Figure 1. Features of Williams Syndrome. A) Characteristic facial presentation over a range of ages, adapted from Kozel *et al*. (2021).[1] B) Two examples of the tendency for individuals with WS to focus on details without proper integration, compared to individuals with Down Syndrome (DS), modified from figures in Bellugi *et al.* (2000).[12]

This distinct cognitive profile of relative strengths and weaknesses defines individuals with WS, which includes a surprising disconnect between general cognitive ability and language, especially vocabulary acquisition.[12,13] Despite lower than average IQ and a delay in language development, individuals with WS are quite talkative; the WS deletion appears to relatively spare expressive linguistic ability, while producing deficits in visuospatial processing, which are reflected in their language, among other measures.[12,13] Individuals with WS will often misuse prepositions when describing the relationship between objects.[12] They also pay more attention to details but lack integration of the whole, which is illustrated in their ability to draw individual parts of house but not orient them correctly with respect to each other (**Figure 1B**).[12]

Surprisingly, these global visuospatial processing deficits do not affect their ability to discriminate between or remember human faces, though their facial processing is atypical.[14] The interpersonal language and facial recognition strengths of the WS cognitive profile support the typical hypersocial personality observed in WS. In addition to loquaciousness and sociability, other common cognitive sequalae include hyperactivity and anxiety.[15] These elements will be

addressed in more detail below, in addition to other sensorimotor and fear-related neurological aspects of WS.

Social Behaviors

One of the most noted features of the WS personality profile is an extremely friendly, outgoing demeanor, especially in childhood.[16,17] Children with WS have higher levels of sociability compared to developmentally similar individuals with Down syndrome as well as typically developing children.[18] This increased social behavior is evident in increased approach, facial orienting, and emotional reactivity to social stimuli, and does not necessarily indicate proficiency in typical social interactions. Individuals with WS have a greater tendency to approach others regardless of familiarity, though the difference is more pronounced with strangers and in children under 4.[18] This indiscriminate approach is evident in their high ratings of approachability for faces regardless of expression – even very angry faces appear to them quite approachable.[19] The abnormally high attention to faces has even interfered with studies of emotionality in children with WS, as the children were more engaged with the researcher than the task.[20,21]

Quantification of eye tracking, which has been shown to be under genetic control,[22] shows that individuals with WS look longer at eyes and faces,[23–25] which may stem from trouble with disengaging rather than quick engagement.[25,26] Additionally, a recent study investigated a hypothesis of hypo-arousal facilitates longer eye gaze; while they showed individuals with WS have slower saccadic movement (indicative of hypo-arousal), they also revealed a deficit in the typical ability to preferably orient to the eyes during a cued test.[27] This suggests hypo-arousal doesn't necessarily facilitate prolonged eye gaze; it may depend on the stage of social interaction

(initiation vs continuation). These facial processing differences may explain the challenges people with WS face in social interactions despite their noticeably high level of social motivation.

The increased social drive may also be influenced by a difference in processing of fear-relevant facial stimuli. Social behavior measures collected by the Salk Institute Sociability Questionnaire were correlated with amygdalar responses to fearful faces in people with WS.[28] The fMRI measurements revealed a relationship where a decreased response in left amygdala response predicted an increase in social approach behaviors of the individual. This relationship was specific to fearful faces. Difficulty with orbitofrontal cortex regulation of the amygdala may be a factor in the social disinhibition, and would explain another symptom, specific phobias, that is common in WS as well.[17]

Anxiety and Fear

Despite being recognized as outgoing and talkative, individuals with WS are also noticeably insecure and anxious.[29] There is a greater prevalence of generalized anxiety disorder and specific phobias in the population, at 10-12% and 39-54% respectively.[30,31] The prevalence of GAD was greater in older subjects. A 4-year study reported persistent high levels of anxiety that were related to performance in social and executive functioning.[32] Their sociability (higher approach, attention to faces, and emotional reactivity) appears to stem from social disinhibition and does not provide them with the skills to navigate the intricacies of social interactions to engage beyond surface-level conversations. Thus, in adulthood, people with WS tend to be more withdrawn.[33] Critically, the relationship between anxiety levels and social skills may be driven by executive dysfunction. This provides a key focus for intervention and treatment of anxiety as

executive disfunction is a symptom of ADHD, the prevalence of which reaches 65% of children with WS.[30]

The nature of specific phobias provides some additional insight into the WS phenotype. Top specific phobias reported in one study include roller coasters, shots, thunderstorms, loud noises, and being in a fight.[34] Sensitivity to sound appears to play a role in some of the fears common to WS, including the loud noises associated with thunder, sirens, fireworks, and vacuum cleaners. Levitin et al. report that in a group of 118 people with WS, 80% experience odynacusis, or a decreased threshold in uncomfortable loudness, and 90% have auditory aversions, which tend to decrease over time unlike the stable aversions present in the autistic or control group.[35] Interestingly, 9% of WS individuals in this sample also had auditory "fascinations", which are sounds they find particularly captivating; this is made even more peculiar because the targets of their fascinations are stimuli that had previously been aversive. Regardless, it is clear that sensory stimuli, especially sound, are significant in triggering specific phobias, and thus sensory features of WS are an important piece of the whole when considering WS neurocognitive behavioral symptoms.

Sensorimotor Features

Hyperacusis is a commonly noted feature of WS and is likely responsible for the phonophobia observed in WS, though one group of researchers clarifies that what is being called hyperacusis is more accurately odynacusis or auditory aversion, which are both highly prevalent in WS.[34,35] In one study of 69 patients, 96% of individuals reported a sound hypersensitivity, especially noticed in their amplified startle response to common sounds.[36] This increased response to sound appeared to influence attention issues and anxiety leading to significant behavioral

problems. In another study of 49 children, 84% had hyperacusis from infancy and were sensitive to sounds at levels 20dB below where control subjects reported discomfort.[37] Hearing loss of high frequencies was also observed, in addition to outer hair cell impairments indicating cochlear dysfunction, despite normal hearing thresholds and middle-ear function.[37,38] In one study of 38 Swedish individuals with WS, females were reported as more sensitive to sound and likely to have sound-related fears.[34] While the auditory system is obviously impacted in WS and should continue to be studied to further investigate possible sex differences, it is not the only sensory system affected.

A variety of sensorimotor differences are also characteristic of WS across domains including balance, gait, tone, and sensory modulation.[17,39] Gross developmental milestones are delayed in children with WS, who are slower to achieve head support, sitting without support, and walking without support, in addition to acquiring over five meaningful words.[9] The low motor tone seen in younger children is replaced with hypertonia as they get older.[39,40] Individuals with WS also commonly show motor coordination deficits in walking, especially over uneven surfaces or descending stairs, and in using tools requiring gross and fine motor abilities.[36,41,42] While these balance deficits tend to improve as individuals age,[33] coordination and gait issues persist.[39] Motor coordination deficits are especially different from nonaffected individuals when visual information was incorporated (e.g., eyes open).[43] Sensory vision is also impacted by the WSCR deletion, but these vision deficits (strabismus, decreased visual acuity, etc.) are distinct from the challenges individuals have with visuospatial tasks suggesting a sensory processing issue.[44] Indeed, integration of depth perception seems to be impaired more than simple depth perception, though that is also impaired in WS.[41,44] Dorsal stream visual dysfunction has been suggested to explain the processing deficits in WS observed in an extensive battery of test measuring vision and

visuospatial abilities.[44,45] Sensory processing dysfunction may even help explain certain aspects of sociability and anxiety, and could be considered a core feature of WS, though this hypothesis is relatively unexplored.[46]

1.1.2 Duplication: 7q11.23 Duplication Syndrome

The research on Dup7 is decades behind that of WS. Though prevalence of the reciprocal genetic syndromes should be similar, its discovery occurred by chance ~40 years after WS was described, when the elastin gene was found to be duplicated in a boy being tested for 22q11 deletion.[5] In addition to delayed discovery, underdiagnosis is a problem, likely due to the milder and less unique suite of characteristics that result from the WSCR duplication compared to deletion. Once Dup7 was characterized, though, similarities between the systems affected by both syndromes were obvious. Like WS, altered facial features have been noted in Dup7, though these craniofacial abnormalities tend to be minor and not as pervasive or distinct as in WS.[47] A few of the characteristics most often reported include a prominent forehead, a high/broad nose, a short philtrum, and thin lips.[47,48] These features are in some ways opposite those in the WS facial suite, and cardiovascular features are similarly opposed, with aortic dilation more likely in Dup7.[48–50]

Developmentally, speech and motor delays are the most prominent symptoms while visuospatial development is largely spared.[47,51] Behavioral symptoms include anxiety, repetitive and stereotyped movements, and increased difficulty with communication and social interaction.[47] Hyperactivity, aggression, and self-harm were seen more in males, though a comprehensive analysis of sex differences has not been done. These characteristics may or may not be representative of the entire Dup7 population, however, since they are based on only a limited

sample, and we see great variability in WS symptoms as well. As awareness, and thus diagnosis, of Dup7 increases, the true range and prevalence of characteristics will become clearer.

Social Behaviors

Social behavior is obviously impacted in Dup7, which is associated with an increased prevalence of autism, in addition to significant speech delays.[47] The social characteristics of Dup7, at first glance, appear to be the opposite of the hypersociability described in WS, however significant similarities are also present. Duplication of the WSCR is associated with speech delays, which are also seen in WS, though autistic behaviors including decreased social interactions and the presence of repetitive interests are more obvious in Dup7.[47] These features result in an increased likelihood of an autism diagnosis.[52] Interestingly, a diagnosis of autism is also more likely in WS. The typical WS deletion is more associated with a so-called "odd-type" autism, where the motivation subscore of the Social Responsiveness Scale is not as severe as the other subscores, but larger or otherwise atypical deletions are associated with classical autism (i.e., all subscores reach clinical significance).[53,54]

Anxiety and Fear

Anxiety and ADHD are symptoms common to both WS and Dup7, though individuals with the duplication appear to have more of an issue with aggression, with instances of oppositional disorder also noted in the literature.[48,55] Anxiety was the feature most likely to raise concerns in parents, affecting 38 of 53 children in a large survey of individuals with Dup7.[48] Separation anxiety and social anxiety were the most commonly mentioned concerns.

Sensorimotor Features

Motor delays were another common feature in both Dup7 and WS. Both syndromes show developmental delays in motor abilities. Cerebellar dysfunction and hypotonia affect over half of people with Dup7, and abnormal gait affects is common.[48] Seizures and epilepsy are more common in Dup7 than WS (though atypical deletions can increase the likelihood).[51,56] In a study of 53 children with Dup7, 26% of parents surveyed reported a high tolerance for pain in their children, [48] which appears opposite to the sensitivity (in response to sound, at least) seen in WS.

The striking differences in some of the features in WS and Dup7, along with obvious similarities, makes the region responsible an interesting target to study the genetic basis of these symptoms, especially as some elements appear very responsive to copy number. The system most sensitive to copy number in Dup7 appears to be the central nervous system, with a high number of neurological symptoms.[48]

1.1.3 Genotype to Phenotype Connections in WS

While WS and Dup7 are rare, the CNVs in the WSCR provide a unique opportunity to connect genotype with phenotype. For example, elastin, a gene in the center of the WSCR, has long been connected to abnormal cardiac phenotypes, even outside of WS or Dup7. By leveraging the variable ploidy (haploid, triploid) of the region in WSCR CNVs, specific single nucleotide polymorphisms (SNPs) in the elastin gene were found to be associated with aortic arteriopathies.[57] The WSCR also provides potential for investigating the dose-dependence of genes; while the same systems are often affected in WS and Dup7, opposite phenotypes can arise. In the case of the aortic arteriopathies, stenosis (narrowing) was characteristic of WS patients while dilation is more prominent in those with Dup7.

While the cardiovascular features of WS and Dup7 appear to be largely monogenic, mainly stemming from dysregulation of the elastin gene, social behavior (which also appears to be sensitive to copy number) may be dependent on multiple genes within the region. Thus, it is more complicated to search for the genetic causes of the aberrant behavior caused by deletion or duplication in the WSCR. Atypical deletions have been especially important in identifying candidates to explore further. While 95% of WS deletions span 25 genes or ~1.5 Mb, a smaller number of cases are caused by a 1.8 Mb deletion, which includes two extra genes (*GTF2IRD2* and *NCF1*), and even fewer cases result from smaller deletions, which leave some genes unaffected.

Atypical Deletions

The atypically large or distal deletions have been shown to increase seizure susceptibility, and cognitive ability, in addition to the increased likelihood of autism mentioned earlier.[53,58–61] Shorter atypical deletions have been useful in narrowing the candidates for the cognitive and behavioral profiles of WS to the telomeric end of the WSCR (**Figure 2**).[62] As early as 1999, the telomeric end was hypothesized to be largely responsible for the main features of WS, when an atypically short deletion sparing the centromeric end of the region, past *ELN*, resulted in the full WS phenotype.[63]

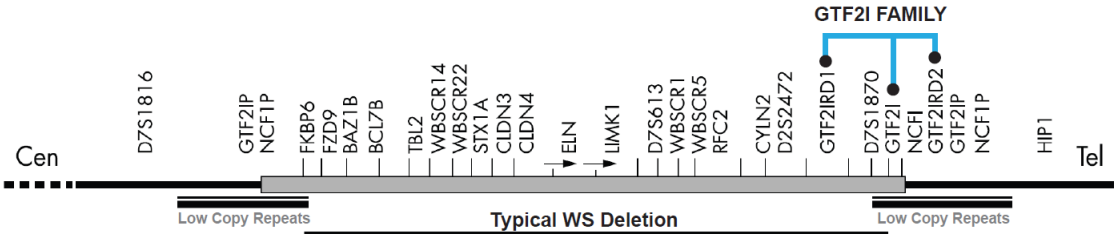

Figure 2. Representation of the WSCR. The typical 1.5 Mb deletion causing WS is shown as a solid line below the chromosome representation marked with the assorted genes of the region. The *GTF2I* family genes (*GTF2IRD1*, *GTF2I*, and *GTF2IRD2*) at the telomeric end (Tel) are indicated. *GTF2IRD2* is not typically deleted. *GTF2IP*, found on either end of the WSCR, is a pseudogene. Figure adapted from Karmiloff-Smith *et al.* (2003).[62]

The *GTF2I* family of genes encoding transcription factors has been a particular focus in the literature. The craniofacial features common to WS have been linked to *GTF2IRD1* in a variety of studies,[64,65] and atypical deletions sparing genes within the telomeric end of the WSCR support the idea that *GTF2IRD1* and *GTF2I* play a role in the social and cognitive symptoms of WS.[66] A boy with an atypical Williams Syndrome deletion that did not include *GTF2IRD1* or *GTF2I* presented with SVAS, mild WS facial features, and slight cognitive impairment but did not show the typical WS behavioral or cognitive profile (i.e., lacked visuospatial deficits).[67] A girl with a shorter telomeric deletion that did not include *GTF2I* and maybe *GTF2IRD1* also lacked the atypical increase in social approach but did have the typical cardiovascular and craniofacial characteristics of the syndrome.[18]

As sequencing technology improved and breakpoints were able to be determined more specifically, studies were able to map genotype-phenotype correlations more precisely. In a study leveraging an atypical deletion to tease apart functions of *GTF2IRD1* and *GTF2I*, a female patient's atypical deletion was caused by a breakpoint between *GTF2IRD1* and *GTF2I*, so *GTF2IRD1* was included in the deletion but *GTF2I* was spared.[65] In this study, the patient had canonical WS facial features, cardiovascular deficits, abnormal gait, and attraction to music, but lacked the typical hypersocial behavioral profile as well as attention issues, hypersensitivity to sound, anxiety, or language and motor delays. This study suggests that *GTF2IRD1* plays the biggest roles in the WS craniofacial features and the visuo-spatial construction issues present in WS individuals, while gaze and social attention were attributed to *GTF2I*.

Other studies disagree, indicating *GTF2IRD1* as responsible for behavioral features. In a family with a similar atypical deletion (where *GTF2I* was spared, but *GTF2IRD1* deleted), the hypersociability was present but without accompanying intellectual disability.[1,66,67] The atypical

deletion from *ELN* to *GTF2IRD1* was associated with SVAS, an outgoing personality, and mild WS facial characteristics, but without visuospatial abnormalities in family members with the deletion.[66,67] These studies provide support for involvement of *GTF2IRD1* in social behavior, contrary to the previous case study.

These contradictions point to some of the limitations of only using atypical deletions to map phenotype to genotype. In addition, atypical deletions are by definition less common, making up less than 5% of the WS population, and inconsistent collection of data precludes widespread analysis of certain features like social behavior.[53,68] Thus, while atypical CNVs can be useful in establishing genes of interest, their rarity limits statistical confidence in any associations uncovered and furthermore cannot establish causation, only correlation. Human iPSCs have been used to overcome some of these challenges, allowing for single gene deletions and rescues,[69,70] but have their own limitations, including the lack of biologically relevant context and inability to represent behavioral features or assess the effect of potential pharmacological interventions directly on those behavioral differences. Thus, mouse models are imperative for assessing pharmacological interventions, such as oxytocin system manipulation, which I discuss in Chapter 2, and when combined with deep phenotyping of relevant models, as I show in Chapter 3, mouse models can provide important information on the roles of individual genes, and highlight anomalies in behavioral circuits that can then be leveraged to assess therapeutic approaches.

1.2 The Many Roles of Oxytocin

Oxytocin (OT), meaning ‘quick birth’ in Greek, is a neurotransmitter most well-known as the ‘love hormone’, but has a much wider role in biology. At the turn of the 20th century, this small molecule, which is produced in the hypothalamus and peripherally released from the pituitary

gland, was linked to increasing blood pressure and inducing uterine contractions, where it ultimately derived its name.[71,72] In 1910, it was found to also induce milk secretion.[73,74] It wasn't until 1927 that oxytocin was officially separated from vasopressin, which shares similar functions, though at the time they were called pitocin and pitressin.[75] Its nine peptides were determined in 1953 and first synthesized in 1954 by du Vigneaud, which help to secure him a Nobel Prize in chemistry.[76–80] The genes responsible for producing the oxytocin precursor (*OXT*) and oxytocin's single receptor (*OXTR*) were identified in the late 20th century.[81,82]

Since then, a lot of work has been done with the small nonapeptide, and oxytocin has been connected to more than just childbirth-related purposes. While the public focus has largely been on social behavior and oxytocin is currently being explored to treat neurodevelopmental disorders with characteristic social behaviors,[83,84] oxytocin also impacts anxiety, feeding, pain perception and learning and more.[85] In one incredibly comprehensive review, the list of features affected by oxytocin includes cardiovascular regulation, analgesia, motor activity, thermoregulation, gastric motility, osmoregulation, sexual, maternal, and social behaviors, stress-related behaviors, feeding and grooming, memory and learning, and opioid tolerance and dependence.[86] With such a wide sphere of influence, fully reviewing the effects of oxytocin is not feasible here, thus I will focus on oxytocin's effects on social, anxiety/fear, and sensorimotor behaviors, as I did above for the WSCR syndromes.

1.2.1 Oxytocin in the Central Nervous System

Social Behaviors

Oxytocin has been connected to numerous pro-social behaviors through human and animal studies. In the classical studies of voles, oxytocin was shown to be critical for pair bonding and

oxytocin receptor patterning was distinct in monogamous compared to polygamous vole species.[87–90] In an oxytocin knockout mouse, social deficits include decreased pup vocalizations after maternal separation and failure to recognize conspecifics that should be familiar.[91] Oxytocin's role in eye gaze, which is critical to navigating social interactions, provides a possible explanation for the differences in social engagement when OT is dysregulated. A double-blind study with a placebo control showed how oxytocin administration intranasally increased gaze time and number of fixations on human faces.[92] Unfortunately, this study was only done in male subjects, so whether this effect also applies to females is unknown. Sex effects to oxytocin would not be unexpected, when even dog breed can affect the oxytocin system: A study in dogs revealed breed differences in reactivity to oxytocin in the context of social interaction with owners.[93] Specifically, Border Collies were more susceptible to the effects of oxytocin as measured by gaze-relevant behaviors (looking at owner or experimenter). In addition to eye gaze behaviors, reaction to the eye gaze of others is also affected by oxytocin. An imaging study connected oxytocin and a SNP in its receptor to amygdala reactivity in response to direct gaze. Individuals with the CT variant were more reactive compared to people with the TT variant.[94]

Anxiety and Fear

While oxytocin is generally considered to be anxiolytic, or anxiety-reducing, its relationship with fear and fear learning is more complex. Fear and anxiety are related but defined by distinct situations; Fear requires a threat, while anxiety occurs in the absence of any discrete threat. In a review of oxytocin's complicated role in fear and anxiety behaviors, it is argued that OT can both reduce maladaptive anxiety but facilitate adaptive fear.[95] This relationship may not be so straightforward, however, as oxytocin was also shown to reduce amygdala reactivity when

fear-inducing stimuli were presented.[96] This complexity is obvious in oxytocin's effect on fear learning, which varies with stress, brain structure, and time of administration.[95] In the central amygdala, OT works to reduce contextual fear responses,[97] but its receptors in the bed nucleus of the stria terminalis is connected with facilitation of the fear response to a discrete cue.[98]

Sensorimotor Features

Oxytocin is also connected to sensory perception and motor activity. In a study in ovariectomized rats, administration of OT increased motor activity and lowered corticosterone and nociception.[99] Automatic motor stimulation is also affected by OT,[100] and delivery of OT can prevent the motor coordination impairment that is typically induced with alcohol.[101] In fact, more broadly, oxytocin is connected to many interoceptive modalities including taste, touch, appetite, desire, nausea, olfaction, pain, sleepiness, and feelings of warmth.[102] OT's receptors are located in important centers of the autonomic nervous system and OT has the ability to interact with other neurotransmitters (dopamine, GABA, glutamate, acetylcholine, norepinephrine) to apparently modulate numerous systems. As a sense of your own body precludes working closely with others, it is not surprising that synchrony of interpersonal movements is also impacted by oxytocin; interestingly this effect is greater when individuals have higher scores in empathy,[103] making an interesting connection between motor coordination and social abilities.

While oxytocin seems to be a promising candidate for numerous systems related to neurodevelopmental disorders, interindividual variability should not be overlooked. In a study on human aggression, while OT did not have a main effect, its presence increased the positive correlation between certain antisocial personality traits and aggressive behavior.[104] Also, though out of the scope of this thesis, oxytocin could also be involved in the gastrointestinal symptoms

related to WS and Dup7, as it is critical for gut health. A closer look at oxytocin dysregulation in WS reveals both alluring leads, but also key gaps that experimental biology might address.

1.2.2 Oxytocin Dysregulation in WS

Oxytocin system dysregulation has been suggested in WS by several studies, with observations ranging from differences in the neuropeptide levels in blood and saliva to receptor expression and epigenetic modifications to the *OXT* gene promoter.[105–108] The earliest study to test the hypothesis that OT is altered in WS occurred in 2012. Dai et al. measured the blood plasma of 13 individuals with WS in the context of exposure to music and cold.[105] Utilizing an indwelling catheter, they were able to take samples throughout a 45 minute period and compare OT and vasopression (AVP) levels at baseline and after the sound and tactile stimuli were presented. Both neuropeptides were higher at baseline in WS relative to age-, sex-, and ethnicity-matched controls. Of the 10 timepoints where blood was collected, OT was higher at all of them, while AVP was only increased at 4 of the 10. In addition, OT responses to the stimuli were greater and more variable in the WS samples than controls, while AVP differences were not significantly different.

In 2015, studies emerged identifying dysregulation of the oxytocin receptor (OXTR) in the WSCR and a potential mechanism causing the change. Haas and Smith utilized data from a previously published dataset on WS transcriptome expression in skin fibroblasts from Henrichsen et al. (2011) to uncover differential expression of the *OXTR*.[107,109] *OXTR* expression was significantly increased in 8 WS compared to 9 controls, while the AVPR1A receptor for AVP just missed the significance threshold. With these data, they hypothesized an epigenetic changes underlying expression differences based on the association of WBSCR22 with methyltransferase

production. In the same year, a separate group published a paper detailing methylation differences enriched at CTCF binding sites across the genome in a dose-dependent manner in WS and Dup7.[110] The following year, Haas et al. published a paper assessing the relationship between DNA methylation of the *OXT* gene and human sociability, though not using WS samples. They show increased methylation in the *OXT* promoter region from saliva samples taken from "healthy" participants was associated with an anxious attachment style, while lower methylation (and presumably greater *OXT* expression) was correlated with better facial processing and more secure attachment styles.[108]

Oxytocin was also measured in a "healthy" population in response to an empathy-inducing video to determine whether its reactivity was associated with a variant in the WS gene *GTF2I*, which had been previously identified as a SNP of interest in social behavior.[106,111–113] Participants had saliva collected before and after watching a video validated to induce an empathy reaction. OT in the saliva was measured via an enzyme-linked immunosorbent assay specific for oxytocin (not AVP). Percent change of OT levels varied between time points (before and after the video) and between individuals with different alleles of SNP rs13227433 found in *GTF2I*, suggesting a link between the Gtf2i SNP and OT reactivity. Whether this finding is relevant in the context of the entire WS deletion is not yet determined.

Thus, overall, while oxytocin is a promising candidate for the WS locus pathology, only correlational studies with humans have been done thus far. In addition, OT levels are extremely variable making them hard to accurately measure, and blood and saliva OT levels may not be representative of the central OT levels that would be impacting neurological phenotypes. Working with model organisms, such as mice, can allow a modicum of control to better assess the impact of OT, especially regarding pharmacological manipulation of the OT system. To fill the void of

direct testing of the OT system in a mouse model relevant to WS, I chose to assess the effect of an oxytocin antagonist on fear learning and memory in a mouse model of the WSCR deletion (Chapter 2).[105,107,114–116] I focused on fear conditioning as oxytocin has been shown to both heighten adaptive fear responses and dampen anxiety behaviors, an interesting combination of effects. In addition, the previously demonstrated increase in social approach was not replicated in our hands at the time.[117] The limitations of previous studies highlight the need for deeper characterization of WS models across more domains, a problem I address in Chapter 3. Below, I review current models relevant to the WSCR, with a particular focus on larger deletions and single gene deletions of the *GTF2I* family, which has been implicated in the cognitive and behavioral profiles of WS to better illustrate the available tools and elements missing from current practice.

1.3 WSCR Animal Models

1.3.1 Benefits and Limitations

Animal models are invaluable to uncovering mechanisms underlying behavioral features of complex neurological disorders, especially those like WS and Dup7 that are rare in the human population. In addition to increasing sample sizes, there are numerous tools for genetic, behavioral, and pharmacological manipulations in mice that are not possible or ethical to employ with human patients. The WSCR is syntenic and largely conserved in mice, situated in reverse orientation on chromosome 5 and lacking the low-copy repeats flanking the human WSCR.[118] Without those repeating domains, the region in mice is easier to characterize and less susceptible to the large CNVs found in the human genome, making the mouse a useful model to reliably characterize this region. Indeed, numerous lines have been developed ranging from single gene deletions to deletions of large parts of the WSCR.[119]

While these mouse lines and other tools are very useful, cognitive and behavioral features are not modeled in a straightforward manner. For example, you cannot model anxiety or autism in a mouse, as these are purely human conditions. However, careful design of tasks can approximate anxiety-like or autistic-like behaviors that may provide critical information about the mechanisms fundamental to these different, but ideally related behaviors.[120] As a single task may not approximate all aspects relevant to human behaviors, a broad survey of related measures may be necessary to provide the level of depth required to explain complex phenotypes.[121–123] Statistical interpretation is also a critical factor in interpreting behavioral results. Prior to my work for this thesis, I surveyed literature in the field of autism and noted a common error in social approach analysis; different groups were being compared based on statistical calculations that are only applicable for within-group analysis. This erroneous comparison is explained and the appropriate statistical workflow for such behavioral paradigms is outlined in my paper, which I present in the Appendix.[124]

Once proper behavioral and statistical paradigms have been determined, the right animal model can reveal the conserved circuits allowing the study of human disease or complex disorders without the restrictions of a human model. Here I review a few of the most relevant models for the WSCR and highlight a single gene, *GTF2IRD1*, which will be a focus of my work in Chapter 3.

1.3.2 WSCR Mouse Models

Complete Deletion

The mouse line with the greatest construct validity for WS is the Complete Deletion (CD) model, also referred to as Del(5Gtf2i-Fkbp6)1Vcam in the Mouse Genome Informatics (MGI) database. In this line, the most common deletion seen in individuals with WS was recreated

through Cre-mediated recombination between *Gtf2i* and *Fkpb6* on mouse chromosome 5, where the syntenic WSCR lies.[125] Homozygous deletion of the region is lethal in mice, but the more relevant hemizygous CD mice are viable, despite slightly lower survival rates, and recapitulate many of the features of WS.

The cardiac features of WS are reflected in an increased arterial pressure and cardiac hypertrophy in the CD mice. A decreased mandible size is representative of the craniofacial differences. CD mice also exhibit a decrease in brain weight, cell density in the basolateral amygdala, dendritic length, CA3 hippocampal volume, and YFP+ neurons in motor, somatosensory, and hippocampal CA1 regions of the brain, while an increase in immature neural density was observed in the dentate gyrus. While no difference in gait or activity in an open field task were observed, CD mice did have a decrease in motility tonus strength and latency to fall on a Rotarod task, revealing possible balance or sensorimotor integration issues, which have also been investigated in WS.[43] Social approach behavior was also assessed in this model, though the increase observed was based only on quantifying male behavior with a single, non-standard task.

The CD model has a high degree of construct validity, by closely mirroring the genetic cause of WS, making it an ideal candidate to assess possible treatments.[126] *Gtf2i* intracisternal therapy on the model showed a decrease in the typically high social approach behaviors.[127] Another study used the model to address the short term memory and cardiac hypertrophy with a green tea extract.[128] While this model may be the most relevant for the human condition, it wasn't developed until 2014, and a deep characterization of its various behavioral phenotypes has not been provided, a gap I fill in Chapter 3.

Prior to development of this CD model, only partial deletions or single gene deletions had been researched. These models, while lacking the ideal construct validity of the CD model, offer

valuable insight the impact of specific genes and which regions in the locus may mediate the key WS phenotypes, such as the distinct cognitive and behavioral profiles.

Distal/Proximal Deletions

The model that provided additional support for which segment of the WSCR had the greatest impact on the WS phenotype split the region into two larger deletions, a proximal deletion (PD; closer to the centromeric breakpoint and representative of the telomeric end of the human WSCR) that lacked the genes between *Gtf2i* and *Limk1*, and a distal deletion (DD), which encompassed the telomeric end, which is related to the centromeric region in humans.[129] Using these two deletion models in combination and alone, it was shown that motor phenotypes were worse when the entire region was deleted, while sociability and startle response were more associated with the proximal end. This provided support that *Gtf2i* was likely involved with the social aspects of WS, and numerous single gene models were made to investigate the claim and the roles of the associated *Gtf2ird1*.

Monogenic Deletions

At least 11 of the 26 genes commonly deleted at the WS locus have knockout mouse models developed.[119] While these models have been imperative in isolating the functions of the individual genes, which provide insight to their potential impact on features of WS and Dup7, reviewing them all is beyond the scope of my work, thus I will narrow my focus to the *GTF2I* family, especially *GTF2IRD1*, as I explain below.

1.4 GTF2IRD1

GTF2IRD1 along with its close family member *GTF2I* have been implicated in key WS phenotypes based on atypical deletion case studies, as noted previously. These genes are likely involved in the behavioral and cognitive profiles in WS and warrant future investigation.

1.4.1 The *GTF2I* Family

The TFII-I protein family includes *GTF2I*, *GTF2IRD1*, and *GTF2IRD2*. These proteins contain I-repeats, which are helix-loop-helix-like domains not seen in other helix-loop-helix proteins. Each gene is alternatively spliced producing multiple isoforms. While the GTF2I protein, TFII-I, has been extensively researched, biochemical information on the GTF2IRD1 protein (also known as BEN, MusTRD1, or CREAM-1) is less advanced.[130] In addition, while mouse lines modeling the duplication and deletion of *Gtf2i* have been developed, only deletions of the related *Gtf2ird1* have been attempted thus far, and were met with varied success, thus I will focus on *Gtf2ird1* from here.[64,117,131–139]

1.4.2 *GTF2IRD1* and WS

As noted above, *GTF2IRD1* encodes a general transcription factor that is ubiquitously expressed. It was initially described and characterized in 1999.[140] Two transcripts produced by alternative splicing were observed, though we know now these represent two groups of many more alternatively spliced transcripts. They hypothesized its function to be similar to that of GTF2I, as a transcription factor and putative negative regulator, which were later confirmed.[139] GTF2IRD1 actually engages in negative autoregulation.[132] In 2002, the mouse ortholog for *GTF2IRD1* was

characterized, revealing significant conservation between the species; the intronic regions were more compact in the mouse *Gtf2ird1* but the coding sequence was largely unaltered.[141]

GTF2IRD1 has been connected to various WS phenotypes, often in conjunction with the highly related *GTF2I* due to their close proximity in the WSCR. Many of these associations have been ascertained by utilizing atypical human deletions, as discussed previously, which do not provide proof of causation. The difficulty in separating *GTF2I* from *GTF2IRD1* and the acknowledged limitations of correlational findings revealed a need for independent analysis of these genes in animal models to fully understand the relationship between *GTF2IRD1* and the WS profile.

1.4.3 *GTF2IRD1* mouse models

Some of the various phenotypes connected to *GTF2IRD1* include visuospatial processing,[142] craniofacial development,[64] fear, aggression, and social behaviors.[131] In 2005, *GTF2IRD1* was implicated in the craniofacial features and growth deficits seen in WS by utilizing heterozygous and homozygous *Gtf2ird1*-null animals created by a transgenic insertion that induced a deletion of the region, including the first exon of *Gtf2ird1*.[64] This same *Gtf2ird1*-null mutant also revealed anxious and hypoactive phenotypes.[137] A different *Gtf2ird1* mutant created via replacement targeting did not find any anxiety-related phenotypes, but reported increased sociability and decreased aggression and fear behaviors, in addition to body weight.[131] The differences in these various models may be a result of differences in creating the models or the background strain on which they were developed. In some former *Gtf2ird1* models, downstream initiation produces a viable product, which was often missed due to poor antibody availability to verify protein levels in many of these earlier models.[139]

The unreliability of the earlier deletion models of *Gtf2ird1*, combined with its relevance to many WS phenotypes, warrants continued investigation of the gene. To avoid the issues with deleting *Gtf2ird1*, I use a novel transgenic line in Chapter 3 to tease apart its function by rescuing its expression on a WS-relevant background –the CD mice described previously– allowing me to isolate its function within the larger deletion. The CD line provides important context for single gene experiments. Allelic interactions (seen in motor phenotype) seem to be common in the region and any results from a single gene in the region may not be valid when combined with the entire deletion. This new *Gtf2ird1* transgenic line can also be used to assess the effects of *Gtf2ird1* overexpression alone, which may connect to features observed in Dup7.

1.5 Conclusion

The WSCR is a valuable genetic background that can be used to investigate the genetic origins of the myriad symptoms associated with CNVs at the locus. When applying our understanding to neurodevelopmental disorders including WS and Dup7, however, context is imperative. Single gene models, usually modeling a deletion, can provide some insight but especially when trying to elucidate mechanisms underlying complex behaviors like social approach, these models may be insufficient. A more appropriate model to use would have better construct validity, such as the Complete Deletion mouse that models the most common WS deletion. In Chapter 2, I attempted to identify OT's effects on conditioned fear by administering an oxytocin antagonist directly to the central nervous system, using the CD model. Characterization of the phenotypes produced by such a model is necessary to best utilize the resource. Deep phenotyping of motor, social, anxiety, and fear-relevant features, as I did in Chapter

3, can provide important insights; for example, it could help determine which behavioral tasks may be most sensitive or informative in future pharmacological studies.

Chapter 2: Oxytocin receptor activation does not mediate associative fear deficits in a Williams Syndrome model

Kayla R. Nygaard, Raylynn G. Swift, Rebecca M. Glick, Rachael E. Wagner, Susan E. Maloney, Georgianna G. Gould, and Joseph D. Dougherty

From:

Oxytocin receptor activation does not mediate associative fear deficits in a Williams Syndrome model.

Nygaard KR, Swift RG, Glick RM, et al. Oxytocin receptor activation does not mediate associative fear deficits in a Williams Syndrome model. *Genes, Brain and Behavior*. 2022; 21(1):e12750. doi:10.1111/gbb.12750

2.1 Abstract

Williams Syndrome results in distinct behavioral phenotypes, which include learning deficits, anxiety, increased phobias and hypersociability. While the underlying mechanisms driving this subset of phenotypes is unknown, oxytocin (OT) dysregulation is hypothesized to be involved as some studies have shown elevated blood OT and altered OT receptor expression in patients. A "Complete Deletion" (CD) mouse, modeling the hemizygous deletion in Williams Syndrome, recapitulates many of the phenotypes present in humans. These CD mice also exhibit impaired fear responses in the conditioned fear task. Here, we address whether OT dysregulation is responsible for this impaired associative fear memory response. We show direct delivery of an OT receptor antagonist to the central nervous system did not rescue the attenuated contextual or cued fear memory responses in CD mice. Thus, increased OT signaling is not acutely responsible for this phenotype. We also evaluated OT receptor and serotonin transporter availability in regions related to fear learning, memory and sociability using autoradiography in wild type and CD mice. While no differences withstood correction, we identified regions that may warrant further

investigation. There was a nonsignificant decrease in OT receptor expression in the lateral septal nucleus and nonsignificant lowered serotonin transporter availability in the striatum and orbitofrontal cortex. Together, these data suggest the fear conditioning anomalies in the Williams Syndrome mouse model are independent of any alterations in the oxytocinergic system caused by deletion of the Williams locus.

2.2 Introduction

Williams Syndrome (WS), a multisystemic neurodevelopmental disorder, is caused by a 1.5–1.8 Mbp hemizygous deletion on chromosome 7q11.23, altering the copy number of 26–28 contiguous genes in the WS critical region (WSCR). The complex phenotypic characteristics of WS include craniofacial dysmorphology, connective tissue abnormalities and cardiac problems such as supravalvular aortic stenosis and peripheral artery stenosis. In addition, WS is characterized by distinct cognitive features, including intellectual disability, profoundly impaired visuospatial construction,[143] atypical facial processing, deficits in motor coordination and control, odynacusis[35] and impaired auditory processing.[17,24,36,40,42,144–148] Interestingly, however, individuals with WS possess relatively intact expressive language and verbal skills,[12,149] as well as heightened sensitivity and emotional response to music.[35,150] One of the most striking phenotypes of individuals with WS is hypersociability and strong social motivation,[16,151,152] despite high non-social anxiety[32] and deficits in social cognition and awareness.[153]

A substantial body of research indicates the neuropeptide oxytocin (OT) plays a key role in mediating the regulation of social behavior and cognition, fear conditioning and extinction, observational fear,[154] fear modulation via social memory[155] and anxiety in humans and rodents.[156–158] Given the aberrant social behavior and anxiety in individuals with WS, recent studies have

tested the hypothesis that OT is dysregulated in WS. Indeed, one study found elevated blood levels of OT in individuals with WS compared with controls.[105] However, the findings on the OT receptor (OXTR) have been contradictory. One study suggested increased gene expression,[107] while another demonstrated downregulation and hypermethylation of *OXTR* in WS.[114]

While we did not see altered social behavior in a recent application of the standard social approach task, we did see differences in freezing during a conditioned fear task.[117] Another mouse model, which deletes the entire WS-homologous region, has also shown alterations in fear conditioning,[159] and individuals with WS have heightened phobias and non-social anxieties. Alterations in the brain OT system play an important role in social fear conditioning, contextual fear-induced freezing and social fear extinction.[160,161] Additionally, peripheral administration of an OT receptor agonist has been shown to inhibit fear-induced freezing,[162] and evoked OT release via channelrhodopsins also results in attenuation of fear.[97]

In this study, we investigated whether OT dysregulation is a mechanism underlying the fear conditioning phenotype following deletion of the WSCR using the mouse experimental system. We employed the model reflecting the most common deletion found in WS patients: the hemizygous loss of the entire genomic region between the *Gtf2i* and *Fkbp6* genes.[159] These heterozygous Complete Deletion (CD) mice show reduced freezing in fear conditioning recall, which is consistent with the expected consequences of OT elevation. Therefore, we probed whether OT activity could be responsible for the decreased expression of associative fear memory in CD mice. Further, to complement the prior human studies of OXTR expression in peripheral cells,[107,114] we tested whether OXTR expression differs in CD versus wild type (WT) mice across the brain, but we found no differences after statistical correction in this system, nor in a second neurotransmitter system (serotonin, 5HT), which had previously been shown to cooperate with OT

in social learning,[158] and can be influenced by OT.[163] Together, these data suggest there is not a direct role for the OT system in associative fear learning in WS.

2.3 Results

2.3.1 Complete deletion mice show impaired contextual and cued fear conditioning

OT has been shown to modulate the freezing response of rodents in conditioned fear tasks. Specifically, central administration of OT results in decreased levels of freezing in response to avoidance-associated cue or context.[164] Thus, because of the suggested increase in OT production in WS, we sought to determine whether the CD mouse model also had altered associative fear learning. Previously a decrease in overall freezing time during fear conditioning was shown in the CD model using only male mice;[159] here we replicate and expand on these findings by confirming the phenotype in both sexes.

We found that CD mice responded to shock during conditioning (Day 1) by increasing freezing to the same extent as WT mice (main effect of minute, $F(2, 36) = 77.81$, $p = 8.5 \times 10^{-14}$). There was no main or interaction effect of genotype observed (**Figure 1B**, complete statistical analysis available in **Table S1**). WT and CD mice both exhibit significantly higher freezing ($F(3, 36) = 9.6$, $p = 8.5 \times 10{-}5$) in the first 2 min of the contextual memory test (context) compared with the first 2 min of training (baseline), indicating each group successfully associated the fear stimuli with the context (**Figure 1C**).

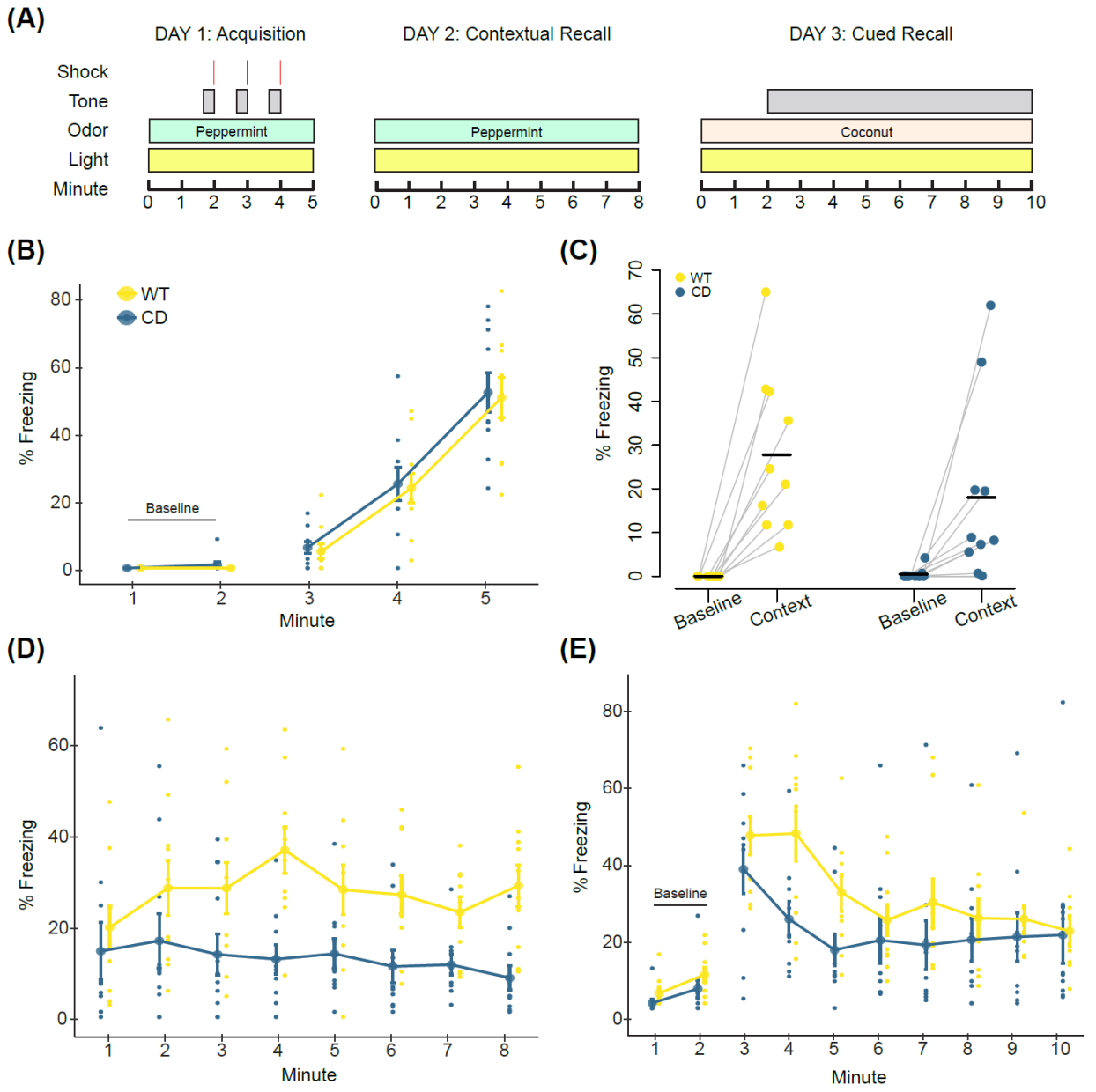

Figure 1. Complete Deletion mice have altered associated fear responses in a conditioned fear task. A) Overview of the conditioned fear task protocol. **B)** Day 1. CD and WT mice show increased freezing with subsequent footshock deliveries. **C)** All mice increased freezing in the context associated with the footshock. Baseline is the average of the first two minutes of Day 1. Context is the average of the first two minutes of Day 2. Black bars indicate the mean average percent freezing. Data points of individual mice are connected. **D)** Day 2. All mice increased freezing to context relative to Day 1 baseline, though CD mice freeze less than WT mice. **E)** Day 3. CD mice have significantly decreased freezing relative to WT mice during minute 4 of tone delivery ($p=0.036$). WT: n=10; CD: n=10. Connected data points in **B**, **D**, and **E** are means ± SEM. Individual scores are represented by colored circles in the background.

On Day 2, CD mice froze less compared with WT mice when placed in the same context (chamber/odor) used for fear conditioning (**Figure 1D**), main effect of genotype, $F(1,$

18) = 7.95, p = 0.011), indicating impaired contextual associative fear memory. The response to the conditioned cue was also decreased in CD mice but not as broadly, with a main effect of minute (F[9162] = 22.0195, p = 2 × 10−16), and a borderline minute by genotype interaction (**Figure 1E**, F(9,162) = 1.9151, p = 0.053), but no main effect of genotype. Post hoc analysis revealed the mean freezing of CD mice was almost 24% less than WT freezing at minute 4 (p = 0.036). Overall, CD mice show a deficit in contextual fear conditioning that is consistent with elevated OT activity.

2.3.2 CD conditioning deficits are not reversed by central infusion of an oxytocin receptor antagonist

We next sought to determine if CD mice had elevated OT levels in the blood, as had been reported in patients.[105] We used the same ELISA approach, but did not see a significant difference between genotypes (t = 0.003, df = 21.664, p = 0.9976; CD (n = 13): M = 1138.191 pg/ml, SD = 503.682; WT (n = 11) M = 1138.791 pg/ml, SD = 478.916). However, even WT mice had a remarkable range of OT levels in blood (82–1731 pg/ml), likely driven by the periodic nature of OT release, coupled with its short half-life (<5 min).[165] Thus, it can be hard, at a single point in time, to detect and make conclusions about average OT levels. Furthermore, blood OT levels may not reflect OT levels in the brain. Therefore, we took an experimental approach; if elevated central OT was responsible for decreased associative fear learning in CD mice, then blocking central OXTR activity should reverse the phenotype. We implanted CD and WT mice with ICV cannulas to directly administer an OXTR peptide antagonist (OTA) and block all receptor activity during the fear conditioning procedure (**Figure 2A**).

There were no baseline freezing differences between CD and WT animals, as measured in minutes 1 and 2 of the first day of testing (F(1, 39)=1.58, p = 0.22). During the conditioning phase, we found an interaction of genotype and treatment. The OTA had an opposite effect on freezing

in CD and WT animals. Specifically, WT mice receiving the OTA froze more than their CD counterparts (**Figure 2B**, $F(1, 39) = 7.32$, $p = 0.0101$).

During contextual fear recall, both CD and WT animals showed evidence of learning, as freezing increased in all groups from Day 1 baseline compared with the first 2 min of Day 2. Overall, a main effect of genotype on contextual fear memory (**Figure 2C**, $F(1, 39) = 19.96$, $p < 6.6 \times 10{-}5$) reflects a significant reduction in freezing within CD mice compared with WT mice, regardless of treatment. While this replicated results from our first experiment, there was no main effect of treatment or a genotype by treatment interaction, thus administration of the OTA did not significantly alter contextual fear responses. This was also true on Day 3 for cued fear responses (**Figure 2D**), where there was only a main of effect of time ($F[2273] = 26.45$, $p = 2.2 \times 10{-}16$) and an interaction between time and genotype ($F[2273] = 5.88$, $p = 2.26 \times 10{-}6$) driven by the decreased freezing of CD mice compared with WT mice at minute 4 ($p = 0.0083$). Thus, we show OT signaling at the OXTR does not account for the impaired associative fear response in this model.

2.3.3 Autoradiography reveals no changes in oxytocin receptor density or distribution in CD mice

In parallel to the experimental approach, we investigated OXTR availability in the mouse brain through a discovery-based approach, as elevated levels in the amygdala might influence fear conditioning.[165] Particularly, given opposite directions of effect of the OTA in WT and CD mice on Day 1 freezing (**Figure 2B**), we suspected genotype differences in OXTR expression in regions related to fear learning. Furthermore, it is of interest to study this binding given the findings of OXTR dysregulation in brains of humans with the WSCR deletion.[107,114]

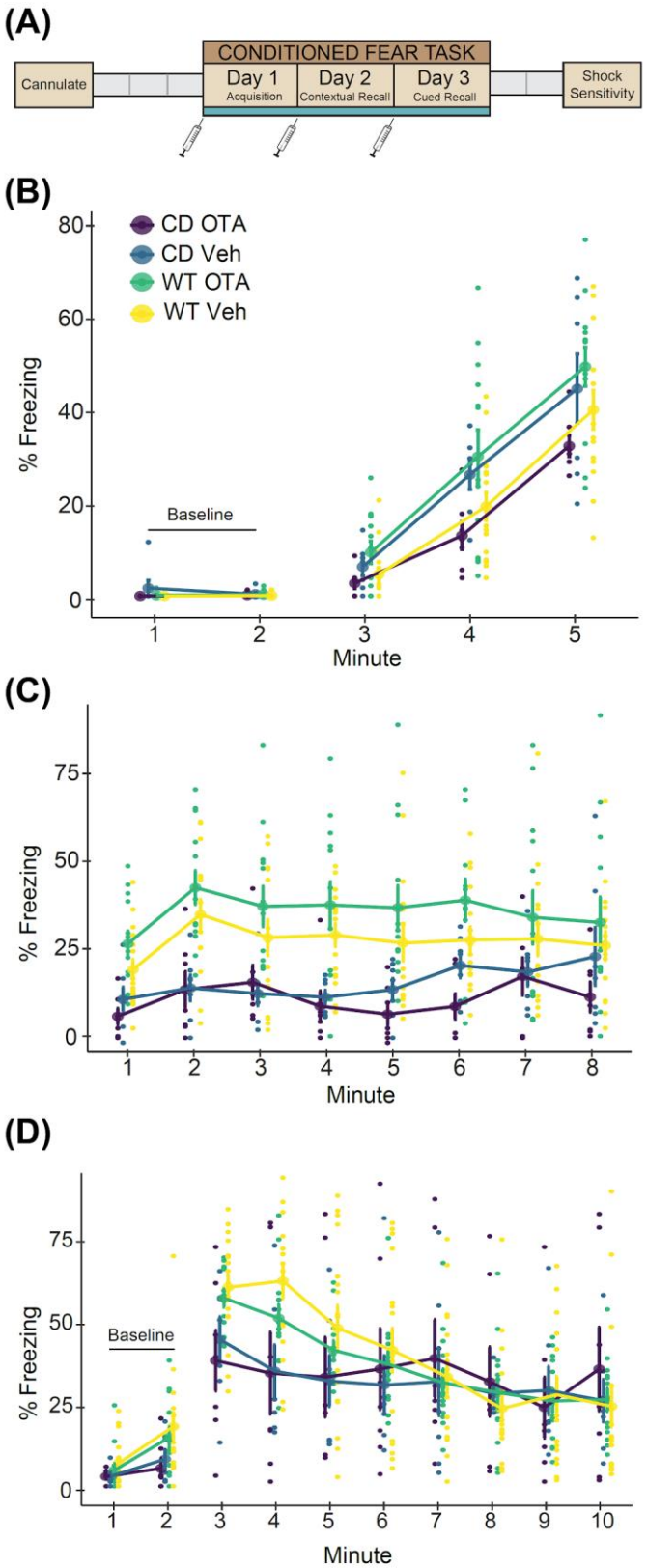

Figure 2. Central administration of an oxytocin receptor antagonist does not rescue reduced contextual or cued fear responses in CD mice. A) Schematic of the experiment. Grey boxes show rest days. Syringes indicate ICV infusions of OTA or

vehicle, which occurred at least 1 hour prior to testing. **B)** Day 1 of conditioned fear. CD and WT mice show increased freezing with subsequent footshock deliveries. WT OTA-treated mice freeze significantly more than CD OTA-treated mice. **C)** Day 2. All mice show increased freezing to context (minutes 1 and 2) relative to Day 1 baseline. CD mice freeze less than WT mice but there is no main or interaction effect of treatment. **D)** Day 3. CD mice have significantly decreased freezing relative to WT during minute 4 of tone delivery (p=0.008), but there is no effect of treatment. WT Veh: n=16; WT OTA: n=13; CD Veh: n=7; CD OTA: n=7. Connected data points are means ± SEM. Individual scores are represented by smaller unconnected circles. Veh = vehicle; OTA = oxytocin receptor antagonist.

Therefore, we conducted an autoradiography study using radiolabeled OVT ligand on coronal sections of WT and CD brains (**Figure 3A**). As well as measuring regions relevant to fear conditioning (BLA and LA[166,167]; LSN[160]), we measured areas of where OT has been shown to affect sociability or memory (AON[168]; LSN[169]; ACC[170,171]; CA2/3[172,173]; Pir[174]; PVN[175]; and CPu[158]; **Figure 3B**), as hypersociability and cognitive impairments are characteristic of WS. We found no significant differences in OXTR binding between genotypes within regions of interest in CD and WT brains when corrected for multiple testing (**Figure 3C**, **Table S2**). There was a nominally significant change in the LSN ($p = 0.034$), but it did not meet the corrected experimentwise critical alpha level ($\alpha = 0.006$). Given the role of the LSN in fear and anxiety and future focused studies of this region may be warranted.

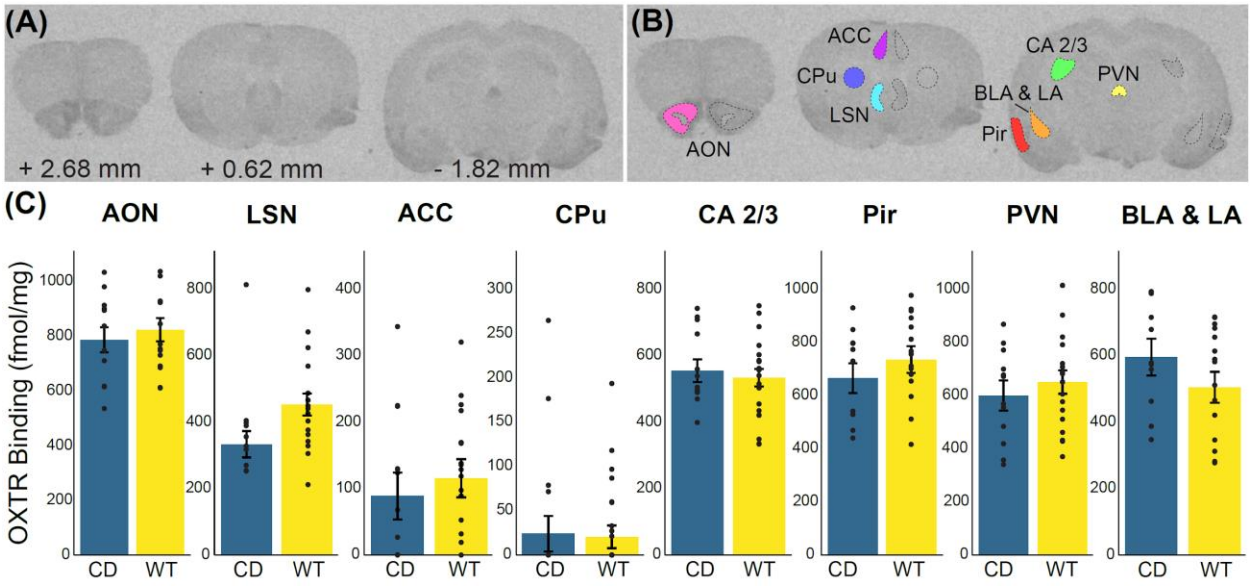

Figure 3. No significant differences in oxytocin receptor density in Complete Deletion mice. A) Coronal sections from a representative mouse brain used in iodinated ornithine vasotocin analog ([^{125}I]OVT) autoradiography analysis with corresponding distance from bregma. **B)** Example tracing of regions of interest including anterior olfactory nucleus (AON), cingulate cortical

areas 1 and 2 (ACC), striatum (CPu), lateral septal nucleus (LSN), hippocampal CA 2 and 3 regions, paraventricular nucleus (PVN), basolateral (BLA) and lateral (LA) amygdala, and piriform cortex (Pir). **C)** Results of oxytocin receptor autoradiography comparing [^{125}I]OVT binding (fmol/mg of protein) in regions of interest between CD mice and WT mice. Colored bars show means ± SEM (brackets). Individual averaged measurements for each mouse are represented by circles. For each genotype, $n \geq 9$.

2.3.4 Autoradiography reveals no changes in serotonin transporter density or distribution in CD mice

Finally, with no changes in OXTR availability, we examined an additional alternative neurotransmitter: serotonin (5HT). Disruption to the 5HT system in WS has been suggested in prior studies. Specifically, Proulx et al. (2010) determined that there are enhanced 5-HT1A receptor-mediated currents in a WS mouse model with low innate anxiety.[176] More recently, Lew et al. (2020) compared serotonergic innervation in the amygdala between autism and WS in human postmortem samples, concluding that there is decreased innervation in WS brains compared with neurotypical brains.[177] Therefore, we focused on the SERT, as it should provide a measure of serotonergic innervation to different structures.

We measured SERT binding in several brain regions (**Figure 4A**). We focused on amygdalar regions relevant to the human postmortem studies[177] and fear conditioning phenotypes,[178] assessing BLA, central amygdala (CeA) and LA regions independently based on findings that they differ in the amount of serotonergic innervation in some species.[179] We then included other areas where SERT has been implicated in behaviors relevant to the WS phenotype, patient findings and knowledge of 5HT biology. These included the nucleus accumbens (NAc) based on 5HT's role on social reward in this region,[158] the BNST for its role in adaptive anxiety,[180] and additional hypothalamic and cortical regions of interest, including the ACC, lateral parietal association (PtA), orbitofrontal cortex (OFC) and the peduncular part of the lateral hypothalamus (PLH) which is where the medial forebrain bundle is found, representing a major ascending pathway for nearly all 5HT axons (**Figure 4B**).[181]

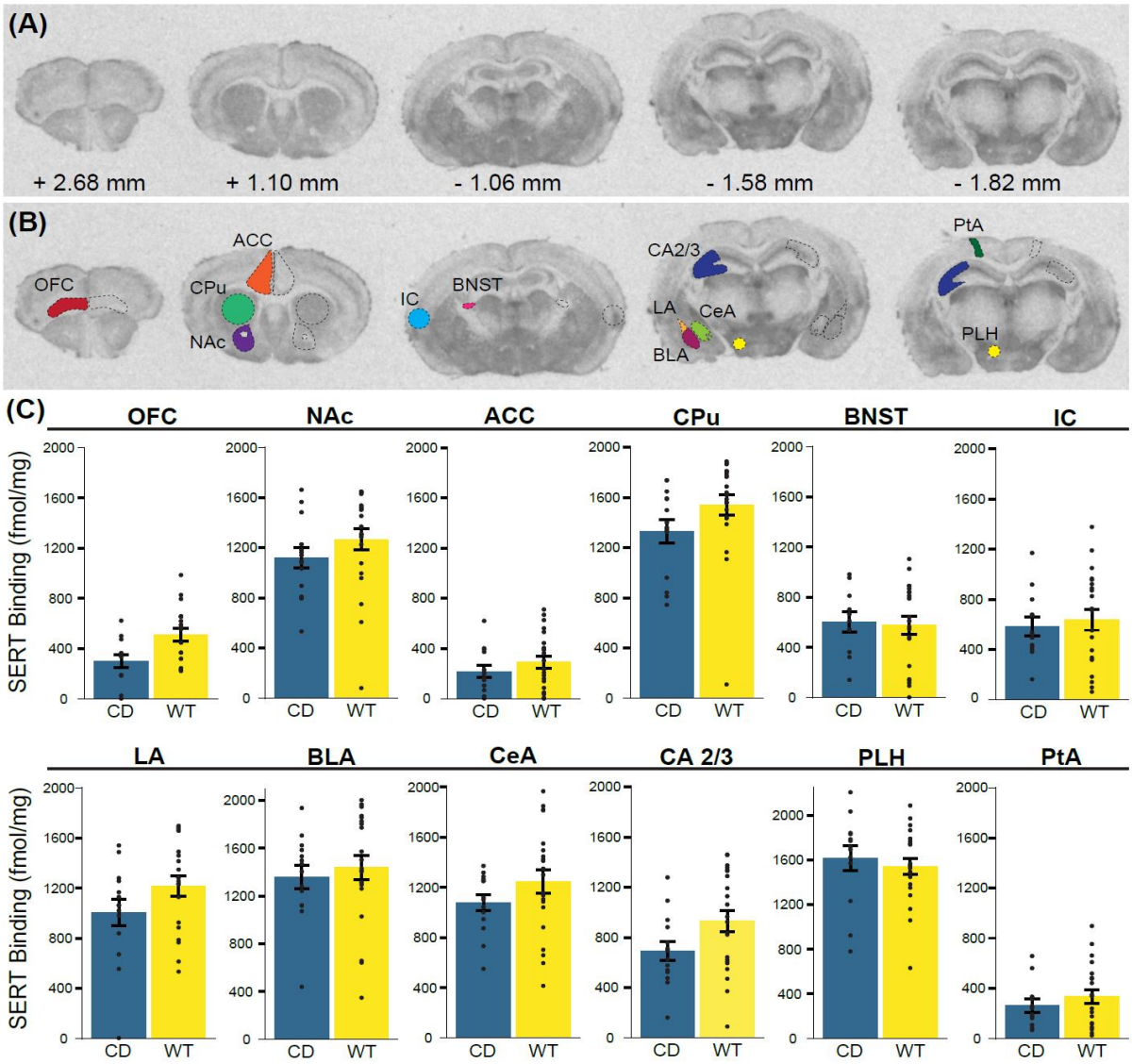

Figure 4. No significant differences in serotonin transporter density in Complete Deletion mice. A) Coronal sections with corresponding distance from bregma from a representative mouse brain used in [^{125}I]RTI-55 autoradiography analysis. **B)** Example tracing of regions of interest including orbitofrontal cortex (OFC), anterior cingulate cortex (ACC), striatum (CPu), nucleus accumbens (NAc), the bed nucleus of the stria terminalis (BNST), insular cortex (IC), hippocampal CA 2 and 3 (CA 2/3) regions, basolateral (BLA), lateral (LA), and central (CeA) amygdala, lateral parietal association (PtA), and the peduncular part of the lateral hypothalamus (PLH). **C)** Results of serotonin transporter (SERT) autoradiography comparing [^{125}I]RTI-55 binding (fmol/mg protein) in regions of interest between CD mice and WT mice. Colored bars show means ± SEM (brackets). Individual measurements for each mouse (regional average) are represented by circles. For each genotype, n ≥ 5.

Overall, there was a nonsignificant cross-region effect of genotype ($F(12) = 2.41$; $p = 0.07$), with no effect of sex ($F(12) = 0.49$; $p = 0.89$) (**Figure 4C**). Nonsignificant decreases in SERT density in CD mice compared with WT were found in the CPu

($p = 0.067$) and OFC ($p = 0.060$), and these may be potential regions of interest for future, focused studies better powered to study smaller effects.

2.4 Discussion

We and others have previously observed altered fear conditioning in WS models.[117,159] Expanding on previous studies that used only male mice, we found, using both sexes, that CD animals on a C57BL/6J background had a suppressed fear response to the context and cue presented in the fear conditioning paradigm. Thus, CD mice enable investigation of the underlying circuit disruptions mediating this phenotype. We hypothesized that the altered associative fear memory response in CD mice was because of the increased availability of OT based on human findings of elevated OT in WS.[105] We focused our efforts on functional studies, which would be definitive with regards to a role for OT in fear conditioning of CD mice. Using an intraventricular cannula, we treated mice with an OT receptor antagonist during each day of conditioned fear to attempt to counteract any possible effect of increased OT production on the fear response. The OTA did not alter the subdued freezing response in CD mice. Therefore, the associative fear conditioning phenotype that results from loss of the WS critical region is not mediated by OXTR activity.

Our results tell us less about the role of OT in WT mice. While many prior studies show OT modulating fear conditioning,[97,162,164,167,182] Pisansky et al. (2017) found OT enhances fear in a social paradigm of observational fear learning, but does not affect non-social fear learning.[154] We did not see a significant difference in WT mouse behavior in response to OTA ($p < 0.083$, Day 2), but as there was a small difference in the means of WT mice given Vehicle compared with those given OTA, there may be an effect below our power to detect. Therefore, we are hesitant to weigh

in on this debate. In the end, the discrepancy across studies could be a result of dosage. Gunduz-Cinar et al. (2020) show different concentrations of OT have opposing effects on other fear-related tasks.[183] We have also observed an effect of genetic background on fear conditioning in the CD mice, with elevated freezing in CD mutants on an FVB/AntJ x C57BL/6J F1 hybrid background.[117] These hybrids show different levels of baseline freezing in response to cues even in WT animals, suggesting the impact of the CD deletion interacts with other genes in the genome to modulate conditioned fear effects. Thus, if OT is not involved, a genetic screen for interacting loci using mouse strain panels may help identify the relevant pathways. Another option would be to investigate genes based on dysregulated expression in models of WS, such as the serotonin 5HT1B receptor, which is among the top 10 dysregulated genes in a cell model of WS.[70]

In addition, given some of the prior work suggesting OXTR receptor gene expression and methylation in WS patient blood cells, we were curious if receptor availability was also altered in the brain following deletion of these genes. We used autoradiography because it can measure the availability of the receptor at the surface, which should reflect protein level and localization changes in addition to changes in gene expression. Further, it provides an opportunity for spatially informed analyses. As such, it is the best single measure for assessing if the OXTR is modulated in a conserved way by these mutations. We assessed regions previously associated with fear conditioning and those where OT had been shown to modulate social reward (CPu). Overall, we found no significant changes in OXTR binding in the CD mouse brain compared with controls, although there was a nominally significant difference in the LSN prior to correction, in the same direction of effect as seen in WS patient blood cells. The LSN is an interesting center integrating a variety of fear and anxiety signals, for example, playing a role in how stressful social cues are received.[184] In addition, the LSN has been associated with fear-enhancing effects of the OXTR,

but as modulation of the OXTR did not alter contextual conditioned fear responses, it is thought to be through indirect means.[185] These data motivate future studies focusing on the role of the LSN in behavioral abnormalities observed in the CD mouse model.

Despite ruling out a direct role for OT in fear conditioning deficits in this model, altered OT might play a role in the increased social motivation in this population. While it is of interest scientifically to assess this, hypersociality is not as much of a concern therapeutically as other phenotypes, such as learning deficits, ADHD, phobias and anxiety. Beyond OT and serotonin, dopamine has also been implicated in WS,[42] and has been previously connected to fear conditioning,[186] other anxiety-avoidance tasks,[187] and ADHD-related hyperactivity.[188] Thus, it is possible deletion of the WSCR disrupts dopamine signaling to result in these behavioral alterations. Understanding the roles of any of these systems in patient-related phenotypes of the CD mice might help highlight potential treatments in WS, as a wide range of therapeutics working on these systems are currently available.

2.5 Materials and Methods

2.5.1 Animals

CD mice contain a hemizygous deletion of the WSCR and were maintained on the C57BL/6J background (Jackson #000664).[159] Animals were bred by crossing CD heterozygotes to C57BL/6J WT animals to produce heterozygous CD experimental mice along with WT littermates for the control group. Tissue collection and genotyping PCR occurred in the second postnatal week. Mice were housed by sex and treatment, when relevant, and were kept on a 12:12 h light/dark schedule with food and water provided ad libitum. All studies were approved by and conducted in accordance with the Institutional Animal Care and Use Committee at Washington

University in St. Louis. All behavioral testing occurred during the light phase and was conducted by a female experimenter. Four independent cohorts were used in this study. Cohort 1 included 13 CD and 11 WT male mice from 8 independent litters and was used to assess blood OT levels via ELISA. Cohort 2 comprised 10 CD (4 females (F), 6 males (M)) and 10 WT (8 F, 2 M) mice from four independent litters and were behaviorally examined as adults (postnatal day [P] 97–106) with the conditioned fear task. Cohort 3 comprised 14 CD (8 F, 6 M) and 29 WT (12 F, 17 M) mice from 16 independent litters and served to evaluate the role of the OT system in associative fear and avoidance learning as adults (P68–118). Cohort 4, comprising 22 WT (12 F, 10 M) and 14 CD (5 F, 9 M) from 11 independent litters, was used to evaluate OXTR and serotonin transporter (SERT) expression in specific brain regions of interest. Tissue was collected post-mortem to confirm initial genotyping results.

2.5.2 Oxytocin ELISA

Blood was drawn from the retro orbital sinus of isoflurane-anesthetized mice at P30 using heparinized glass capillary tubes. Samples were collected in 1.8 ml EDTA-coated tubes, spun at 1600 g for 5 min at 4°C, then split into two aliquots, placed on dry ice and stored at −80°C until use. An ELISA kit was used for colorimetric quantification of OT per the manufacturer's protocol (ADI-900-153A, Enzo Life Sciences, Farmingdale, NY, USA). Prior to use, samples were diluted 1:3 with 200 µl of assay buffer. Absorbance measurements were read at 405 nm and OT concentration was calculated using a standard curve produced using the provided OT standards.

2.5.3 Conditioned fear task

Associative fear and avoidance learning were evaluated in the CD mice using the conditioned fear paradigm (**Figure 1A**), as described in our previous studies.[189] Briefly, each mouse was habituated to and tested in an acrylic apparatus, which measured 26 cm × 30 cm × 30.5 cm tall and contained a metal grid floor, an LED light bulb and an inaccessible peppermint odorant, housed within a sound-attenuating chamber (Actimetrics, Wilmette, IL, USA). The chamber light turned on at the start of each trial and remained illuminated for the duration. On Day 1, the testing session was 5 min. An 80 dB white noise tone sounded for 20 s each at 100 s, 160 s and 220 s. A 1.0 mA shock was paired with the last 2 s of the tone. The baseline freezing behavior (first 2 min) and freezing behavior during the last 2 min was quantified via the computerized image analysis software program FreezeFrame (Actimetrics). This measure allowed for simultaneous visualization of behavior while adjusting a "freezing threshold," which categorized behavior as freezing or not freezing during 0.75 sec intervals. Freezing was defined as no movement except for normal respiration and data were presented as percent of time spent freezing. Testing on Day 2 lasted 8 min, during which no tones or shocks were presented. This procedure enables evaluation of freezing behavior in response to contextual cues associated with the shock stimulus from Day 1. For the 10 min testing session on Day 3, the context was changed to an opaque Plexiglass-walled chamber containing a different (coconut) odorant. The 80 dB tone began at 120 s and lasted the remainder of the trial. Freezing during habituation to the new context was quantified across the first 2 min. Freezing behavior to the auditory cue associated with the shock stimulus from Day 1 was quantified for the remaining 8 min. Each day of testing, males were run first, followed by females. Assigned boxes were counterbalanced by genotype. Between runs, the apparatus was cleaned with 70% ethanol (Days 1 and 2) or 0.02% chlorhexidine diacetate solution (Day 3; Zoetis,

Parsippany-Troy Hills, NJ, USA). Animals were put in a holding cage until all cagemates had been tested, then animals were returned to their home cage. Shock sensitivity was evaluated after testing as previously described to verify differences in freezing were not the result of altered sensitivity to the shock stimulus itself.[190]

2.5.4 Intracerebroventricular infusion of oxytocin receptor antagonist during conditioned fear task

The surgical area of adult mice was shaved a day prior to insertion of the guide cannula to facilitate the intracerebroventricular (ICV) injections. Mice were anesthetized with 2.5–5% isoflurane and placed in a stereotaxic apparatus. Prior to the procedure, mice received a local anesthetic, 1 mg/kg of Buprenorphine SR (ZooPharm, Laramie, WY, USA), and an antibiotic, 2.5–5 mg/kg of Baytril (Bayer Healthcare LLC, Shawnee Mission, KS). An incision was made along the skull to visualize bregma to lambda. The periosteum was removed by lightly scratching the surface of the skull and the area was cleaned three times with a betadine solution (Purdue Products L.P., Stamford, CT, USA) on sterile cotton swabs followed by a quick hydrogen peroxide swab. The guide cannula was placed in a stereotaxic cannula holder (#51636–1, Stoelting, Wood Dale, IL, USA). Using a rapid, fluid motion, the 26-gauge unilateral guide cannula (C315GS-5/SPC, Plastics One, Roanoke, VA, USA) with dummy cap was inserted at the following coordinates: M/L = +1, A/P = −0.4, D/V = −2.2, based on prior work.[191–193] The guide cannula was cut to a length of 2 mm so that it entered the lateral ventricle. The dummy cap (C315DCS-5/SPC) and internal cannula (C315IS-5/SPC) were cut to protrude 0.2 mm from the end of the guide. C&B Metabond dental cement (Parkell, Edgewood, NY, USA) was mixed on a chilled ceramic dish and used to secure the cannula to the skull and seal the surgical area. The dental cement dried

completely before the animal was transferred to a recovery cage. Animals were housed together after fully awake and provided 0.25 mg of the chewable anti-inflammatory Rimadyl (Bio-Serv, Flemington, NJ, USA). During daily monitoring, dummy caps were replaced and tightened as needed. Mice were euthanized at the first sign of distress or damage to the surgical area and had at least 3 days for recovery prior to testing.

All mice received 1-μl infusions at least 1 h before each day of the conditioned fear task. Each mouse was given either vehicle (artificial cerebrospinal fluid solution, Tocris Bioscience, Bristol, UK; WT n = 16, CD n = 7) or an OT receptor antagonist (OTA) (desGly-NH_2,d(CH_2)$_5$ [Tyr(Me)^{2}Thr^4]OVT, Bachem, Torrance, CA, USA; WT n = 13, CD n = 7). The OTA, dissolved in vehicle at 1 ng/μl, is a peptidergic ornithine vasotocin analog chosen because of its broad applicability, long half-life and prior use in ICV injections.[194–196] The solutions were unilaterally delivered into the lateral ventricles through the 33-gage internal cannula via a PlasticsOne Cannula Connector (C313CS) over the course of 1 min using a Quintessential Stereotaxic Injector (Stoelting #53311) and a 1 μl Hamilton syringe. After injection, 15–30 s passed before removing the internal cannula to ensure proper diffusion. The conditioned fear task was performed as described above. Following completion of behavioral testing, cannula placement was confirmed by injecting enough dye to flood the ventricles and immediately euthanizing the animal via isoflurane overdose. Brains were extracted and sliced coronally at the injection site with a razor blade. Infusion of the dye into the ventricles was then confirmed by eye and samples that missed the ventricles were excluded from the final analysis.

2.5.5 Quantitative autoradiography of OXTR and SERT in mouse brain

Naïve mice were rapidly euthanized by cervical dislocation. Brains were removed, placed in an ice-cold saline solution for 1 min, then excess saline solution was wicked onto a paper towel. Brains were frozen on crushed dry ice and then stored at −80°C until sliced into 20 µm coronal sections in a cryostat (Leica Biosystems 1850, Buffalo Grove, IL, USA) at −16 to −18°C. Slides were thaw-mounted onto gelatin-coated microscope slides, vacuum-desiccated overnight (18 h) at 4°C, then stored at −80°C until use. Adjacent sections were used for two distinct ligands, [^{125}I]OVT and [^{125}I]RTI-55, to assess OXTR and SERT availability, respectively.

OXTR quantitative autoradiography

Binding of iodinated ornithine vasotocin analog ([^{125}I]OVT) to OXTR in the mouse brain was performed as described previously,[197] with minor modifications. Mounted sections were thawed for 30 min at 22–23 °C, then pre-incubated for 30 min in 50 mM Tris HCl buffer pH 7.4 at 22–23 °C. Next, sections on slides were incubated for 90 min in upright cytomailers filled with 10 ml buffer containing 10 mM $MgCl_2$, 0.1% bovine serum albumin and 50 pM [^{125}I]OVT (NEX2540, PerkinElmer, Boston, MA, USA). Non-specific binding was obtained by incubating representative adjacent sections on slides from the series in buffer containing unlabeled OT (1 µM, Ascent Scientific, Bristol, UK). Sections on slides were then washed twice for 5 min each in glass staining dishes containing 300 ml of 4°C buffer and were dipped for 2 s in 4°C deionized water. Slides were dried on a benchtop slide warmer for 1 h or until sections were opaque and any droplets had evaporated.

SERT quantitative autoradiography

Slides were defrosted at 22–23°C for 30 min and pre-incubated in 30 mM sodium phosphate, 120 mM sodium chloride buffer, pH 7.4 at 22–23°C, for 30 min. For incubation, 100 mM sucrose, 100 nM GBR12909 (to block binding to dopamine transporter) and 50 pM [^{125}I]RTI-55 (NEX272, Perkin Elmer) were added to the buffer. To measure non-specific binding, 10 μM mazindol was added to a subset of slide mailers containing a representative set of duplicate slides. All unlabeled ligands were from Sigma. Incubation was carried out for 2 h at 22–23 °C. Sections were rinsed twice for 1 min in 4°C buffer (without sucrose), then dipped for 2 s in 4°C deionized water, drained and placed on a slide warmer (Lab-Line, Fisher Scientific, Pittsburgh, PA, USA), at moderate setting (4 on 10 scale) for 2 h.

Exposure and imaging

Sections on slides were exposed to Biomax MR film (Carestream/Kodak, Rochester, NY, USA) in a cassette for 48 h along with tritium standards (ART0123A, American Radiolabeled Chemicals, St. Louis, MO, USA) calibrated to [^{125}I]-incubated brain mash, as previously described.[198] Films were developed using an automatic film processor. Digital images of autoradiograms were captured using a 12-bit CCD monochrome digital camera (CFW-1612 M, Scion, Frederick, MD, USA) with a 60 mm lens (f-stop = 4)(Nikon, Melville, NY, USA) mounted on a copy stand (RS-1, Kaiser Fototechnik, White Plains, NY, USA) with a Kaiser Slimlite Plano LED lightbox. Pixel intensity was calibrated to measure density in units of femtomoles/mg (fmol/mg) protein using a linear function with ImageJ software (https://imagej.nih.gov/ij/download.html).[199]

OXTR data collection

OVT binding to the OXTR was measured in the anterior olfactory nucleus (AON), lateral septal nucleus (LSN), anterior cingulate cortex (ACC), striatum (CPu), hippocampal CA2 and CA3 regions (CA 2/3), paraventricular hypothalamic nucleus (PVN), piriform cortex (Pir), and the combined basolateral and lateral amygdala (BLA & LA) by tracing each region in ImageJ based on coordinates from the Franklin and Paxinos mouse brain atlas.[200] The AON was traced at Bregma 2.68 mm. The LSN, ACC and CPu were traced between Bregma 0.68 mm and 0.26 mm. The hippocampal CA 2/3 region, PVN and BLA & LA were traced between Bregma −1.58 mm and −1.82 mm. To mitigate potential variability across sections, each region was measured at multiple predetermined locations across consecutive sections within each sample. Nine or more animals were measured per genotype (**Table S2**).

SERT data collection

SERT availability was measured in the BLA, LA, central amygdala (CeA), ACC, CPu, CA 2/3, insular cortex (IC), lateral parietal association (PtA), nucleus accumbens (NAc), orbitofrontal cortex (OFC), peduncular part of the lateral hypothalamus (PLH) and the bed nucleus of the stria terminalis (BNST). The BLA, LA and CeA were traced between Bregma −1.58 and −1.82 mm. The ACC was traced at Bregma 0.74 mm, the IC was traced at Bregma −1.06 mm and the NAc was traced at 0.74 mm. The CPu was traced between Bregma 0.74 mm and 0.26 mm. The hippocampal CA 2/3 region was traced between Bregma −1.34 mm and −1.58 mm. The PtA and PLH were traced at Bregma −1.58 mm to −1.70 mm. The BNST was traced between Bregma 0.62 mm and −0.22 mm. Up to four independent measurements were taken at the same predetermined locations on consecutive sections within a sample. At least five animals were measured per genotype (**Table S2**).

2.5.6 Statistical analysis

Analysis for ELISA and the conditioned fear task was completed in R using RStudio (Version 1.2.5019). ELISA analysis utilized the "drc" package.[201] Conditioned fear data were condensed by minute then assessed for normality, homogeneity of variance and outliers. Data were analyzed with a linear mixed effect model using the "lme4" package,[202] with genotype as the main factor, and minute as a repeated measure. Tukey's HSD was employed for post hoc analysis. An effect of sex was screened for but not included in the final analysis, as there was no significant effect on any outcome. Detailed outputs for conditioned fear are included in **Table S1**.

Autoradiography analysis was performed using IBM SPSS Statistics for Windows, Version 26.0. Measurements within a region were averaged to compute the binding value of that region for each sample. These values were normalized by film to compute the total binding value for each region by subtracting the nonspecific binding value of a control sample from the sample binding value. Values less than zero after normalization were treated as zero. Normality, outliers and variance were assessed prior to hypothesis testing. Total OXTR binding values were transformed using a square root transformation to resolve normality violations. A 2×2 Analysis of Covariance with fixed factors of sex and genotype and a covariate of age was performed on each region of interest for OXTR autoradiography. For SERT, no transformations were necessary. A 2×2 multivariate Analysis of Variance was used to assess factors of sex and genotype across 12 regions of interest. Detailed outputs of statistical tests for all autoradiography experiments are in **Table S2**.

2.6 Acknowledgments

We would like to thank N. Kopp for his contributions and support, E. Minakova and D. Burek for surgical training and H.A. Stephens, Y.R. Sanchez, M.C. Nunn and B.L. Pierce for their expert technical assistance in tissue preparation. CD mice were a generous gift from V. Campuzano. The Animal Behavior Core at Washington University in St. Louis graciously shared their resources. Financial support was provided by the NIMH 5R01MH107515-05 (JDD), NSF DGE-1745038 (KRN), NICHD P50HD103525 (IDDRC@WUSTL) and John L. Santikos Charitable Foundation of the San Antonio Area Foundation and The Morrison Trust (GGG). The authors declare no competing financial interests.

2.7 Supplementary Tables

Table S1. Statistical Analysis for Conditioned Fear Task in Figures 1 and 2.

CONDITIONED FEAR TASK							
MICE	AGE	SAMPLE SIZE (n):	Wild Type (WT)		Complete Deletion (CD)		
	P97-P106		10		10		
			Female	Male	Female	Male	
			4	6	8	2	
Figure	Parameter (unit)	Comparison	Descriptive Statistics		Statistical Analysis		
		Independent Variables	Average ± SEM	Median (2Q,3Q)	Statistical Test	Significance	
Figure 1B	Acquisition - Baseline (% Freezing)	Baseline Minute 1	WT: 0 ± 0	WT: 0 (0, 0)	Linear mixed model, Random effect: Animal; Anova to test fixed effects; Tukey's HSD multiple comparison within minute	Genotype: F(1,18) = 1.355, p = 0.2596 Time: F(1,18) = 1.355, p = 0.2596 Time*Genotype Interaction: F(1,18) = 1.355, p = 0.2596	
			CD: 0 ± 0	CD: 0 (0, 0)			
		Baseline Minute 2	WT: 0 ± 0	WT: 0 (0, 0)			
			CD: 0.974 ± 0.8367	CD: 0 (0, 0)			
	Acquisition - Conditioned Stimulus (% Freezing)	Conditioned Stimulus Minute 3	WT: 4.824 ± 2.1754	WT: 2.43 (0, 5.4225)	Linear mixed model, Random effect: Animal; Anova to test fixed effects; Tukey's HSD multiple comparison within minute	**Time: F(2,36) = 77.809, p = 8.503x10^{-14}** Genotype: F(1,18) = 0.0768, p = 0.7848 Time*Genotype Interaction: F(2,36) = 0.0013, p = 0.9987	D1M3: CD - WT p = 0.997
			CD: 5.978 ± 1.6756	CD: 5.75 (1.665, 7.99)			D1M4: CD - WT p = 0.996
		Conditioned Stimulus Minute 4	WT: 23.239 ± 4.3305	WT: 21.24 (17.835, 28.54)			D1M5: CD - WT p = 0.993
			CD: 24.44 ± 4.9038	CD: 22.62 (17.26, 30.97)			
		Conditioned Stimulus Minute 5	WT: 49.539 ± 5.7833	WT: 52.89 (33.78, 61.3325)			
			CD: 51.038 ± 5.7448	CD: 46.225 (40.69, 67.79)			
Figure 1C	Contextual Fear Memory (Average % Freezing)	Average % Freezing Baseline	WT: 0 ± 0	WT: 0 (0, 0)	Linear mixed model, Random effect: Animal; Anova to test fixed effects; Post hoc comparison within genotypes between context	**Time: F(1,18) = 26.6553, p = 6.517x10^{-5}** Genotype: F(1,18) = 1.0801, p = 0.3124 Time*Genotype Interaction: F(1,18) = 1.3479, p = 0.2608	WT baseline - CD baseline p = 0.999826
			CD: 0.487 ± 0.4183	CD: 0 (0, 0)			**WT baseline - WT context p = 0.0004**
		Average % Freezing Context (First 2 Minutes)	WT: 27.7675 ± 5.8102	WT: 22.82 (12.88, 40.555)			**CD baseline - CD context p = 0.037928**
			CD: 18.059 ± 6.6325	CD: 8.54 (5.98, 19.66)			WT context - CD context = 0.41626
Figure 1D	Contextual Fear (% Freezing)	Contextual Fear Minute 1	WT: 22.758 ± 5.3559	WT: 23.23 (8.535, 27.89)	Linear mixed model, Random effect: Animal; Anova to test fixed effects; Tukey's HSD multiple comparison within minute	**Genotype: F(1,18) = 7.9504, p=0.01135** Time: F(7,126) = 1.6189, p = 0.13586 **Genotype*Time Interaction: F(7,126) = 1.8853, p = 0.07723**	D2M1: CD - WT p = 0.946
			CD: 16.746 ± 7.2904	CD: 7.315 (2.33, 23.4525)			D2M2: CD - WT p = 0.308
		Contextual Fear Minute 2	WT: 32.777 ± 6.933	WT: 27.655 (15.8925, 43.3375)			D2M3: CD - WT p = 0.123
			CD: 19.372 ± 6.9001	CD: 11.335 (6.2025, 26.215)			**D2M4: CD - WT p = 0.0014**
		Contextual Fear Minute 3	WT: 32.747 ± 6.4617	WT: 32.3 (18.805, 42.6975)			D2M5: CD - WT p = 0.147
			CD: 15.898 ± 5.192	CD: 9.98 (4.33, 26.1225)			**D2M6: CD - WT p = 0.0809**
		Contextual Fear Minute 4	WT: 42.403 ± 5.8707	WT: 42.475 (30.63, 50.775)			D2M7: CD - WT p = 0.315
			CD: 14.705 ± 3.7242	CD: 12.445 (7.0875, 20.1325)			**D2M8: CD - WT p = 0.0114**
		Contextual Fear Minute 5	WT: 32.32 ± 6.2531	WT: 34.96 (19.4425, 42.145)			
			CD: 16.087 ± 3.8355	CD: 12.195 (8.63, 22.015)			
		Contextual Fear Minute 6	WT: 31.034 ± 4.7758	WT: 29.425 (23.4225, 44.44)			
			CD: 12.85 ± 4.1328	CD: 7.37 (3.7775, 14.725)			
		Contextual Fear Minute 7	WT: 26.613 ± 3.9524	WT: 31.195 (13.38, 36.14)			
			CD: 13.3 ± 2.6011	CD: 11.75 (8.18, 16.6075)			
		Contextual Fear Minute 8	WT: 33.377 ± 5.3107	WT: 33.66 (20.1125, 43.9975)			
			CD: 9.955 ± 3.1435	CD: 5.33 (3.255, 14.2225)			
Figure 1E		Baseline Minute 1	WT: 4.118 ± 1.3881	WT: 2.21 (1.33, 5.11)	Linear mixed model, Random effect: Animal; Anova to test fixed effects	**Time: F(1,18) = 16.9901, p = 0.00064** Genotype: F(1,18) = 2.1989, p = 0.1554 Genotype*Time Interaction: F(1,18) = 0.326, p = 0.57506	
			CD: 1.422 ± 1.0962	CD: 0 (0, 0.9975)			
		Baseline Minute 2	WT: 9.337 ± 1.9484	WT: 8.185 (5.4225, 11.945)			
			CD: 5.371 ± 2.375	CD: 3.545 (1.66, 5.445)			
	Cued Fear (% Freezing)	Cued Fear Minute 3	WT: 48.178 ± 5.402	WT: 47.005 (33.33, 63.6075)	Linear mixed model, Random effect: Animal; Anova to test fixed effects; Tukey's HSD multiple comparison within minute	**Time: F(9,162) = 22.0195, p < 2x10^{-16}** Genotype: F(1,18) = 2.1895, p = 0.1562 **Genotype*Time Interaction: F(9,162) = 1.9151, p = 0.05315**	D3M3: CD - WT p = 0.905
			CD: 38.751 ± 6.7536	CD: 44.915 (25.96, 50.5075)			**D3M4: CD - WT p = 0.0364**
		Cued Fear Minute 4	WT: 48.717 ± 7.7196	WT: 57.965 (27.655, 63.7175)			D3M5: CD - WT p = 0.368
			CD: 24.823 ± 4.9433	CD: 20.13 (13.9375, 31.8625)			D3M6: CD - WT p = 0.996
		Cued Fear Minute 5	WT: 32.212 ± 5.2023	WT: 31.195 (21.3475, 42.5875)			D3M7: CD - WT p = 0.732
			CD: 16.238 ± 4.4482	CD: 11.285 (8.96, 16.0375)			D3M8: CD - WT p = 0.993
		Cued Fear Minute 6	WT: 24.556 ± 4.3495	WT: 22.345 (14.71, 31.415)			D3M9: CD - WT p = 0.998
			CD: 18.936 ± 6.5093	CD: 8.625 (4.42, 28.98)			D3M10: CD - WT p = 1.00
		Cued Fear Minute 7	WT: 29.486 ± 6.6092	WT: 22.395 (17.5325, 29.205)			
			CD: 17.577 ± 6.7742	CD: 12 (4.2025, 18.365)			
		Cued Fear Minute 8	WT: 25.108 ± 5.3123	WT: 21.46 (12.6475, 33.1075)			
			CD: 19.055 ± 5.9164	CD: 15.045 (6.5225, 26.11)			
		Cued Fear Minute 9	WT: 24.845 ± 3.5911	WT: 23.5 (19.135, 25.55)			
			CD: 19.868 ± 6.7176	CD: 16.595 (4.885, 21.1525)			
		Cued Fear Minute 10	WT: 21.466 ± 3.5788	WT: 19.555 (13.6675, 27.445)			
			CD: 20.401 ± 7.8675	CD: 9.555 (5.5575, 26.225)			

**p-values < 0.1 are bolded*

CONDITIONED FEAR TASK with ICV TREATMENT							
MICE	AGE	SAMPLE SIZE (n):	WT - Vehicle	WT - OTA	CD - Vehicle	CD - OTA	
	P68-P118		16	13	7	7	
			Female | Male	Female | Male	Female | Male	Female | Male	
			8 | 8	4 | 9	5 | 2	3 | 4	
Figure	Parameter (unit)	Comparison	Descriptive Statistics		Statistical Analysis		
		Independent Variables	Average ± SEM	Median (2Q,3Q)	Statistical Test	Significance	
Figure 2B	Acquistion - Baseline (% Freezing)	Baseline Minute 1	WT-Veh: 0 ± 0	WT-Veh: 0 (0, 0)	Linear mixed model, Random effect: Animal; Anova to test fixed effects	Time: F(1,39) = 0.2699, p = 0.6063 Genotype: F(1,39) = 1.5766, p = 0.2167 Treatment: F(1,39) = 0.3356, p = 0.5657 INTERACTIONS: Freeze*Genotype: F(1,39) = 2.3043, p = 0.1371 Freeze*Treatment: F(1,39) = 1.5034, p = 0.2275 Genotype*Treatment: F(1,39) = 2.2488, p = 0.1418 Freeze*Genotype*Treatment: F(1,39) = 2.4906, p = 0.1226	
			WT-OTA: 0.1362 ± 0.1362	WT-OTA: 0 (0, 0)			
			CD-Veh: 1.7071 ± 1.7072	CD-Veh: 0 (0, 0)			
			CD-OTA: 0 ± 0	CD-OTA: 0 (0, 0)			
		Baseline Minute 2	WT-Veh: 0.0831 ± 0.0831	WT-Veh: 0 (0, 0)			
			WT-OTA: 0.2731 ± 0.1916	WT-OTA: 0 (0, 0)			
			CD-Veh: 0.3814 ± 0.3814	CD-Veh: 0 (0, 0)			
			CD-OTA: 0.19 ± 0.19	CD-OTA: 0 (0, 0)			
	Acquisition - Conditioned Stimulus (% Freezing)	Conditioned Stimulus Minute 3	WT-Veh: 4.7894 ± 1.4969	WT-Veh: 2.43 (0, 6.5225)	Linear mixed model, Random effect: Animal; Anova to test fixed effects; Tukey's HSD multiple comparison within minute	Genotype: F(1,39) = 1.7043, p = 0.19938 Treatment: F(1,39) = 0.5815, p = 0.45032 **Time: F(2,78) = 155.1855, p < 2x10^{-16}** INTERACTIONS: **Genotype*Treatment: F(1,39) = 7.3211, p = 0.01** Genotype*Time: F(2,78) = 0.3276, p = 0.72163 Treatment*Time: F(2,78) = 0.0271, p = 0.97329 Genotype*Treatment*Time: F(2,78) = 1.8232, p = 0.16833	D1M3: CD Veh - WT Veh p = 0.9999
			WT-OTA: 9.6577 ± 2.5153	WT-OTA: 10.67 (0, 17.7)			D1M3: WT OTA - WT Veh p = 0.9842
			CD-Veh: 6.4557 ± 1.8984	CD-Veh: 7.52 (2.67, 8.865)			D1M3: CD OTA - WT Veh p = 0.9999
			CD-OTA: 2.7886 ± 1.2213	CD-OTA: 3.11 (0, 3.76)			D1M3: WT OTA - CD Veh p = 0.9998
		Conditioned Stimulus Minute 4	WT-Veh: 19.855 ± 3.079	WT-Veh: 15.115 (9.297, 27.762)			D1M3: CD OTA - CD Veh p = 0.9998
			WT-OTA: 30.9408 ± 5.9432	WT-OTA: 25.78 (8.44, 46.9)			D1M3: CD OTA - WT OTA p = 0.9658
			CD-Veh: 26.9214 ± 3.2769	CD-Veh: 29.2 (22.83, 31.71)			D1M4: CD Veh - WT Veh p = 0.9497
			CD-OTA: 13.38 ± 3.0608	CD-OTA: 12.44 (8.21, 16.41)			D1M4: WT OTA - WT Veh p = 0.3034
		Conditioned Stimulus Minute 5	WT-Veh: 41.3312 ± 4.3103	WT-Veh: 36.665 (29.78, 53.11)			D1M4: CD OTA - WT Veh p = 0.9709
			WT-OTA: 50.8969 ± 4.3256	WT-OTA: 51.56 (48.21, 58.48)			D1M4: WT OTA - CD Veh p = 0.9993
			CD-Veh: 46.0129 ± 7.6612	CD-Veh: 46.67 (28.89, 63.33)			D1M4: CD OTA - CD Veh p = 0.5260
			CD-OTA: 33.2943 ± 2.3647	CD-OTA: 31.56 (29.78, 34.97)			**D1M4: CD OTA - WT OTA p = 0.0779**
							D1M5: CD Veh - WT Veh p = 0.9970
							D1M5: WT OTA - WT Veh p = 0.5101
							D1M5: CD OTA - WT Veh p = 0.8949
							D1M5: WT OTA - CD Veh p = 0.9968
							D1M5: CD OTA - CD Veh p = 0.6124
							D1M5: CD OTA - WT OTA p = 0.0765
not pictured	Contextual Fear Memory (Average % Freezing)	Average % Freezing Baseline	WT-Veh: 0.04156 ± 0.0415	WT-Veh: 0 (0, 0)	Linear mixed model, Random effect: Animal; Anova to test fixed effects; Post hoc comparison within genotypes between context not considering treatment	**Time: F(1,39) = 196.2863, p = 2.2x10^{-16}** **Genotype: F(1,39) = 22.5684, p = 3.216x10^{-5}** Treatment: F(1,39) = 1.0977, p = 0.3012 INTERACTIONS: **Time*Genotype: F(1,39) = 26.7991, p = 7.159x10^{-6}** Time*Treatment: F(1,39) = 1.4433, p = 0.2369 Genotype*Treatment: F(1,39) = 1.9603, p = 0.1694 Time*Genotype*Treatment: F(1:39) = 1.3381, p = 0.2544	CD Baseline - WT Baseline p = 0.9983
			WT-OTA: 0.2046 ± 0.1577	WT-OTA: 0 (0, 0)			**WT Context - WT Baseline p = 0**
			CD-Veh: 1.044 ± 1.0443	CD-Veh: 0 (0, 0)			**CD Context - CD Baseline p = 0.0022**
			CD-OTA: 0.095 ± 0.095	CD-OTA: 0 (0, 0)			**CD Context - WT Context p = 2.9878x10^{-9}**
		Average % Freezing Context (First 2 Minutes)	WT-Veh: 27.758 ± 3.122	WT-Veh: 25.183 (20.630, 32.709)			
			WT-OTA: 35.01 ± 3.6576	WT-OTA: 31.29 (21.97, 45.13)			
			CD-Veh: 13.513 ± 3.2704	CD-Veh: 12.19 (9.75, 16.925)			
			CD-OTA: 10.959 ± 3.3	CD-OTA: 5.540 (4.768, 18.942)			
Figure 2C	Contextual Fear (% Freezing)	Contextual Fear Minute 1	WT-Veh: 20.1638 ± 2.7597	WT-Veh: 15.705 (12.952, 27.655)	Linear mixed model; Animal id random effect; Anova to test fixed effects; Tukey's HSD multiple comparison within time only considering genotype	**Genotype: F(1,39) = 19.9632, p = 6.6x10-5** **Time: F(7,273) = 3.2942, p = 0.002221** Treatment: F(1,39) = 1.1557, p = 0.288966 INTERACTIONS: Time*Genotype: F(7,273) = 1.3534, p = 0.225352 Time*Treatment: F(7,273) = 0.2155, p = 0.981667 Genotype*Treatment: F(1,39) = 2.4705, p = 0.124 Time*Genotype*Treatment: F(7:273) = 0.5370, p = 0.806138	**D2M1: CD - WT p = 0.07126**
			WT-OTA: 27.2877 ± 3.5031	WT-OTA: 25.22 (18.58, 38.94)			**D2M2: CD - WT p = 9.7218x10^{-5}**
			CD-Veh: 11.9529 ± 3.3111	CD-Veh: 11.11 (7.74, 15.045)			**D2M3: CD - WT p = 0.00826**
			CD-OTA: 7.27 ± 2.2227	CD-OTA: 6.64 (2.875, 9.735)			**D2M4: CD - WT p = 0.00037**
		Contextual Fear Minute 2	WT-Veh: 35.3519 ± 4.1518	WT-Veh: 32.82 (26.1925, 45.69)			**D2M5: CD - WT p = 0.00116**
			WT-OTA: 42.7385 ± 4.6455	WT-OTA: 39.56 (30.67, 55.31)			**D2M6: CD - WT p = 0.00886**
			CD-Veh: 15.0729 ± 3.5278	CD-Veh: 17.33 (9.73, 18.805)			D2M7: CD - WT p = 0.13113
			CD-OTA: 14.6486 ± 4.9996	CD-OTA: 8.89 (4.655, 23.06)			D2M8: CD - WT p = 0.19205
		Contextual Fear Minute 3	WT-Veh: 29.0006 ± 4.8913	WT-Veh: 24.385 (13.3875, 47.9)			
			WT-OTA: 37.5815 ± 5.5208	WT-OTA: 35.56 (21.24, 49.56)			
			CD-Veh: 13.5443 ± 3.2708	CD-Veh: 13.72 (8.43, 15.3)			
			CD-OTA: 16.6529 ± 4.6414	CD-OTA: 13.33 (9.09, 17.955)			
		Contextual Fear Minute 4	WT-Veh: 29.6837 ± 3.2344	WT-Veh: 34.955 (18.557, 37.997)			
			WT-OTA: 37.9792 ± 6.3412	WT-OTA: 42.67 (18.67, 54.22)			
			CD-Veh: 12.5529 ± 1.8664	CD-Veh: 12.39 (8, 16.85)			
			CD-OTA: 10.0829 ± 4.1256	CD-OTA: 7.52 (4.445, 9.53)			
		Contextual Fear Minute 5	WT-Veh: 27.4537 ± 4.936	WT-Veh: 19.245 (12.875, 37.1025)			
			WT-OTA: 37.1962 ± 6.0902	WT-OTA: 29.2 (25.22, 44.89)			
			CD-Veh: 14.61 ± 3.0534	CD-Veh: 17.26 (9.54, 20.795)			
			CD-OTA: 7.8471 ± 2.9955	CD-OTA: 3.98 (1.99, 13.085)			
		Contextual Fear Minute 6	WT-Veh: 28.3244 ± 3.6164	WT-Veh: 23.23 (18.67, 37.4)			
			WT-OTA: 39.2915 ± 5.8313	WT-OTA: 37.17 (33.78, 48.89)			
			CD-Veh: 21.2857 ± 3.3255	CD-Veh: 24.44 (15.085, 26.99)			
			CD-OTA: 10.01 ± 3.4947	CD-OTA: 9.33 (2, 16.2)			
		Contextual Fear Minute 7	WT-Veh: 28.5981 ± 4.621	WT-Veh: 25.11 (15.83, 37.18)			
			WT-OTA: 34.5492 ± 7.312	WT-OTA: 28.89 (13.72, 55.56)			
			CD-Veh: 19.4743 ± 3.8466	CD-Veh: 22.57 (11.08, 23.895)			
			CD-OTA: 18.2829 ± 5.1421	CD-OTA: 19.03 (10.4, 23.275)			
		Contextual Fear Minute 8	WT-Veh: 26.7512 ± 3.9562	WT-Veh: 26.06 (16.225, 34.667)			
			WT-OTA: 33.1485 ± 6.9856	WT-OTA: 20.89 (20.09, 48.21)			
			CD-Veh: 23.7514 ± 7.9418	CD-Veh: 17.78 (8.93, 31.38)			
			CD-OTA: 12.6029 ± 4.1604	CD-OTA: 10.27 (3.34, 19.115)			
	Cued Fear Baseline (% Freezing)	Baseline Minute 1	WT-Veh: 6.7538 ± 1.7976	WT-Veh: 3.98 (1.33, 9.1825)	Linear mixed model, Random effect: Animal; Anova to test fixed effects	**Time: F(1,39) = 24.2008, p = 1.641x10^{-5}** **Genotype: F(1,39) = 5.2402, p = 0.02757** Treatment: F(1,39) = 0.7865, p = 0.38059 INTERACTIONS: **Time*Genotype: F(1,39) = 3.3477, p = 0.07495** Time*Treatment: F(1,39) = 0.2473, p = 0.62176 Genotype*Treatment: F(1,39) = 0.0549, p = 0.8159 Time*Genotype*Treatment: F(1,39) = 0.0429, p = 0.83691	
			WT-OTA: 4.6977 ± 1.8858	WT-OTA: 3.1 (0, 4)			
			CD-Veh: 2.91 ± 1.0616	CD-Veh: 1.77 (1.33, 3.765)			
			CD-OTA: 2.8486 ± 0.8253	CD-OTA: 3.1 (1.325, 4.22)			
		Baseline Minute 2	WT-Veh: 17.6231 ± 4.0732	WT-Veh: 12 (6.8575, 21.6825)			
			WT-OTA: 14.3885 ± 3.1284	WT-OTA: 10.62 (6.19, 21.78)			
			CD-Veh: 7.9814 ± 2.8593	CD-Veh: 4 (1.555, 14.16)			
			CD-OTA: 5.2657 ± 2.8701	CD-OTA: 2.65 (0, 7.105)			

Figure 2D	Cued Fear (% Freezing)	Cued Fear Minute 3	WT-Veh: 58.7388 ± 3.9924	WT-Veh: 61.78 (50.165, 69.247)	Linear mixed model, Random effect: Animal; Anova to test fixed effects; Tukey's HSD multiple comparison within time for genotype alone	**Time: F(2,273) = 26.447, p = $2.2x10^{-16}$** Genotype: F(1,39) = 1.0398, p = 0.3142 Treatment: F(1,39) = 0.0596, p = 0.8085 INTERACTIONS: **Time*Genotype: F(7,273) = 5.8774, p = $2.262x10^{-6}$** Time*Treatment: F(7,273) = 1.1599, p = 0.3261 Genotype*Treatment: F(1,39) = 0.1607, p = 0.6907 Time*Genotype*Treatment: F(2,273) = 0.4794, p = 0.8492	**D3M3: CD - WT p = 0.080759**
			WT-OTA: 55.6 ± 2.6732	WT-OTA: 56.44 (48.44, 64.6)			**D3M4: CD - WT p = 0.008288**
			CD-Veh: 43.3286 ± 6.6848	CD-Veh: 49.12 (33.335, 55.53)			D3M5: CD - WT p = 0.433508
			CD-OTA: 37.1 ± 8.9436	CD-OTA: 38.22 (21.505, 52.345)			D3M6: CD - WT p = 0.971907
		Cued Fear Minute 4	WT-Veh: 60.6531 ± 5.2684	WT-Veh: 67.18 (46.18, 75.66)			D3M7: CD - WT p = 0.999868
			WT-OTA: 49.6308 ± 4.1193	WT-OTA: 46.46 (44.25, 55.11)			D3M8: CD - WT p = 0.997868
			CD-Veh: 34.0129 ± 7.7645	CD-Veh: 25.66 (19.555, 43.14)			D3M9: CD - WT = 1
			CD-OTA: 33.3957 ± 12.1127	CD-OTA: 16.81 (11.54, 57.385)			D3M10: CD - WT p = 0.985158
		Cued Fear Minute 5	WT-Veh: 46.6925 ± 6.7533	WT-Veh: 50.335 (25.137, 65.795)			
			WT-OTA: 40.2008 ± 2.7712	WT-OTA: 38.94 (32.74, 44.69)			
			CD-Veh: 31.0643 ± 7.5289	CD-Veh: 30.67 (13.495, 41.7)			
			CD-OTA: 32.2329 ± 11.8144	CD-OTA: 21.68 (8.865, 48.67)			
		Cued Fear Minute 6	WT-Veh: 40.1825 ± 6.546	WT-Veh: 39.025 (13.095, 58.22)			
			WT-OTA: 36.0385 ± 3.8887	WT-OTA: 33.19 (26.55, 42.04)			
			CD-Veh: 29.8943 ± 8.6538	CD-Veh: 24.34 (18.22, 29.33)			
			CD-OTA: 34.6143 ± 11.9848	CD-OTA: 23.45 (14, 48.965)			
		Cued Fear Minute 7	WT-Veh: 32.3075 ± 5.1365	WT-Veh: 30.365 (16.595, 46.0175)			
			WT-OTA: 30.5985 ± 4.0338	WT-OTA: 26.55 (21.68, 38.67)			
			CD-Veh: 30.9157 ± 9.9984	CD-Veh: 22.12 (10.635, 47.02)			
			CD-OTA: 37.7671 ± 11.4266	CD-OTA: 25.22 (20.8, 53.045)			
		Cued Fear Minute 8	WT-Veh: 22.945 ± 4.4636	WT-Veh: 16.37 (14.16, 29.3325)			
			WT-OTA: 27.6538 ± 4.5452	WT-OTA: 29.2 (19.56, 35.56)			
			CD-Veh: 27.4043 ± 2.0588	CD-Veh: 26.55 (23.945, 28.83)			
			CD-OTA: 30.9014 ± 10.2019	CD-OTA: 22.67 (14.16, 43.505)			
		Cued Fear Minute 9	WT-Veh: 27.0338 ± 5.3605	WT-Veh: 20.135 (9.975, 46.68)			
			WT-OTA: 25.1954 ± 3.5282	WT-OTA: 20.8 (14.16, 37.17)			
			CD-Veh: 28.29 ± 7.1723	CD-Veh: 19.11 (17.56, 35.015)			
			CD-OTA: 23.23 ± 8.9934	CD-OTA: 16.37 (9.345, 27.775)			
		Cued Fear Minute 10	WT-Veh: 23.5981 ± 6.0351	WT-Veh: 14.51 (6.5875, 24.7775)			
			WT-OTA: 25.5585 ± 3.4353	WT-OTA: 25.89 (16, 32.59)			
			CD-Veh: 25.2943 ± 6.221	CD-Veh: 22.22 (15.775, 30.075)			
			CD-OTA: 34.6057 ± 12.4249	CD-OTA: 23.56 (9.12, 59.11)			
**p-values < 0.1 are bolded, Veh = Vehicle*							

Table S2. Statistical Analysis for OXTR and SERT Autoradiography in Figures 3 and 4.

OXTR Autoradiography Statistical Analysis									
Region	**Full Region Name**	**Genotype**	**Total n**	**Female n**	**Female Mean +/- SEM**	**Male n**	**Male Mean +/- SEM**	**ANCOVA Results[a,b]**	**Corrected p-value**
ACC	anterior cingulate cortex	WT	17	10	99.3212 ± 34.1635	7	132.8717 ± 47.3068	GENO: F(1,22)=0.324, p=0.575	p = 4.6
		CD	10	3	77.4752 ± 55.1005	7	101.0427 ± 41.3338	SEX: F(1,22)=0.375, p=0.546	p = 4.368
AON	aortic optic nucleus	WT	14	9	763.0854 ± 48.2315	5	886.8484 ± 72.0676	GENO: F(1,20)=0.355, p=0.558	p = 4.464
		CD	11	4	793.2109 ± 73.2264	7	779.8056 ± 54.9005	SEX: F(1,20)=0.740, p=0.4	p = 3.2
BLA	basolateral & lateral amygdala	WT	15	10	528.4022 ± 54.7091	5	482.8567 ± 74.3161	GENO: F(1,21)=1.57, p=0.224	p = 1.792
		CD	11	5	613.8502 ± 83.4951	6	578.4987 ± 73.9358	SEX: F(1,21)=0.319, p=0.578	p = 4.624
CA 2/3	hippocampus regions CA 2 & 3	WT	20	12	486.5112 ± 32.1591	8	582.4017 ± 43.729	GENO: F(1,28)=0.234, p=0.632	p = 5.056
		CD	13	5	569.4905 ± 54.1236	8	539.5399 ± 41.4852	SEX: F(1,28)=0.599, p=0.445	p = 3.56
Cpu	caudate putamen/striatum	WT	17	10	17.8168 ± 15.4067	7	23.1842 ± 21.0416	GENO: F(1,21)=0.023, p=0.88	p = 7.04
		CD	9	3	18.6106 ± 28.783	6	30.0852 ± 25.9221	SEX: F(1,21)=0.126, p=0.726	p = 5.808
LSN	lateral septal nucleus	WT	17	10	443.7764 ± 41.7528	7	462.5080 ± 51.0552	**GENO: F(1,22)=5.118, p=0.034**	p = 0.272
		CD	10	3	282.1728 ± 60.8088	7	388.4841 ± 46.9098	SEX: F(1,22)=1.595, p=0.22	p = 1.76
Pir	piriform cortex	WT	15	10	746.1092 ± 58.0171	5	723.9866 ± 81.3129	GENO: F(1,20)=0.847, p=0.368	p = 2.944
		CD	10	4	634.4353 ± 85.2362	6	696.4849 ± 72.3641	SEX: F(1,20)=0.077, p=0.784	p = 6.272
PVN	paraventricular nucleus	WT	20	12	646.8375 ± 55.0879	8	654.8481 ± 69.093	GENO: F(1,27)=0.477, p=0.496	p = 3.968
		CD	12	4	641.9636 ± 96.3819	8	560.0322 ± 62.8069	SEX: F(1,27)=0.27, p=0.608	p = 4.864
By genotype and sex from estimated marginal means; SEM = standard error of the mean, GENO = genotype									
[a] Fixed factors: sex and genotype, covariate: age; [b] Bonferroni Correction, critical α = 0.00625									

SERT Autoradiography Statistical Analysis									
Region	**Full Region Name**	**Genotype**	**Total n**	**Female n**	**Female Mean +/- SEM**	**Male n**	**Male Mean +/- SEM**	**Multivariate ANOVA Results[a,b]**	**Between-Subjects Effects**
ACC	anterior cingulate cortex	WT	22	12	289.4284 ± 61.8965	10	289.1077 ± 79.0262	**Genotype: F(12,12) = 2.141, p = 0.07, Wilks' Lambda = 0.293** Sex: F(12,12) = 0.489, p = 0.885, Wilks' Lambda = 0.672 Genotype*Sex: F(12,12) = 1.301, p = 0.328, Wilks' Lambda = 0.435	GENO: F(1,23)=0.527,p=0.475
		CD	14	5	211.2059 ± 75.0746	9	218.0609 ± 62.2096		SEX: F(1,23)=0.166,p=0.687
BLA	basolateral amygdala	WT	21	11	1542.0245 ± 136.8943	10	1326.6276 ± 149.6572		GENO: F(1,23)=0.002,p=0.965
		CD	14	5	1525.4806 ± 151.2241	9	1266.8352 ± 119.4666		SEX: F(1,23)=0.638,p=0.432
BNST	bed nucleus of the stria terminalus	WT	21	12	611.3552 ± 90.9035	9	531.8985 ± 127.1372		GENO: F(1,23)=0.146,p=0.706
		CD	11	5	722.652 ± 114.8744	6	503.8346 ± 97.9522		SEX: F(1,23)=0.328,p=0.572
CA 2/3	hippocampus regions CA 2 & 3	WT	22	12	978.7105 ± 108.5505	10	873.5517 ± 137.29		GENO: F(1,23)=2.065,p=0.164
		CD	14	5	781.6372 ± 150.0473	9	643.6365 ± 87.8401		SEX: F(1,23)=0.106,p=0.747
CeA	central amygdala	WT	21	11	1374.99 ± 133.9068	10	1108.807 ± 126.8176		GENO: F(1,23)=1.315,p=0.263
		CD	14	5	1125.4029 ± 88.5061	9	1055.1931 ± 89.3567		SEX: F(1,23)=0.227,p=0.638
Cpu	caudate putamen/striatum	WT	22	12	1622.2558 ± 74.8469	10	1446.2245 ± 156.5769		**GENO: F(1,23)=3.7,p=0.067**
		CD	14	5	1409.6564 ± 131.3332	9	1281.6204 ± 128.598		SEX: F(1,23)=0.110,p=0.744
IC	insular cortex	WT	21	12	606.6271 ± 96.5621	9	678.7568 ± 152.4349		GENO: F(1,23)=0.125,p=0.727
		CD	13	5	701.5797 ± 137.3324	8	511.0048 ± 78.1395		SEX: F(1,23)=0.003,p=0.956
LA	lateral amygdala	WT	20	11	1167.0251 ± 109.2435	9	1278.5066 ± 123.0872		GENO: F(1,23)=0.569p=0.458
		CD	14	5	1176.4828 ± 180.1849	9	908.4341 ± 127.2575		SEX: F(1,23)=0.022,p=0.883
NAc	nucleus accumbens	WT	21	12	1379.1824 ± 65.8827	10	1131.0983 ± 160.7995		GENO: F(1,23)=1.326,p=0.261
		CD	14	5	1252.071 ± 149.5267	9	1044.5984 ± 95.8042		SEX: F(1,23)=1.551,p=0.226
OFC	orbital frontal cortex	WT	21	11	486.3343 ± 59.6736	9	539.8918 ± 84.9561		**GENO: F(1,23)=3.903,p=0.06**
		CD	13	5	305.4111 ± 89.3258	8	297.4236 ± 63.7359		SEX: F(1,23)=0.155,p=0.698
PLH	hypothalamus	WT	22	12	1687.884 ± 65.4506	10	1366.4643 ± 112.7004		GENO: F(1,23)=2.64,p=0.118
		CD	13	5	1711.2213 ± 156.022	8	1558.616 ± 161.5287		SEX: F(1,23)=0.608,p=0.444
PtA	lateral parietal association	WT	21	12	297.6349 ± 55.9814	9	385.0822 ± 107.0252		GENO: F(1,23)=0.41,p=0.528
		CD	12	5	342.1413 ± 117.7591	7	206.3785 ± 30.2895		SEX: F(1,23)=0.113,p=0.74
SEM = standard error of the mean, GENO = genotype									

Chapter 3: Extensive characterization of a Williams Syndrome murine model shows *Gtf2ird1*-mediated rescue of select sensorimotor tasks, but no effect on enhanced social behavior

Kayla R. Nygaard, Susan E. Maloney, Raylynn G. Swift, Katherine B. McCullough, Rachael E. Wagner, Stuart B. Fass, Krassimira Garbett, Karoly Mirnics, Jeremy Veenstra-VanderWeele, and Joseph D. Dougherty

3.1 Abstract

Williams Syndrome (WS) is a rare neurodevelopmental disorder exhibiting cognitive and behavioral abnormalities, including increased social motivation, yet also risk of anxiety and specific phobias along with motor delay. WS is caused by a microdeletion of 26-28 genes on chromosome 7, including the GTF2IRD1 transcription factor, which has been suggested to play a role in the behavioral profile of the WS. Duplications of the full region also lead to frequent autism diagnosis, social phobias, and language delay. Thus, genes in the region appear to regulate social motivation in a dose-sensitive manner. A complete deletion (CD) mouse, heterozygously eliminating the syntenic WS region, has been deeply characterized for cardiac phenotypes, but direct measures of social motivation have not been assessed. Furthermore, the role of *Gtf2ird1* in these behaviors has not been addressed in a relevant genetic context. Here, we have generated a mouse overexpressing *Gtf2ird1*, which can be used both to model duplication of this gene alone and to rescue *Gtf2ird1* expression in the CD mice. Using a comprehensive behavioral pipeline and direct measures of social motivation, we provide evidence that the CD locus regulates social motivation along with motor and anxiety phenotypes, but that *Gtf2ird1* complementation is not sufficient to rescue most of these traits, and duplication does not decrease social motivation. However, *Gtf2ird1* complementation does rescue light-aversive behavior and performance on

select sensorimotor tasks, perhaps indicating a role for this gene in sensory processing or integration.

3.2 Introduction

Microdeletion of the 7q11.23 region results in the neurodevelopmental disorder known as Williams Syndrome (WS). Hemizygosity of the 26-28 genes in this region, also known as the Williams Syndrome Critical Region (WSCR), causes multisystemic symptoms which match some features and mirror others from reciprocal 7q11.23 Duplication Syndrome (Dup 7), revealing the importance of gene dosage in the pathophysiology of these disorders.[69,110,203] Both syndromes result in altered craniofacial features, cardiac issues, motor coordination deficits, and behavioral challenges.[1,48] Many people with WS exhibit hypersociability and tend to approach strangers with little apprehension, though the lack of social anxiety does not preclude a more generalized anxiety and occasional extreme phobias, both of which are more prevalent in WS than the general population.[18,35,47] Unfortunately, the underlying mechanisms of these behavioral differences are not well understood and thus there are no targeted treatments to help individuals with WS navigate the expectations of society, similar to the struggle autistic individuals face. However, unlike the complex etiology of idiopathic autism, the discrete genetic foundation of WS provides a unique opportunity to uncover these mechanisms, as a relatively small deletion leads to such a recognizable behavioral profile.

As WS and Dup7 are rare, understanding the complex etiology and circuit pathology underlying behavioral phenotypes in humans, or with human brain samples, is challenging. While cellular phenotypes can be investigated in iPSC models,[69] animal models are still required to uncover the link between gene dosage and behavioral phenotypes. Fortunately, the WSCR is

syntenic in mice, and a complete deletion (CD) mouse model has been developed that recapitulates many features of WS.[159]

An initial survey of various features in the CD mouse line discovered mild cardiac deficits, craniofacial anomalies, and some alterations in behavior. Specifically, the authors reported deficits in motor performance in a Rotarod task and a decreased habituation to social stimuli in an open field social interaction test.[159] However, the classic three-chamber social approach task showed no difference in CD mice (albeit on a FVB/AntJ x C57BL/6J hybrid background);[117] the FVB strain generally shows less social approach than C57BL/6J,[124] suggesting this background may be less sensitive for CD social phenotyping. However, the previous measures showing increased social interest were only conducted in males and social motivation, the amount of work an animal is willing to do to engage with a conspecific, was not directly measured. Likewise, while motor learning has been assessed on the Rotarod apparatus,[159] less work has characterized motor strength or coordination generally and the results were not consistent on the hybrid background with slightly different test parameters.[117] Finally, anxiety has also been a difficult domain to assess consistently in mice, as transient emotionality can affect the results.[122] Similar mouse models deleting a single gene in the WSCR, *Gtf2ird1*, show opposite results in anxiety-like behavior.[131,137] Overall, a deeper phenotyping of these domains would be of use, especially to best address the effects of WSCR copy number variation at the level of mechanisms or circuits.

Prior to the development of the CD line modeling the full deletion, single gene deletions were the most common approach in trying to elucidate function, and *Gtf2ird1* is one gene that has been implicated in a variety of hallmark WS phenotypes, from craniofacial to cognitive and behavioral differences. *Gtf2ird1* is often implicated alongside its neighbor and family member *Gtf2i*, as these genes occur in tandem in the WSCR and are rarely found separately affected by

atypical deletions, thus it is difficult to isolate their effects using human studies alone. While both genes are conserved in the mouse genome, there has been more difficulty with reliably producing a *Gtf2ird1* knockout animal. Alternative and frame-shifted start codons allow truncated versions of the protein to be expressed, even preserving much of its function outside of its negative autoregulation.[139] Verifying *Gtf2ird1* expression, or lack thereof, was also unreliable prior to the development of decent antibodies.

To avoid the trouble of deleting this elusive gene, we adopt a different strategy to gage the influence of *Gtf2ird1* on relevant phenotypes; we designed a study to both assess the impact of *Gtf2ird1* while also providing an extensive characterization of the CD model, as both of these contributions would benefit our understanding of the WSCR. Thus, we present a novel *Gtf2ird1* transgenic expression line, which we use to thoroughly assess the role of *Gtf2ird1*; we test the hypothesis that *Gtf2ird1* plays a dose-dependent role in the cognitive and behavioral symptoms of WS, concurrently examining the effects of *Gtf2ird1* overexpression on the background of both C57BL/6J and CD mouse lines. We used a comprehensive battery of tasks designed to elucidate the contributions of *Gtf2ird1* to the WS phenotype. Simultaneously, using the same extensive suite of behavioral measures, we provide a detailed assessment of the CD mouse, providing key information on phenotypes related to motor, anxiety, fear, and social behaviors.

As hypersociability is a key feature of WS, we characterized social behavior of these mice by measuring a suite of relevant behaviors using 3-Chamber Social Approach, Open Field Social Approach,[159] Social Motivation Operant, Maternal Isolation-induced Pup Ultrasonic Vocalization (USV), and Resident Intruder tasks. To measure anxiety-relevant behaviors in the CD mice, we utilized Open Field, Elevated Plus Maze, Light/Dark Box, and Novel Object avoidance tasks, and assessed fear learning and recall with the Conditioned Fear task. In addition to atypical sociability

and anxiety, WS results in a variety of sensory and motor symptoms. These characteristics were measured here by performance in a sensorimotor battery as well as Rotarod, Acoustic Startle/Pre-Pulse inhibition, and Marble Burying tasks.

We replicated and extended the previously reported social differences in the CD mice, showing enhanced social approach and motivation, in addition to sensorimotor differences and greater avoidance behavior in some anxiety-related tasks. An open-space aversion noticed in multiple paradigms was not apparent when a social stimulus was available. Finally, we rule out *Gtf2ird1* as being the sole mediator of the social changes, as duplication of this gene did not decrease these behaviors, nor did its complementation of the complete deletion rescue any notable social phenotypes. However, it does appear to mediate aspects of light-induced anxiety-related behaviors (Light/Dark Box) and sensorimotor coordination (Platform, Rotarod), as complementation can ameliorate the deficits observed in the CD mice, suggesting a role for *Gtf2ird1* in sensorimotor processing.

3.3 Results

3.3.1 A novel overexpression mouse rescues *Gtf2ird1* expression in the context of a complete deletion of the syntenic Williams Syndrome Critical Region

Evidence from atypical deletions show the telomeric end of the Williams Syndrome Critical Region (WSCR) as important for most of the key WS features. Two specific genes, *Gtf2i* and *Gtf2ird1*, within the telomeric end are suspected to play important roles in the cognitive and behavioral profiles of WS. While a *Gtf2i* mouse line has been developed, no such line exists for *Gtf2ird1*. Here we fill that gap with a novel transgenic line that expresses *Gtf2ird1* and test the

hypothesis that *Gtf2ird1* is critical for features of WS by rescuing its expression on the most relevant background, the complete deletion of the WSCR.

To determine the role *Gtf2ird1* plays in the WS behavioral repertoire, we first generated and validated a novel mouse overexpressing the general transcription factor GTF2IRD1 (TG-Gtf2ird1-HA) via its endogenous regulatory elements, engineered using a bacterial artificial chromosome with an HA tag that was inserted just prior to a stop codon of *Gtf2ird1* (**Fig 1A**). The line was validated through qPCR and Western blot analysis of heterozygous animals which reveal increased production of *Gtf2ird1* RNA (**Fig 1B**; $t=-5.247$, $p<0.001$) and protein (**Fig 1C**; $t=-1.991$, $p=0.048$, one-tailed).

Next, we demonstrated the ability of the TG-Gtf2ird1-HA mouse to rescue *Gtf2ird1* expression in the Complete Deletion (CD) mouse, a line that effectively deletes the syntenic Williams Syndrome Critical Region,[159] by crossing heterozygous TG-Gtf2ird1-HA and CD animals to produce four distinct progeny: wildtype (WT), TG-Gtf2ird1-HA (TG), Complete Deletion (CD), and the putative rescue (TG/CD), which combines the transgene and the complete deletion (**Fig 1D**). Molecular validation via qPCR confirmed *Gtf2ird1* overexpression in TG heterozygotes and decreased expression in CD animals relative to WTs, while RNA expression in the TG/CD group was not significantly different from WT, indicative of a molecular rescue (**Fig 1E**; $F(3,20)=22.190$, $p<0.001$). To ensure altered expression was specific to *Gtf2ird1*, we also measured relative expression of the nearby related gene, *Gtf2i*. The overexpression of *Gtf2ird1* did not significantly alter RNA expression of *Gtf2i*, which was significantly lower on the CD background regardless of transgene presence (**Fig 1F**; $F(3,20)=12.818$, $p<0.001$). To confirm protein expression was also affected, we ran a Western blot, probing for GTF2IRD1 using GAPDH as a control (**Fig 1G**). CD GTF2IRD1 protein levels were significantly lower than WT, TG, and

TG/CD GTF2IRD1 levels, which did not differ from each other significantly (**Fig 1H**; F(3,8)=9.918, p=0.005).

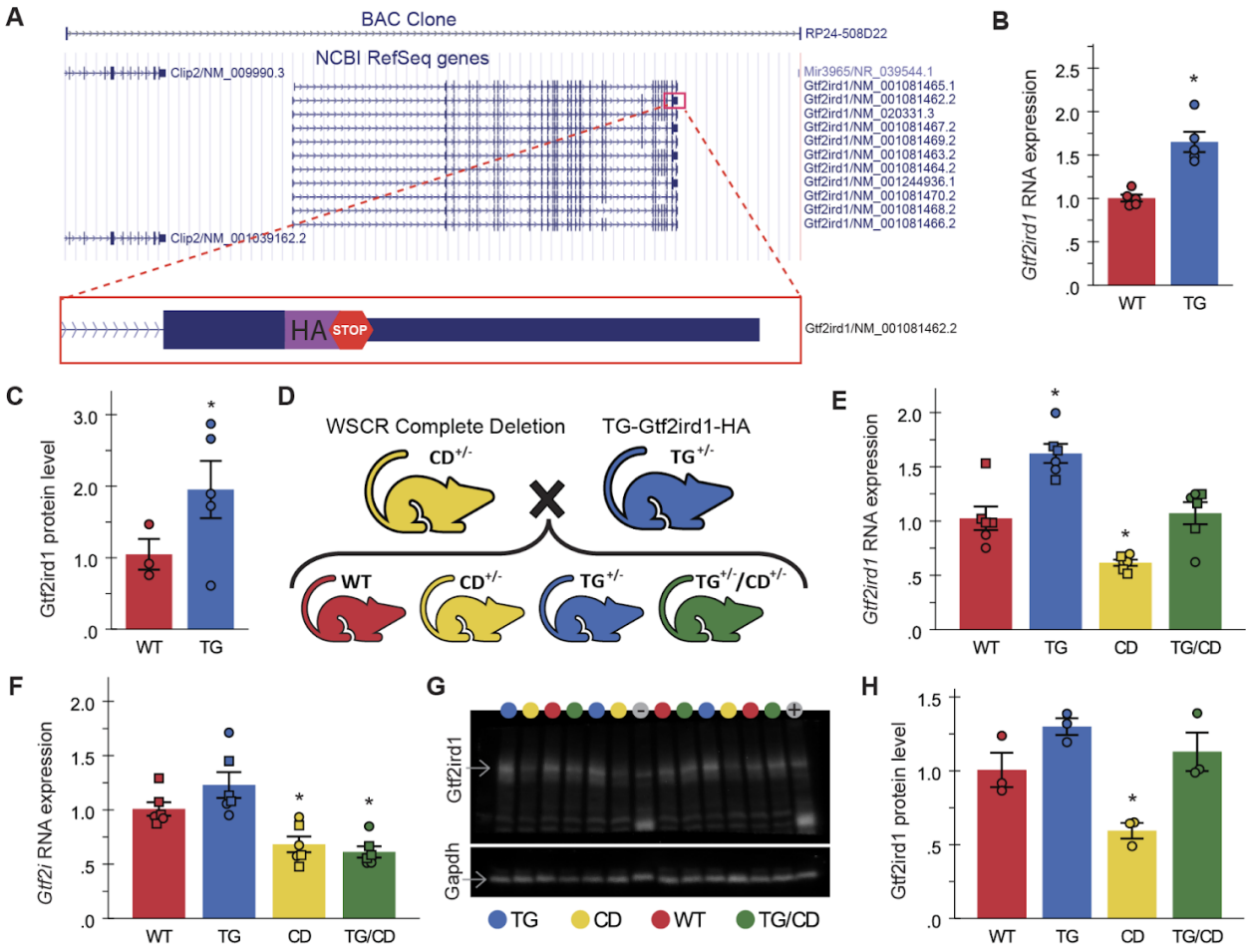

Figure 1. Novel *Gtf2ird1* overexpression mouse rescues *Gtf2ird1* RNA and protein levels in a Complete Deletion mouse modeling deletion of the syntenic Williams Syndrome Critical Region. A) Location of the BAC clone used to create the TG-Gtf2ird1-HA mouse line along with a cartoon of the HA-tag added just prior to the stop codon of one of the *Gtf2ird1* isoforms. **B)** RNA expression of *Gtf2ird1* relative to *Gapdh* in WT and TG littermates. **C)** Relative GTF2IRD1 protein levels in WT and TG littermates, n=3 WT, 5 TG. **D)** Heterozygous (+/-) CD and TG animals were crossed to directly compare WT, TG, CD, and TG/CD progeny. **E)** *Gtf2ird1* RNA expression in progeny from cross outlined in panel D, n=6 per genotype. **F)** *Gtf2i* RNA expression from the same animals as in E. **G)** Western blot of TG x CD progeny probed with antibodies for GTF2IRD1 and GAPDH, colored circles above the lanes indicate genotype, - and + represent negative and positive controls for the transgene. **H)** GTF2IRD1 protein levels quantified from the blot in panel G, n=3 per genotype. All RNA and protein levels were normalized to *Gapdh* expression. For E and F only, square = male, circle = female; * = p<0.05.

Having thus validated the expression of the *Gt2ird1* allele and complementation of *Gtf2ird1* levels in the CD background, we then utilized this same breeding scheme to generate a set of litter-matched behavioral cohorts for comprehensive behavioral testing (**Table 1**, see

Methods), enabling a study of main effects of each allele, as well as detection of interactions. We likewise included sex in all subsequent analyses, and report sex effects when significant.

3.3.2 *Gtf2ird1* restoration ameliorates select sensorimotor coordination deficits in Complete Deletion mice

Both WS and Dup7 are associated with strength deficits and motor delays. While G*tf2ird1* has been connected to the WS craniofacial phenotype and is suspected to play a role the unique cognitive profile (which includes visuospatial processing deficits) and behavioral features of WS, its role in sensorimotor features of WS has not been thoroughly defined. To address both the impact of *Gtf2ird1* on these features and the complete WS deletion in CD mice, we devised a comprehensive assessment of sensorimotor abilities, which also provided information necessary to properly interpret tasks relying on adequate motor performance. The wide-ranging compilation of tasks addressed a variety of basic motor abilities and more complex tasks requiring integration of sensory information (in mice, coordinated movement often is informed by their whiskers, rather than their eyes).[204] We split relevant tasks across two cohorts; in one cohort (**Fig 2A**, *above midline*), we tested activity in over 1-hour in an open field apparatus and natural digging behaviors observed in the Marble Burying task. Animals in the other cohort (**Fig 2A**, *below midline*) were tested using Rotarod, Pre-Pulse inhibition, and a sensorimotor battery, which included Walk, Inverted Screen, Pole, Platform, and Ledge tasks to assess a variety of movement related abilities, such as motor initiation, strength, coordination, and balance, as well as sensory processing.

At P30, results were not significant across a 1-hour open field activity task (**Fig 2B,C**). At P60, motor initiation, coordination, balance and strength were tested using a battery of sensorimotor tasks. There was no difference in motor initiation in a Walk task (**Fig 2D,E**), though

differences in balance and strength were observed. Specifically, the CD animals were unable to hold on to an inverted screen as long (**Fig 2F,G**; $H(3)=30.208$, $p<0.001$) but climbed down the pole faster than WT animals (**Fig 2H,I**; $H(3)=16.709$, $p<0.001$). A balance deficit was observed in CD animals as fewer animals in this group were able to remain on a thin, acrylic ledge for a full minute (**Fig 2J**; $H(3)=29.487$, $p<0.001$) or on a small platform just large enough for the mice to stand atop (**Fig 2K,L**; $H(3)=16.919$, $p<0.001$). Interestingly, rescue of *Gtf2ird1* partially restored performance on the Platform task, as TG/CD animals stayed on the platform significantly longer than their CD counterparts ($p=0.046$). To examine motor coordination more directly, we used the Rotarod task (**Fig 2M**), which revealed another partial rescue (**Fig 2N**; CD*TG interaction: $F(1,69)=6.977$, $p=0.01$). While all mice learned the task and generally improved over subsequent trials, CD and CD/TG animals had a shorter latency to fall relative to WT and TG animals ($F(1,69)=35.227$, $p<0.001$). The interaction between CD and TG alleles (e.g., rescue) was most apparent in females (**Fig 2O**; Sex*CD*TG: $F(1,69)=4.461$, $p=0.038$).

In the Marble Burying task, both CD and CD/TG mice buried far fewer marbles than the WT and TG groups (**Fig 3P,Q**; $H(3)=34.458$, $p<0.001$). This finding is confounded by a matching decrease in total distance traveled, represented here by the lower number of center entries by the CD groups (**Fig 2R;** $F(1,90)=76.712$, $p<0.001$). Thus, the fewer marbles buried may simply be a factor of hypoactivity, though what is causing the hypoactivity here and not in the open field task is not known. The effect is perhaps enhanced by the novel bedding used in the Marble Burying task, which was not present in the other apparatus.

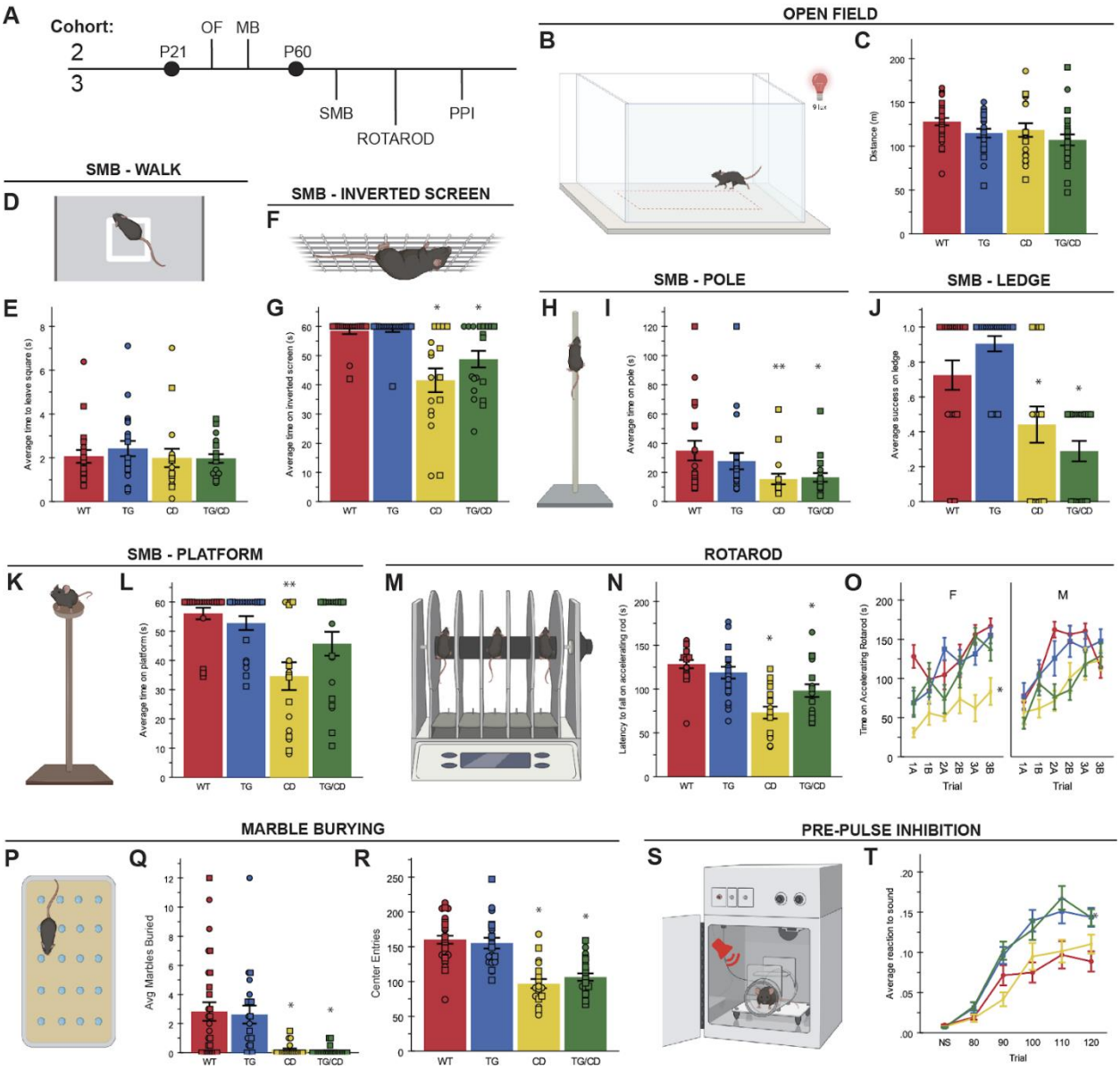

Figure 2. *Gtf2ird1* restoration affects a subset of sensorimotor deficits observed in the CD mice. A) Tasks were split between two cohorts: OF = Open Field, MB = Marble Burying, SMB = sensorimotor battery, PPI = Pre-Pulse Inhibition. **B)** A 50x50cm arena under red illumination at 9 lux was used for the 1-hour open field task. **C)** No significant differences were observed between groups in total distance travelled in the open field task. **D)** In the "walk" task, the time for mice to exit the white square in the center of a large open space was measured as a proxy for motor initiation. **E)** No differences in motor initiation were observed. **F)** The inverted screen task measures how long a mouse can hold on for up to 60 seconds **G)** CD and TG/CD mice were not able to hold on to the screen for the full minute. **H)** Mice were placed on a textured pole for up to 120 seconds. **I)** CD and TG/CD mice were significantly faster to leave the pole. **J)** CD and TG/CD mice were unable to stay on an acrylic ledge for a full minute. **K)** Time on a small platform was measured. **L)** CD animals had a decreased latency to fall that was ameliorated with the presence of the transgene. **M)** The Rotarod apparatus used to measure ability to stay on a moving rod. **N)** Animals with the CD allele fell off an accelerating rod faster than WT, but presence of the transgene improved the outcome. **O)** This partial rescue was especially clear in females. **P)** Mice have 30 minutes to explore a chamber with 20 evenly spaced marbles. **Q)** Animals with the CD allele (CD and TG/CD animals) buried significantly fewer marbles than WT and TG animals. **R)** CD and TG/CD animals travel less distance overall, reflected here in significantly fewer center entries. **S)** In the Acoustic Startle/Pre-Pulse Inhibition paradigm, mice are exposed to an acoustic stimulus while confined within a sound-attenuated box on a force plate to measure the startle reflex. **T)**

Transgenic expression of *Gtf2ird1* resulted in a greater startle response (in Newtons) across various sound levels; each trial reflects sound level in decibels, NS = no sound. * $p<0.05$, ** $p<0.005$ relative to WT.

Sensory sensitivity is another feature of WS that warrants investigation, as WS individuals are more reactive to sounds.[35,37] In the Acoustic Startle/Pre-Pulse Inhibition (PPI) task (**Fig 2S**), animals are presented with acoustic stimuli designed to induce the startle response in mice. The CD allele alone did not influence response to an acoustic startle stimulus, but mice harboring the TG allele responded with greater startle magnitude force to all levels of sound (**Fig 2T**; $F(1,45)=12.898$, $p<0.001$). These data may indicate a unique feature of unbalanced *Gtf2ird1* expression relative to the rest of the WSCR. There was also an interaction between CD and TG alleles and sex (*not shown*; Sex*CD*TG: $F(1,45)=6.295$, $p=0.016$), suggesting a possible difference in sensitivity to genotype based on sex. The PPI trials revealed an interaction between CD and TG alleles on sensorigating ability, with the TG/CD animal showing lower percent inhibition than the other groups (*not shown*; CD*TG: $F(1,45)=5.227$, $p=0.027$).

3.3.3 Restoring *Gtf2ird1* expression in CD mice rescues light-avoidant but not center-avoidant anxiety-like behaviors

Anxiety is another feature common to WS and Dup7, though the specific forms differ. Non-social anxiety and increased prevalence of phobias are over-represented in the WS population, while Dup7 is characterized by greater social anxiety and separation anxiety, with no clear phenotype related to fear. As there are no specialized treatments for these symptoms among patients, having a well characterized model for preclinical screening of therapeutics may eventually lead to better care. Thus, we thoroughly assessed non-social anxiety-like avoidance features in the CD mouse model to identify tasks sensitive to this mutation and evaluate the potential impact of *Gtf2ird1*.

Anxiety-like behavior is measured in rodents by quantifying passive avoidance behavior in low-threat situations in which perceived danger is diffuse and uncertain (PMID: 33005134). These situations for a rodent include open spaces and brightly lit spaces. One common trigger for passive avoidance behavior in rodents is the center space of an open field (**Fig 2B**). Similarly, the Elevated Plus Maze (EPM) measures an animal's passive avoidance of the open arms. In contrast, light and open space is leveraged together in the Light/Dark Box task, and a decrease in time in the light side of the box has been used to indicate passive avoidance behavior, rather than relying on avoidance of the open space alone. Together, these tasks should inform us of the anxiety-like, passive avoidance features of the CD mice and whether *Gtf2ird1* expression has any effect on these behaviors.

In the 1-hr Open Field task conducted under red light at 9 lux, both CD and TG alleles resulted in a decrease in center time (**Fig 3A,C**, CD: $F(1,86)=13.175$, $p<0.001$; TG: $F(1,86)=7.935$, $p=0.006$). As the overall distance traveled was not different between groups (**Fig 2C**), these results are consistent with heightened anxiety-like behavior. Regardless of genotype, females spent less time in the center than their male counterparts (**Fig 3B**, $F(1,86)=12.210$, $p<0.001$). In contrast to the avoidance behaviors observed in the open field, there were no observed differences in the percent time spent in the open arms of the EPM under complete darkness (**Fig 3D,E**).

Interestingly, during the Light/Dark Box task (**Fig 3F**), we observed a significant interaction of CD and TG alleles on the percent time spent in the light (**Fig 3G**; $F(1,73)=7.460$, $p=0.008$). CD animals spent significantly less time in the light relative to their WT ($p=0.009$) and TG/CD ($p=0.018$) counterparts, while TG/CD animals were not significantly different from the WT group, reflective of the TG allele rescuing CD deficits in this task. Thus, *Gtf2ird1* complements the CD mutation for this phenotype.

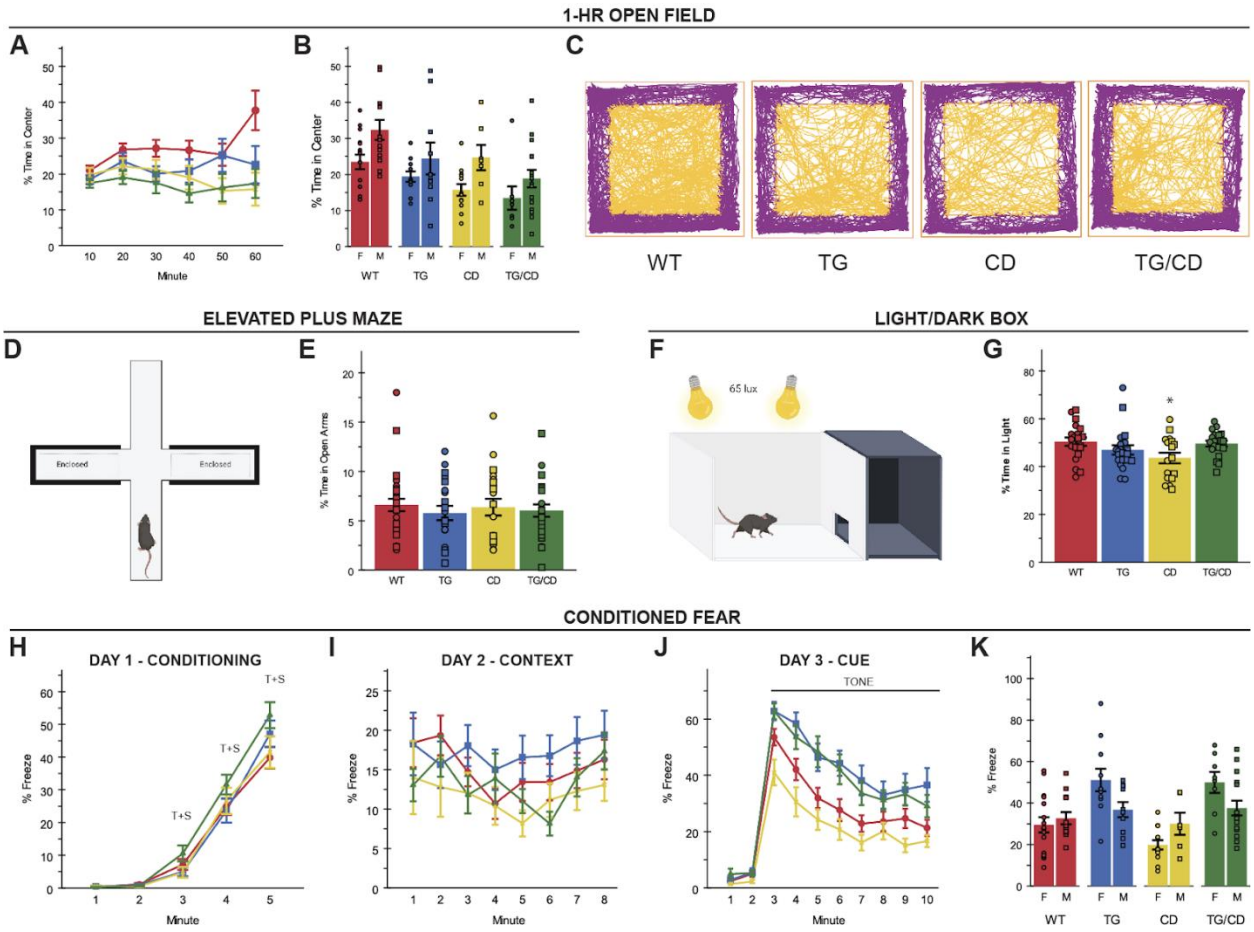

Figure 3. ***Gtf2ird1*** **corrects CD-induced decreased time in light but not in center space. A)** Transgene and CD allele decrease time spent in center of an open field. **B)** Females consistently spend less time in the center compared to males. **C)** Representative track plots of open field task, chosen based on group mean. **D)** Diagram of elevated plus maze apparatus. **E)** No difference in percent time spent in open arms of the EPM was observed. **F)** Diagram of the Light/Dark Box task. **G)** Decreased time in light side caused by CD allele is rescued with expression the *Gtf2ird1* transgene. **H)** No significant differences in freezing were observed during the training day of the Conditioned Fear task. **I)** Day 2 of the Conditioned Fear task also revealed no significant differences. **J)** TG allele increases percent freezing during cued recall in the 3rd day of Conditioned Fear. **K)** The TG affect on percent time freezing during cued recall is greater in female mice. T+S = Tone + Shock

In contrast to the passive avoidance of anxiety-like measures, fear responses, which have a component of anxiety, are active avoidance behaviors quantified in situations where a threat is imminent and well-defined (PMID: 33005134). We used the fear conditioning task to further evaluate active avoidance and associative memory by quantifying freezing behavior in response to a shock paired with a novel auditory cue and spatial context. No differences in freezing response to the pairing of the shock and tone+context were observed between genotypes on Day 1 (**Fig 3H**).

However, females froze more than males overall (Day 1, min 3-5; $F(1,85)=5.606$, $p=0.02$). This sex effect was also observed during contextual fear recall on Day 2 when mice were re-exposed to the spatial context to test hippocampal-dependent spatial conditioning (Fig **3I**; $F(1,85)=5.650$, $p=0.02$). The CD mice also showed reduced freezing, similar to our previous reports, but did not pass the significance threshold due to low power.

During amygdalar-dependent cued fear recall on Day 3, animals with the TG allele overexpressing *Gtf2ird1* showed increased freezing in response to the auditory cue (**Fig 3J**; $F(1,85)=29.860$, $p<0.001$). This increased freezing was especially pronounced in females with the TG allele (**Fig 3K**; $F(1,85)=11.861$, $p<0.001$). Mice with only the CD allele show reduced freezing behavior compared to all other groups, replicating our previous effect,[115] although the comparison to WT mice did not quite pass the significance threshold ($p=.078$), likely due to power here. Shock sensitivity was comparable across groups.

3.3.4 Enhanced social approach and motivation is independent of *Gtf2ird1*

Finally, given the interesting contrasting social motivation phenotypes in WS and Dup7 patients,[18,47] we conducted a comprehensive phenotyping of social behavior in our cohorts. To identify early signs of social behavior changes, we first assessed pup social communication via a maternal isolation-induced ultrasonic vocalization (USV) paradigm (**Fig 4A,B**). Given elevated aggression in Dup7 patients,[55] we included standard measures of social dominance (tube test) and aggression (resident intruder). Sociability differences in the CD model was originally identified in a modified single chamber version of social approach (Open Field Social Approach), rather than the typical 3-chamber task widely used in ASD models.[120,159] Previous work in our lab failed to identify differences in the 3-chamber task alone, though on a C57BL/6J x FVB hybrid background

that shows lower social approach in general.[117,124] Thus, our comprehensive battery here included a deliberate precise replication of the Open Field Social Approach conditions as a baseline control,[134,159] the standard 3-chamber social approach assay,[120,205] and finally a 14-day social operant task we recently designed to be a direct measure of social motivation in rodents.[206,207]

Early differences in communication were evident across the three days of the USV task. CD animals demonstrate a delay in USV production and TG/CD animals show sustained deficits in the number of calls made (**Fig 4C**). These differences do not appear to be due to gross developmental delays, as weights were comparable between CD and TG/CD animals (**Fig 4D**). Also, at P14, all pups had the ability to flip themselves upright, with no difference in time to right (**Fig 4E**). This suggests there is an early disruption to social communicative behavior with haploinsufficiency for the WSCR. In adults, previous research showed decreased dominance in the Tube Test and aggression in the Resident Intruder paradigms. Though CD animals seem to win less, there was no statistically significant difference in Day 1 of the Tube Test paradigm (**Fig 4F**; $H(7)=11.102$, $p=0.134$). In addition, there were also no significant differences in attacks (**Fig 4G**; CD: $F(1,32)=1.167$, $p=0.288$). However, the total numbers of wins and attacks, respectively, were low in both paradigms regardless of group, which may have had an impact.

Similar to Segura-Puimedon *et al.* (2014), in the Open Field Social Approach task we found CD animals spent more time investigating the social stimulus mouse relative to WT animals ($p=0.013$).[159] In fact, we observed increased social approach in all groups relative to WT levels (**Fig 4H**, $H(3)=8.916$, $p=0.03$). The significance of this difference appeared to be driven by the lack of habituation, as WT levels of approach fell after 10 minutes while the other groups remained more interested in the social stimulus (**Fig 4I**, $H(3)=13.160$, $p=0.004$).

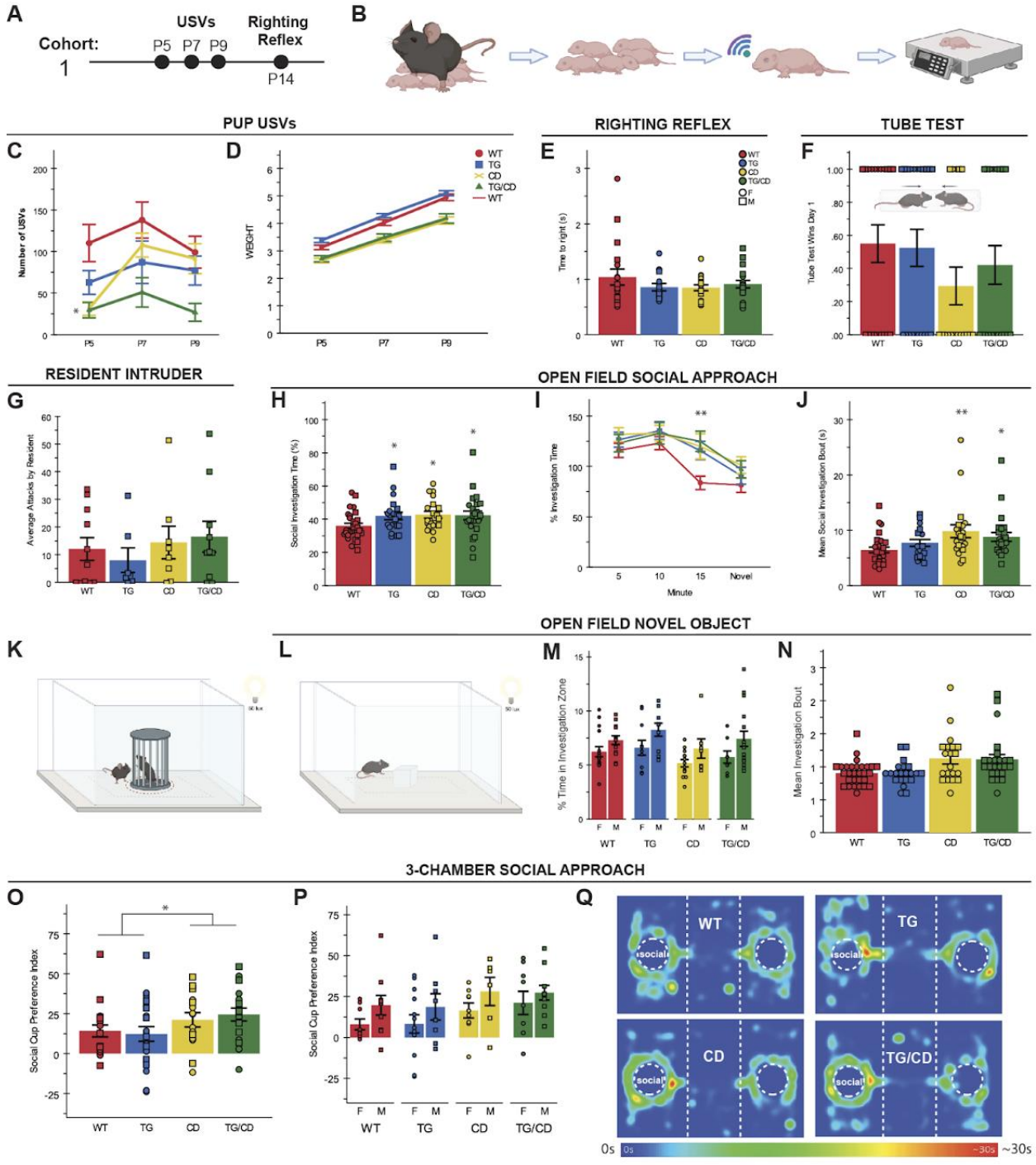
A
Cohort:
1
USVs
P5 P7 P9
Righting Reflex
P14
B
PUP USVs
RIGHTING REFLEX
TUBE TEST
C
Number of USVs
D
WEIGHT
E
Time to right (s)
F
Tube Test Wins Day 1
WT
TG
CD
TG/CD
RESIDENT INTRUDER
OPEN FIELD SOCIAL APPROACH
G
Average Attacks by Resident
H
Social Investigation Time (%)
I
% Investigation Time
Minute
Novel
J
Mean Social Investigation Bout (s)
OPEN FIELD NOVEL OBJECT
K
L
M
% Time in Investigation Zone
N
Mean Investigation Bout
3-CHAMBER SOCIAL APPROACH
O
Social Cup Preference Index
P
Social Cup Preference Index
Q
WT
TG
CD
TG/CD
social
0s
~30s
SOCIAL OPERANT
R
S
FR1 Mean Total Rewards
T
FR3 Mean Total Rewards
U
PR Breakpoint

Figure 4. Increased social approach and motivation in CD model of WS deletion is independent of *Gtf2ird1*. A) Cohort 1 was used to test USVs over postnatal (P) days 5, 7, and 9 and righting reflex at P14. **B)** Maternal separation induced pup USV workflow: Remove dam, place pups in warming chamber without moving them, measure their temperature while in the nest, individually records each pup's USVs, then weigh. **C)** Calling rate was altered in mice with CD alleles, though TG/CD mice sustained the deficit over all three days. **D)** Pup weight after USV recordings show decreased weight in pups with the CD allele. **E)** No differences observed in righting at P14. **F)** No significant differences measured in the Tube Test for Social Dominance task. **G)** No differences in average number of attacks by resident on intruder. **H)** TG, CD, and TG/CD animals spent more time investigating a social stimulus. **I)** WT show habituation at 15 minutes that is not seen in the other groups. **J)** Only mice with the CD allele show longer mean bouts of investigation. **K)** Representation of the Open Field Social Approach apparatus. **L)** Open Field Novel Object apparatus. **M)** Novel object avoidance seen in CD and TG/CD animals. **N)** CD allele caused increased mean investigation bout in the novel object task. **O)** Social preference index in the 3-chamber social approach task was greater in animals harboring the CD allele. **P)** Male animals had higher social preference for the social cup than females. **Q)** Representative heat maps of 3-chamber social approach. **R)** Social Motivation Operant apparatus. **S)** CD allele results in greater mean rewards during FR1. **T)** Males with the CD allele had increased mean rewards during FR3. **U)** The breakpoint during PR3 was greater in CD and TG/CD animals. circle = female, square = male, * <0.05, ** < 0.005 relative to WT group.

While all groups showed sustained interest in the stimulus, the mean social investigation bout was only higher in CD and TG/CD animals (**Fig 4J**, $H(3)=13.16$, $p=0.006$). Regardless, it is clear that the deletion of the WSCR increases social approach behaviors here as measured in investigation time and average investigation bout.

To control for the potential impact of novelty on the Open Field Social Approach task (**Fig 4K**), we next tested the reaction to another novel object, a translucent square placed in the center of the familiar open field arena (**Fig 4L**). CD and CD/TG animals spent significantly less time near the novel object (**Fig 3M**; $F(1,86)=4.203$, $p=0.043$). In addition to the effect of the CD allele, females of all groups spent less time investigating the novel object as well ($F(1,86)=11.901$, $p<0.001$). Though less time was spent investigating the object, mean investigation bout was higher in animals with the CD deletion, just as it was in the Open Field Social Approach task (**Fig 3N**; $H(3)=11.271$, $p=0.01$).

The increased social approach also holds in the classic 3-Chamber Social Approach task. Deleting the WS region results in a greater preference for the social cup, compared to an empty cup (**Fig 4O,Q**; $F(1,71)=5.527$, $p=0.022$). This preference was especially pronounced in males across all genotypes (**Fig 4P**; $F(1,71)=6.061$, $p=0.016$). The effects on social novelty were not as straightforward, and no differences across groups were significant (*not shown*).

Finally, we applied our social operant task to precisely investigate social motivation (**Fig 4R**). Animals were trained to nosepoke to open a door on a to reveal a novel mouse as a social reward. During training, or fixed ratio 1 (FR1) one poke results in one reward. As animals reach criteria (sufficient total nosepokes, active:inactive ratio, and interactions per reward), they progress to a fixed ratio of 3 (FR3) for three days. After 3 days of FR3 (for learners) or 10 days at FR1 (nonlearners), each animal undergoes testing at a progressive ratio of 3 (PR3), where an animal's maximum motivation, or breakpoint, is assessed by continuously increasing the difficulty to attain each reward by 3 nosepokes. Learners with the complete deletion allele (CD and CD/TG genotypes), reached a higher number of total rewards during FR1 (**Fig 4S**; $F(1,47)=14.07$, $p<0.001$). During FR3, an interaction between sex and CD allele emerged, showing that only males with the CD allele reach more rewards (**Fig 4T**; CD*Sex: $F(1,44)=4.932$, $p=0.032$). Subsequently, the CD allele results in a higher breakpoint during PR3 (**Fig 4U**; $H(3)=8.092$, $p=0.044$). These results show that complete deletion of the WSCR in mice results in higher social motivation, meaning they are willing to work harder to keep accessing the social stimuli. While this appears to be driven by an increase in males, the limited number of animals who met criteria to be considered learners (and thus be included in our analysis) constrains our power to assess sex effects.

3.4 Discussion

In this extensive characterization of mouse models relevant to the WSCR, the consequences of complete deletion of the region were obvious. We found a widespread effect of WSCR deletion in the CD model, causing deficits in sensorimotor abilities (i.e., performance in Inverted Screen, Ledge, Platform, Rotarod, Marble Burying tasks), select anxiety-like behaviors

(i.e., time in Open Field center, and time in light side of the Light/Dark Box), and enhanced social interest (i.e., approach behaviors in open field and 3-chamber set-ups, and increased motivation in the social operant paradigm). We also presented a novel transgenic line expressing a gene of interest, *Gtf2ird1*, and used this line to genetically rescue *Gtf2ird1* expression in the CD line to examine its ability to rescue any of the atypical phenotypes presented. The transgenic *Gtf2ird1* line highlighted a role for *Gtf2ird1* in sensorimotor coordination (as evidenced in the Platform and Rotarod tasks) and potentially sensory processing more generally (sound sensitivity and potentially light). While *Gtf2ird1* affected a few features clearly, most features where not significantly impacted by its rescue or overexpression, suggesting either *Gtf2ird1* is not involved or is only part of the underlying etiology and would require prohibitively large samples to observe its small effect in those domains.

Interestingly, *Gtf2ird1* seems linked with a hyper-response to sound (in force produced during acoustic startle trials during PPI, and in increased freezing during contextual fear) is clearly correlated with expression of the transgene. This could mean one of two things: either the HA-tagged beta isoforms are altered in a way that specifically affects these phenotypes, or uneven *Gtf2ird1* expression relative to the other genes in the WSCR is causing dysregulation relevant to these phenotypes. Either way, we can derive useful information from the issue. If it is an issue of the HA tag, we learn that a specific subset of *Gtf2ird1's* isoforms is definitely involved in responsivity to sound, or if it's an issue of discordant expression within the WSCR, we learn there is definitely interplay between *Gtf2ird1* and at least one other gene in the region.

Beyond its usefulness here, the novel transgenic *Gtf2ird1* line may also provide opportunities to research into the various roles *Gtf2ird1* isoforms may play in typical development. *Gtf2ird1* is an extensively alternatively spliced gene, with numerous uncharacterized isoforms.

Our TG-Gtf2ird1-HA mouse tags less than half of the isoforms (only the beta variants that contain the full exon 30)[208]; leveraging this fact, we can investigate how these two groups of isoforms differ. Utilizing an HA antibody in a pull-down assay would effectively separate these isoform groups for downstream analysis to compare their functions, particularly in regard to genomic binding.

Synthesizing the results of this study beyond the contributions of *Gtf2ird1*, we show the Complete Deletion mouse may not be a straightforward model for anxiety-related phenotypes relevant to the Williams Syndrome deletion. Avoidance of the center in the Open Field task suggesting an anxiety-like phenotype, which was replicated in the decreased time in the light side of the Light/Dark Box, but no differences were apparent in the EPM. Whether the EPM is not sensitive to the particular anxiety-inducing features relevant to WS or whether the EPM requires specific parameters for peak performance needs to be determined. It is possible that the differences observed in the Open Field and EPM tasks (**Fig 3**) may be due to differences in the length the tasks (60 vs 5 min), the limited time animals spent in the EPM open arms at all (<10% of time on average), or the age of the animals in both tasks (not yet adults). Without answers to why a phenotype wasn't observed in the EPM, use of that task and interpretation of other anxiety-relevant phenotypes may need to be done cautiously.

In conclusion, while linking individual genes to specific features is difficult due to varied expressivity despite identical genetic lesions, the WS critical region (WSCR) still provides a unique genetic landscape to investigate genotype-phenotype connections and the pathological effect of copy number variations. Modeling single gene deletions of candidates within the WSCR limits our ability to address polygenic features, as they may not provide an accurate picture of the consequences (or possible functional restoration) within the context of the full deletion. The CD

mouse modeling the full WSCR deletion is the most relevant genetic context, providing excellent construct validity to study polygenic traits of WS and can be used to study a broad suite of characteristics, as shown here. Future studies focusing on mechanism discovery for behavioral symptoms could be a way to provide meaningful answers despite complicated etiology, as suggested by Kozel *et al.* (2021).[1]

3.5 Materials and Methods

All experimental protocols were approved by and performed in accordance with the relevant guidelines and regulations of the Institutional Animal Care and Use Committee of Washington University in St. Louis and were in compliance with US National Research Council's Guide for the Care and Use of Laboratory Animals, the US Public Health Service's Policy on Humane Care and Use of Laboratory Animals, and Guide for the Care and Use of Laboratory Animals.

3.5.1 Gtf2ird1 Transgenic Mouse Creation

We selected a bacterial artificial chromosome (BAC) clone (RP24-508D22) which contained the entirety of the 100 kb *Gtf2ird1* gene, and 89 kb (60 kb at the 5' end and 28 kb at the 3') of flanking regulatory sequence (e.g., the *Gtf2ird1* promoter, etc.), but no additional intact genes or their promoters. This was then recombineered using standard methods to insert an HA tag in-frame directly before the stop codon of the beta isoforms.[208,209] Specifically, we used homologous recombination via transient expression of RecA, followed by Neomycin selection of an inserted FRT flanked cassette. The selection cassette was then removed after the expression of FLPe, leaving on a single FRT site downstream of the stop codon. Transgenic mice (TG)

overexpressing *Gtf2ird1* (TG-Gtf2ird1-HA) were created by injecting this modified BAC into C57BL/6NTac mouse oocytes and transplanting these eggs into pseudo pregnant surrogates to carry them to term. Transgene-specific primers (BgenoF3 – CAACATTCCCAAGCGCAAGAG and BgenoR3 – GATAACTGATCGCGGCCAGC, which produce a 440 bp product in TG animals and no product in WT animals) were used for identification of TG founder animals. BAC copy number was determined to be 2-4. Multiple founders were evaluated to confirm transgenic RNA production by RT-PCR, and a single line was taken forward for evaluation. Lines were backcrossed to C57BL6/J for over 4 generations prior to commencing experiments.

3.5.2 Husbandry

All mice used in this study were maintained and bred in the vivarium at Washington University in St. Louis on a 12/12 hour light/dark cycle with food and water provided freely. Three distinct mouse lines were used: C57BL/6J (WT, RRID:IMSR_JAX:000664), the Complete Deletion (CD) mouse modeling deletion of the Williams Syndrome critical region,[159] and a novel transgenic line (TG) overexpressing *Gtf2ird1* with an HA tag. CD and TG lines were maintained as heterozygotes by crossing to WT animals. Heterozygous CD and TG mice were crossed to produce behavioral cohorts containing WT, TG, CD, and TG/CD littermates to best compare across genotypes. Animals were housed by genotype and sex at weaning. Tissue was collected from pups for first genotyping and after death to verify genotype via PCR amplification.

3.5.3 Molecular Validation

Molecular analysis to assess RNA and protein levels via RT-qPCR and Western blotting was performed as previously described.[139] Brains were collected from pups ~E13.5 for initial

characterization of the novel line and just prior to postnatal day 21 (P21) for validation of the crosses. RNA expression and protein levels were assessed relative to *Gapdh* using primers and antibodies described previously.[139] Expression of the transgenic allele was verified by a Western blot using an antibody to the HA tag, which is included only on some transcripts due to alternative splicing of the last exon.

3.5.4 Behavioral Testing

For behavioral analysis, three separate cohorts of mice were used to assess a variety of characteristics with the fewest number of animals possible (**Table 1**). All tasks were run by female experimenters. Tasks within each behavioral battery were ordered from least to most stressful to minimize the effect one task had on subsequent tasks. Adolescent and adult mice were handled for 5 days prior to starting the first behavioral task and mice in Cohorts 2 and 3 were marked with a non-toxic, permanent marker during weight collection to easily distinguish them during testing. Males were run before female animals. Tail samples were collected post-mortem to verify genotypes.

Table 1. Behavioral cohort sample size and task order.

Cohort	***1 - 031120***	***2 - 020420***	***3 - 021521***
Max n	91	**94**: 48F, 46M	**77**: 40F, 37M
Task 0	Temp/Weight (P5,7,9)	Weight (P26-35)	Weight (P50-86)
Task 1	USVs (P5,7,9)	Open Field (P27-36)	Sensorimotor Battery (≥P56)
Task 2	Righting Reflex (P14)	Social Approach (P28-39)	Rotarod
Task 3		Marble Burying (P30-40)	Light/Dark Box
Task 4		Elevated Plus Maze (P33-45)	3-Chamber Social Approach
Task 5		Novel Object (P38-47)	Tube Test
Task 6		Social Operant (P40-64;90-112)	Pre-Pulse Inhibition
Task 7		Conditioned Fear (P59-70;119-128)	Resident Intruder (P88-123)

COHORT 1

Maternal separation induced ultrasonic vocalizations and righting reflex

Briefly, pups were tested in their colony room by the same female experimenter (REW) on postnatal (P) days 5, 7, and 9. Mice were identified and genotyped by toe clipping, which was performed after the P5 recording session. All recordings occurred after 12 PM CST between March and September of the same year. Prior to recording, the parents were removed, and pups in the nest were placed in a warming chamber at 33°C without removing them from their nest to maintain a surface body temperature of 31.1–37.5°C. After 10 minutes to acclimate body temperature, each pup was placed in an empty cage in a sound attenuating box (36x64x60 cm) and recorded for 3 minutes. The Avisoft UltraSoundGate CM16 microphone was positioned 5 cm from the top of the cage and an Avisoft UltraSoundGate 116H amplifier (gain = 8, 16 bits, sample rate = 250 kHz) was used for all measurements. Ultrasonic vocalizations (USVs) were recorded using the Avisoft-RECORDER software. Raw WAV files were processed using a custom MATLAB pipeline to extract call numbers and spectral and temporal call features.[206,210]

In addition to USVs, weight and temperature were recorded for each mouse at each time point. A non-contact HDE Infrared Thermometer was used to take the temperature of each mouse before they were removed from the nest for USV recording. Mice were weighed after recording. Pinnae detachment was also assessed at P5, and eye opening was documented at P14. At P14, the righting reflex was evaluated for each mouse by measuring the time for pup to right itself after being held on its back for 5 seconds. Three trials, limited to 1 minute, were performed for each mouse, averaged for analysis, and direction of righting was noted.

COHORT 2

All tasks performed on the 3 batches of cohort 2 took place within a sound- and scent-attenuated white opaque box (70.5 x 50.5 x 60 cm) to minimize external stimuli. Males were run

prior to females when possible, and a run order was created to counterbalance groups across apparatuses and runs. Between trials, 70% ethanol was used to clean glass marbles and wire cups and 0.2% Nolvasan was used for everything else.

Open Field

Mice were placed in a 50 x 50 x 45 cm plexiglass enclosure under red light at 9 lux and allowed to explore for 60 minutes, adapted from our previously published methods.[206] Any-Maze software (Stoelting, Co) tracked the movement via the body center, beginning when the doors to the chamber were closed via video captured with an overhead CCTV camera. A center zone was designated as the middle 50% of the total chamber area.

Open Field Social Approach

Replicating previously published methods,[134,159] we examined social approach behavior in an open field setting, under white light at 50 lux. In the center of the Open Field enclosure, a novel, sex- and age-matched stimulus animal was placed under a wire pencil cup (with a clear plastic cup on top to prevent climbing). The experimental animal was then added to the chamber and allowed to explore and interact with the social stimulus. After 15 minutes, recording was stopped, the experimental animal was removed to a clean holding chamber, and the stimulus animal was switched out with a novel mouse. The experimental animal was returned to the chamber for 5 additional minutes of recorded exploration. Any-Maze video tracking was used for both trials, and measured time spent in an investigation zone defined as 2cm out from the circumference of the center cup.

Elevated Plus Maze

Anxiety-like behaviors were tested as previously described.[117] Briefly mice were placed in the center of the apparatus, which contained two open and two closed arms, and allowed to explore for 5 minutes in the dark. This was repeated for two more days. Trials were recorded under red light illumination with an overhead camera using Ethovision software (Noldus Information Technology) to track movement.

Marble Burying

Mice were introduced to a novel, transparent enclosure (47.6x25.4x20.6 cm) with 20 evenly spaced, clear marbles on clean, novel, autoclaved aspen bedding. Animals were allowed to explore freely for 30 minutes and were tracked via Any-Maze software. After the animals were removed, two independent scorers recorded the number of marbles buried (defined as at least 75% covered with bedding). These scores were averaged for analysis.

Open Field Novel Object Exploration

Adapting previously published methods,[159] we examined novel object exploration in an open field setting to control for potential novel effects in the Open Field Social Approach task. A translucent cube was placed in the center of the same Open Field chamber used for Open Field, and Open Field Social Approach, also under white light at 50 lux. Mice were placed in the apparatus to explore freely for 20 minutes while movement was recorded and tracked using Any-Maze software. A 2-cm investigation zone was defined around the object, in addition to center and perimeter areas.

Social Motivation Operant Conditioning

Based on previously published methods,[206] social motivation –how hard an animal will work for access to a social partner– was assessed using our social motivation operant assay. Our 16-day paradigm allows for assessment of both social reward seeking and social orienting, two components of social motivation.[211] An operant conditioning chamber was modified to include a door that raised in response to a nosepoke in the active hole to provide 12 seconds of access to a novel sex- and age-matched partner stimulus mouse. To assess social reward seeking, the number of active (i.e., elicits a reward) versus inactive nosepokes were quantified. To assess social orienting, the behavior of the animal was tracked using Ethovision (Noldus) and number of interactions with the stimulus mouse and time spent near the stimulus mouse were quantified. Following two days of habituation (door remained open and the nosepoke holes were not accessible), mice received at least 3 days of fixed ratio 1 (FR1) conditioning, where 1 nosepoke in the active hole resulted in a reward. Mice that had at least 40 active nosepokes, 75% accuracy (active:inactive), and 65% interactions during rewards were considered to have met conditioning criteria and progressed to a fixed ratio of 3 (FR3), where 3 nosepokes were required to receive the reward. After 3 days of FR3 (or 10 days of FR1 for mice who failed to reach criteria), mice were tested in a progressive ratio of 3 (PR3), where the first reward was provided after 3 active nosepokes and each subsequent reward required 3 additional nosepokes to obtain. The breakpoint was measured as the number of rewards a mouse was able to acquire before 30 minutes of nosepoke inactivity.

Conditioned Fear

To assess associative and anxiety-related memory, mice were tested in a Conditioned Fear task over 3 days following previously published methods.[139,206,212] Shock sensitivity was evaluated as we previously described.[139]

COHORT 3

Sensorimotor Battery

Adult mice were evaluated with a battery of sensorimotor measures to assess motor initiation, balance, strength, and coordination using previously published methods.[139,206] The battery included walk evaluation of walk initiation, balance (Ledge and Platform tests), fine motor coordination (Pole test), and strength with coordination (Inclined and Inverted Screen tests).

Rotarod

Motor coordination was assessed using the Rotarod following our previously published methods.[212] Briefly, latency to fall was measured for each mouse in three different situations: a stationary rod (for up to 60 sec), a continuously rotating rod (3.0 rpms; for up to 60 sec), and an accelerating rod (3.0-17 rpms; for up to 180 seconds).

Light/Dark Box

The Light/Dark Box was used to assess anxiety-related avoidance behavior leveraging the mouse's innate preference for dark spaces. Mice were placed in the dark side of a chamber (47.6 x 25.4 x 20.6 cm) and were allowed to explore freely. The light side, which was twice as large as the dark side, was illuminated at 65 lux with incandescent desk lamps. Beam brakes were used to measure time spent in each chamber during the 20-minute task. Time spent in the light side was

used as a proxy for anxiety-like behavior, with more anxious-like mice avoiding the brightly lit open space.

3-Chamber Social Approach

Sociability and preference for social novelty were examined in the Social Approach task, following our previously published methods.[206] The mice received two, 10-min habituation trials: first to the center chamber of the apparatus and then to the entire chamber including the empty social investigation cups. Next, sociability was assessed for 10 min during which a novel age- and sex-matched conspecific was placed under one cup (the side used was counterbalanced across groups). During the fourth 10-min trial, a second, age- and sex-matched novel conspecific was placed under the other cup to assess preference for social novelty. The time spent investigating and number of investigations for each investigation cup, as well as time in and entries into each chamber and total distance traveled, was quantified using Any-Maze video tracking software.

Tube Test of Social Dominance

As social creatures, mice create social hierarchies within their social groups. Thus, laboratory mice acquire social hierarchical rank behaviors within their cage environments between six-eight weeks of age, which can be leveraged to examine normal social dominance behavior. We tested for this normal hierarchical behavior in our mice using the tube test for social dominance following our previously described methods.[206]

Acoustic Startle/Pre-Pulse Inhibition Task

Sensorimotor gating and startle reactivity were assessed using the Acoustic Startle/Pre-Pulse Inhibition (PPI) task following our previously published methods.[139,206] Briefly, acoustic startle to a 120 dB auditory stimulus pulse (40 ms broadband burst) and PPI (response to a pre-pulse plus the startle pulse) were measured concurrently using computerized instrumentation (StartleMonitor, Kinder Scientific) over 65 randomized trials. A percent PPI score for each trial was calculated using the following equation: % PPI = (startle pulse alone − (pre-pulse + startle pulse))/startle pulse alone × 100.

Resident Intruder

Male mice were single housed prior to testing, which was performed as previously described.[117,213] Briefly, after 10 days of single-housing to establish a territory, male cages were placed in a sound-attenuating box in the dark. Using infrared illuminators, the task was recorded by a digital camera on the night vision setting. Over three days, resident males received a different C57BL/6J WT stimulus animal and interactions were allowed for 10 minutes.

For analysis, we used neural networks for pose estimation followed by random forest classifiers from the pose estimates for classification of attack behaviors across each video, and then compared counts of each attack by genotype using SPSS. The random forest classifier program used was simple behavior annotator (simBA) version simba-uw-tf 0.85.3,[214] and the neural network used for body part tracking was deeplabcut (DLC), version 2.2rc3, using the resnetv50.[215]

Specifically, we labeled 240 frames taken from 120 approximately ten-minute videos that were converted from MTS to mp4 using ffmpeg. Each frame was labeled with sixteen unique body parts, eight per animal as according to the simBA 16bp user manual. The DLC neural net trained

using 80% of the labeled frames for approximately 370,000 iterations with default parameters. With a p-cutoff of 0.9 the trained network was able to predict high confidence mouse body part positions within 2.26 millimeters of human-labeled positions in the testing set, and the general quality of labels were confirmed by visual inspection of several videos. Estimates of pose were then exported to .csv for analysis by simBA. SimBA was trained to identify attack behavior (resident attacking intruder, RI) using 180 annotated behavior files, downloaded from https://osf.io/tmu6y/ in addition to four in-house annotated videos. All training files were annotated according to definitions found in the simBA preprint.[214]

In addition to these RI annotated files, a custom script was created to reverse the direction of attack in order to estimate instances of the intruder attacking the resident (IR). Both training sets were trained using 6000 trees, 20% training set, Gini impurity function, number of estimators equal to the squared number of features, and 1 min leaf. Probability thresholds for each model were chosen based on a maximum F1 score curve from the testing set. To ensure IR and RI datasets were mutually exclusive, a custom script was written to calculate the mean of random forest probability of overlapping frames of RI and IR behaviors and keep only the behavior with the larger mean probability across the overlap. Scoring by algorithms was visually inspected by trained behaviorists for a subset of videos to confirm accuracy. Finally, we tested for group differences in attacks in SPSS as described below. All custom code is available here upon request.

3.5.5 Data Analysis

Analyses were conducted in SPSS v27. Prior to analyses, data was screened for missing values and fit of distributions with assumptions underlying univariate analysis. This included the Shapiro-Wilk test on z-score-transformed data and qqplot investigations for normality, Levene's

test for homogeneity of variance, and boxplot and z-score (±3.29) investigation for identification of influential outliers. Means and standard errors were computed for each measure. All variables were examined via 3-way ANOVA with Bonferroni correction to assess the effects of the CD and TG alleles with sex included as a predictor. If appropriate, sex was dropped from the model to achieve best fit. Weight was used as a covariate for data analysis in the Acoustic Startle/Pre-Pulse Inhibition Task. For tasks with multiple timepoints measured per animal, a repeated measures ANOVA was applied if no data points were missing, otherwise a linear mixed model was used with the repeated predictor included as a random factor nested with subject to create hierarchy. To achieve normality for a given variable, outliers with a z-score ± 3.29 were removed or a square root transformation was applied. If normality could not be achieved and/or variance was not homogenous, nonparametric analysis was performed. All graphed data represents the raw values and the standard error of the mean.

Data Availability Statement

All data and detailed protocols are available upon reasonable request.

3.6 Acknowledgments

The authors would like to thank Amanda Titus and the Animal Behavior Subunit of the IDDRC at Washington University School of Medicine for assistance in running select assays, Carly Wender and Gunnar Forsberg for validation of the *Gtf2ird1* overexpressing mice, V. Campuzano for sharing the CD mice, Nastacia Goodwin from the University of Washington - Seattle for technical support for the SimBA program, and the transgenics core at UC Irvine for production of the TG mice. Some elements of the figures were created using BioRender.com. This

work was supported by the NSF (DGE-1745038 to KRN) and the NIMH (MH094604 (JV), R01MH067234 (KM), R01MH107515 (JDD)), and NICHD P50HD103525 (IDDRC@WUSTL).

3.7 Supplementary Tables

Table S1. Statistical information for Figure 1 – Molecular Validation

Figure	Panel	Method	Target	Genotype	n	Sex	n	Mean	SD	SEM	Age	Test	BY	Results	Notes
1	B	qPCR	Gtf2ird1	WT	5	not sexed		1.00	0.09	0.04	E13.5	t-test	Geno	t=-5.247, p<0.001	two-sided, equal variances assumed
				TG	5			1.65	0.26	0.12					
	C	Western	Gtf2ird1	WT	3	not sexed		1.05	0.37	0.22	E13.5	t-test	Geno	t=-1.991, p=0.048	one-sided, equal variances not assumed
				TG	5			1.95	0.90	0.40					
	E	qPCR	Gtf2ird1	WT	6	M,F	4,2	1.03	0.27	0.09	P19/20	ANOVA	Geno	F(3,20)=22.190, p=0.000001	main effect of Genotype; originally ran w/sex but no effect so dropped from model
				TG	6	M,F	3,3	1.62	0.21	0.09					
				CD	6	M,F	3,3	0.61	0.07	0.09					
				TG/CD	6	M,F	3,3	1.07	0.25	0.09					
	F	qPCR	Gtf2i	WT	6	M,F	4,2	1.01	0.15	0.08	P19/20	ANOVA	Geno	F(3,20)=12.818, p=0.000068	main effect of Genotype; originally ran w/sex but no effect so dropped from model
				TG	6	M,F	3,3	1.23	0.29	0.08					
				CD	6	M,F	3,3	0.68	0.18	0.08					
				TG/CD	6	M,F	3,3	0.61	0.13	0.08					
	H	Western	Gtf2ird1	WT	3	not sexed		1.01	0.20	0.10	P0/1	ANOVA	Geno	F(3,8)=9.918, p=0.005	main effect of Genotype significant individual differences: WT>CD, TG>CD, CD<TG/CD.
				TG	3			1.30	0.10	0.10					
				CD	3			0.59	0.09	0.10					
				TG/CD	3			1.13	0.23	0.10					

Table S2. Statistical information for Figure 2 – Sensorimotor Tasks

Figure	Panel	Task	Variable	Geno	n	Sex	n	Mean	SD	SEM	Cohort	Test	BY	Results	Notes
2	C	Open Field	Distance travelled (m)	WT	19	F	16	132.70	25.46	6.35	2 ~P30	ANOVA	CD x TG x Sex	HA: F(1,86)=3.639, p=0.06	no significant differences, even when removing insignificant factors, but HA trends lower.
						M	13	122.21	16.65	4.62					
				TG	21	F	11	121.77	19.55	5.89					
						M	10	107.44	25.71	8.13					
				CD	20	F	13	119.08	36.32	10.07		ANOVA	CD x TG x Sex	First 10 minutes: Sex: F(1,86)=5.366, p=0.023 First 20 minutes: Sex: F(1,86)=4.644, p=0.034 First 30 Minutes: HA: F(1,86)=3.978, p=0.049	sex effect only in 1st 10 or 20 minutes; HA effect in first 30 minutes, but not last 30 min.
						M	7	117.59	33.84	12.79					
				TG/CD	24	F	8	112.02	25.58	9.04					
						M	16	104.75	33.18	8.30					
	E	SMB:Walk	Time to Leave Square (s)	WT	20	F	9	2.14	1.65	0.55	3 >P60	ANOVA	CD x TG x Sex	No significant differences	Used sqrt - didn't fix normality issue though
						M	11	2.00	1.04	0.31					
				TG	21	F	13	2.80	1.86	0.52					
						M	8	1.80	0.76	0.27					
				CD	17	F	9	1.98	2.07	0.69		Kruskal-Wallis	Group	H(7)=3.735, p = 0.810	Since sqrt didn't fix normality issues so K-W is most applicable test.
						M	8	2.01	1.37	0.48					
				TG/CD	19	F	9	1.82	0.76	0.25					
						M	10	2.11	0.99	0.31					
	D	SMB: Inverted Screen	Time on Screen (s)	WT	20	F	9	58.50	4.50	1.50	3 >P60	Kruskal-Wallis	Group	H(7)=36.340, p<0.001	Pairwise comparisons: FCD < MTG, FWT, MWT, FTG; FTG/CD < MWT,< FTG
						M	11	58.36	5.43	1.64					
				TG	21	F	13	60.00	0.00	0.00					
						M	8	57.43	7.26	2.57					
				CD	17	F	9	36.42	14.63	4.88		Kruskal-Wallis	Geno	H(3)=30.208, p<0.001	CD << WT p<0.001 CD << TG p<0.001 TG/CD << WT p=0.002 TG/CD << TG p<0.001
						M	8	47.36	18.20	6.43					
				TG/CD	19	F	9	44.93	12.67	4.22					
						M	10	52.19	11.32	3.58					
	I	SMB: Pole	Time on Pole (s)	WT	20	F	9	38.75	24.06	8.02	3 >P60	Kruskal-Wallis	Group	H(7)=22.347, p=0.002	FCD << F WT & F TG/CD; plus others
						M	11	31.70	35.02	10.56					
				TG	21	F	13	23.90	18.54	5.14					
						M	8	33.96	35.07	12.40					
				CD	17	F	9	10.94	5.52	1.84		Kruskal-Wallis	Geno	H(3)=16.709, p<0.001	CD - TG p = 0.002 CD - WT p<0.001 TG/CD - TG p = 0.05 TG/CD - WT p=0.014
						M	8	20.57	20.85	7.37					
				TG/CD	19	F	9	13.11	5.02	1.67					
						M	10	19.64	17.01	5.38					
	not shown	SMB: Pole	Time to Turn (s)	WT	20	F	9	39.00	26.48	8.83	3 >P60	ANOVA	Group	TG*Sex: F(1,69)=5.352, p=0.024	sqrt transformed but still not normal; F TG < M TG
						M	11	37.22	37.26	11.23					
				TG	21	F	13	17.61	18.32	5.08					
						M	8	39.39	49.76	17.59					
				CD	17	F	9	51.42	36.12	12.04		Kruskal-Wallis	Group	H(7)=9.785, p=0.201	separating by trial, the second trial is more different, though none are significant.
						M	8	30.35	28.47	10.07					
				TG/CD	19	F	9	19.21	26.55	8.85					
						M	10	47.92	45.73	14.46					
	J	SMB: Ledge	Average Success	WT	20	F	9	0.78	0.26	0.11	3 >P60	ANOVA	CD x TG x Sex	Main effect of CD: F(1,69)=34.964, p<0.001 CD*HA interaction: F(1,69)=4.684, p=0.034	CD and TG/CD are lower than WT and TG TG/CD is lower than TG
						M	11	0.68	0.46	0.10					
				TG	21	F	13	0.92	0.19	0.09					
						M	8	0.88	0.23	0.12					
				CD	17	F	9	0.33	0.35	0.11		Kruskal-Wallis	Geno	H(3)=29.487, p<0.001	CD -WT p=0.033 TG/CD - WT p<0.001 TG/CD - TG p<0.001 CD - TG p<0.001
						M	8	0.56	0.50	0.12					
				TG/CD	19	F	9	0.33	0.25	0.11					
						M	10	0.25	0.26	0.10					
	L	SMB	PLAT	WT	20	F	9	59.39	1.83	0.61	3 >P60	Kruskal-Wallis	Group	H(7)=18.527, p=0.01	M CD << F TG and M&F WT; FCD << FWT
						M	11	53.38	11.35	3.42					
				TG	21	F	13	53.53	10.12	2.81					
						M	8	51.61	12.41	4.39					
				CD	17	F	9	34.86	21.44	7.15		Kruskal-Wallis	Geno	H(3)=16.919, p<0.001	CD < TG/CD p=0.046 CD < TG p=0.001 CD < WT p<0.001 *partial rescue?
						M	8	34.40	18.55	6.56					
				TG/CD	19	F	9	49.99	14.06	4.69					
						M	10	41.89	20.68	6.54					
	not shown	Rotarod: Stationary	Latency to Fall (s)	WT	20	F	9	59.37	1.38	0.56	3 >P60	ANOVA	CD x TG x Sex	Sex effect: F(1,69)=5.176, p=0.026 CD effect: F(1,69)=10.541, p=0.002 Sex*CD: F(1,69)=6.159, p=0.016	not normal, violates levene's
						M	11	60.00	0.00	0.00					
				TG	21	F	13	60.00	0.00	0.00					
						M	8	59.04	2.71	0.96					
				CD	17	F	9	54.00	6.64	2.21		Kruskal-Wallis	Group	H(7)=19.69, p=0.006	FCD < F WT & FTG, and all males FTG/CD < FTG, MWT, MTG/CD (only looking at Geno, CD < WT and TG)
						M	8	58.29	4.83	1.71					
				TG/CD	19	F	9	56.26	5.22	1.74					
						M	10	59.53	1.48	0.47					
	not shown	Rotarod: Continuous	Latency to Fall (s)	WT	20	F	9	56.26	4.02	1.34	3 >P60	ANOVA	CD x TG x Sex	CD effect: F(1,69)=13.738, p<0.001 TG effect: F(1,69)=4.179, p=0.045 CD*TG: F(1,69)=3.946, p=0.051	not normal, violates levene's
						M	11	57.76	4.50	1.36					
				TG	21	F	13	57.63	3.24	0.90					
						M	8	56.54	2.83	1.00					
				CD	17	F	9	46.35	11.69	3.90		Kruskal-Wallis	Geno	H(3)=12.304, p=0.006	CD<TG p=0.006, CD < WT p=0.002 TG/CD < WT p=0.045 not powered enough to see sex*cd rescue effect apparent in graph.
						M	8	52.58	6.85	2.42					
				TG/CD	19	F	9	54.96	4.99	1.66					
						M	10	54.65	3.95	1.25					
	N/O	Rotarod: Accelerating	Latency to Fall (s)	WT	20	F	9	129.39	28.41	9.47	3 >P60	ANOVA	CD x TG x Sex	CD effect: F(1,69)=35.227, p<0.001 CD*TG: F(1,69)=6.977, p=0.01 Sex*CD*TG: F(1,69)=4.461, p=0.038	mostly normal, a few exceptions; average is normal: FWT>FCD p<0.001, MWT>MCDp=0.003, MTG>MCDTGp=0.016; CDF<CDMp=0.032; CDF<CDTGFp<0.001; WT>CD p<0.001, TG>TGCD p=0.021, CD<CDTGp=0.009
						M	11	127.89	13.86	4.18					
				TG	21	F	13	116.15	37.50	10.40					
						M	8	123.23	16.20	5.73					
				CD	17	F	9	59.41	24.25	8.08		rmANOVA	CD x TG x Sex	Main effect of trial, p<0.001 trial*CD p=0.03 trial*CD*TG*Sex p=0.025	later trials are less normal, all meet levene's; When split by sex, only females have TG*CD interaction and only males have trial*CD interaction. See right for graph - maybe use that instead of these two? Though F TG 2a not normal
						M	8	88.73	24.45	8.64					
				TG/CD	19	F	9	106.67	32.53	10.84					
						M	10	90.80	29.95	9.47					
	Q	MARBLE BURYING	Marbles Buried	WT	29	F	16	2.6	3.4	0.9	2 ~P35	ANOVA	CD x TG	Significant effect of CD allele: F(1,90)=46.170, p<0.001	used sqrt_MBAvgBuried; removed nonsig. Sex from model; still not passing assumptions for ANOVA so see below.
						M	13	3.1	3.5	1.0					
				TG	21	F	11	3.0	3.5	1.1					
						M	10	2.2	1.8	0.6					
				CD	20	F	13	0.2	0.4	0.1		Kruskal-Wallis	Geno	H(3)=34.458, p<0.001	mean rank: WT 59.47, TG 64.45, CD 33.13, TG/CD 30.19; individual comparisons: <0.001 for TG/CD-WT, TG/CD-TG, CD-WT, CD-TG; WT-TG are not different, TG/CD-CD are not different.
						M	7	0.2	0.6	0.2					
				TG/CD	24	F	8	0.0	0.0	0.0					
						M	16	0.2	0.4	0.1					

R	MARBLE BURYING	Center Entries	WT	29	F	16	162.4	35.9	9.0	2 ~P35	ANOVA	CD x TG x Sex	Significant effect of CD allele: F(1,86)=67.793, p<0.001	raw values; 1 group (F TG had 0.026 for Shapiro-Wilkes but no sex effect so that would disappear when collapsed)
					M	13	157.3	25.6	7.1					
			TG	21	F	11	149.2	31.4	9.5					
					M	10	161.9	40.6	12.8					
			CD	20	F	13	96.3	30.5	8.5					
					M	7	98.1	29.5	11.2		ANOVA	CD x TG	Main effect of CD: F(1,90)=76.712, p<0.001	remove sex from model, lower mean square error
			TG/CD	24	F	8	114.5	25.0	8.8					
					M	16	102.4	26.0	6.5					
not shown	Pre-Pulse Inhibition	Average Startle to 120 dB (N)	WT	15	F	7	0.05	0.04	0.01	3 >P60	ANOVA	Geno x Sex WITH Weight	Sex: F(1,45)=0.37, p=0.546 Geno: F(1,45)=4.815, p=0.005 Geno*Sex: F(1,45)=2.671, p=0.059	Once controlled for weight, sex effects disappear (except in WT p=0.046), only females are different from each other: WT < TG p<0.001 WT < CD p=0.017 WT < TG/CD p=0.002
					M	8	0.15	0.06	0.02					
			TG	16	F	10	0.14	0.04	0.01					
					M	6	0.17	0.03	0.01					
			CD	11	F	6	0.11	0.05	0.02		rmANOVA	CD x TG x Sex WITH Weight	Sex: F(1,45)=0.37, p=0.546 CD: F(1,45)=0.445, p=0.508 **TG: F(1,45)=12.898, p<0.001** **Trial*CD*TG: F(3,135)=4.787, p=0.003**	FWT < FCD p=0.017 FWT < TG p<0.001 Borderline CD*Sex (0.074) and CD*TG*Sex effects (0.081)
					M	5	0.11	0.04	0.02					
			TG/CD	12	F	5	0.14	0.01	0.01					
					M	7	0.16	0.05	0.02					
T	Pre-Pulse Inhibition	Average Combined Startle (N)	WT	15	F	7	0.05	0.02	0.01	3 >P60	rmANOVA	CD x TG x Sex WITH Weight	Sex: F(1,45)=1.555, p=0.219 CD: F(1,45)=1.805, p=0.186 **TG: F(1,45)=20.69, p<0.001** **Sex*CD*TG: F(1,45)=6.295, p=0.016**	FWT < FTG p<0.001 MCD < MCDTG p=0.001 Test weight is also significant - is that bad? Also, data not all normal, levens at 0.02 6 for one
					M	8	0.11	0.05	0.02					
			TG	16	F	10	0.11	0.04	0.01					
					M	6	0.12	0.05	0.02					
			CD	11	F	6	0.07	0.03	0.01					
					M	5	0.07	0.04	0.02		Kruskal-Wallis	Group	H(7)=21.133, p=0.004	FWT - M WT 0.008, MTG 0.01, F TG 0.001, F TG/CD 0.006, M TG/CD <0.001 MCD - MTG/CD 0.017 FCD - FTG 0.048, M TG/CD 0.015
					F	5	0.12	0.01	0.01					
			TG/CD	12	M	7	0.13	0.04	0.02					
not shown	Pre-Pulse Inhibition	Average Percent Inhibition of Startle	WT	15	F	7	32.08	11.70	4.42	3 >P60	rmANOVA	CD x TG x Sex WITH Weight	Sex: F(1,45)=2.538, p=0.118 CD: F(1,45)=3.511, p=0.067 TG: F(1,45)=0.636, p=0.429 **CD*TG: F(1,45)=5.227, p=0.027**	Not all individual sections are normal, but over all they are; TG > TG/CD p=0.007 CD > TG/CD p=0.048 TG vs TG/CD sig at each level CD vs TG/CD sig at trials 1 and 2
					M	8	32.41	12.03	4.25					
			TG	16	F	10	30.09	5.74	1.82					
					M	6	42.78	11.85	4.84					
			CD	11	F	6	28.46	8.28	3.38					
					M	5	37.85	10.38	4.46		Kruskal-Wallis	Group	H(7)=12.787, p=0.077	
			TG/CD	12	F	5	21.65	5.17	2.31					
					M	7	28.07	11.53	4.36					

Table S3. Statistical information for Figure 3 – Anxiety and Fear-Related Tasks

Figure	Panel	Task	Variable	Geno	n	Sex	n	Mean	SD	SEM	Cohort	Test	BY	Results	Notes
	A/B	OPEN FIELD/1-HR ACTIVITY	Percent Time in Center	WT	29	F	16	23.5	8.0	2.0	2 ~P30	ANOVA	CD x TG x Sex	Main effect of Sex: F(1,86)=13.240, p<0.001 Main effect of CD: F(1,86)=11.962, p<0.001 Main effect of TG: F(1,86)=6.685, p=0.011	Using raw values
						M	13	32.4	10.0	2.8					
				TG	21	F	11	19.4	4.8	1.5					
						M	10	24.4	13.9	4.4					
				CD	20	F	13	15.7	5.9	1.6					
						M	7	24.7	9.3	3.5					
				TG/CD	24	F	8	13.5	9.1	3.2					
						M	16	18.9	9.7	2.4					
														Main effect of Sex: F(1,86)=12.210, p<0.001 Main effect of CD: F(1,86)=13.175, p<0.001 Main effect of TG: F(1,86)=7.935, p=0.006	used sqrt transformation
	not shown	OPEN FIELD/1-HR ACTIVITY	Velocity (m/s)	WT	29	F	16	0.037	0.007	0.002	2 ~P30	ANOVA	CD x TG x Sex	TG: F(1,86)=3.682, p=0.058	no overall distances in velocity (not shown);
						M	13	0.034	0.005	0.001					
				TG	21	F	11	0.034	0.005	0.002					
						M	10	0.030	0.007	0.002					
				CD	20	F	13	0.033	0.010	0.003					
						M	7	0.033	0.009	0.004		rmANOVA		Main effect of trial p<0.001	If only looking at the first 20 minutes, there is a sex effect here (0.034), especially in first 10 min (0.016)
				TG/CD	24	F	8	0.031	0.007	0.003					
						M	16	0.029	0.009	0.002					
	not shown	OPEN FIELD/1-HR ACTIVITY	Center Velocity (m/s)	WT	29	F	16	0.064	0.020	0.005	2 ~P30	ANOVA	CD x TG x Sex	Main effect of Sex: F(1,86)=19.699, p<0.001 Main effect of CD: F(1,86)=6.107, p=0.015	Both sexes generally move faster in the center but females still move faster than males in the center; CD moving faster than WT
						M	13	0.046	0.017	0.005					
				TG	21	F	11	0.070	0.014	0.004					
						M	10	0.053	0.019	0.006					
				CD	20	F	13	0.078	0.028	0.008					
						M	7	0.052	0.020	0.007		rmANOVA		Main effect of trial p<0.001 Main effect of Sex: F(1,74)=23.195, p<0.001	some blank values because animals spent 0 time in Center during some bins
				TG/CD	24	F	8	0.086	0.029	0.010					
						M	16	0.063	0.025	0.006					
	not shown	OPEN FIELD/1-HR ACTIVITY	Perimeter Velocity (m/s)	WT	29	F	16	0.030	0.006	0.001	2 ~P30	ANOVA	CD x TG x Sex	Main effect of TG: F(1,86)=6.052, p=0.016	TG allele slower than WT allele…interestingly no sex effect here - only center.
						M	13	0.031	0.006	0.002					
				TG	21	F	11	0.026	0.005	0.001					
						M	10	0.026	0.007	0.002					
				CD	20	F	13	0.026	0.008	0.002					
						M	7	0.029	0.011	0.004		rmANOVA		Main effect of trial p<0.001 Main effect of TG: F(1,86)=6.901, p=0.01	TG effect driven by first 30 minutes
				TG/CD	24	F	8	0.025	0.006	0.002					
						M	16	0.024	0.008	0.002					
	E	ELEVATED PLUS MAZE	Percent Time in Open Arms	WT	29	F	16	6.7	3.8	0.9	2 ~P37	ANOVA	CD x TG x Sex	no significant factors	Using raw values
						M	13	6.5	3.1	0.9					
				TG	21	F	11	6.5	3.4	1.0					
						M	10	5.0	3.3	1.0					
				CD	20	F	13	6.5	4.3	1.2					
						M	7	6.2	3.1	1.2					
				TG/CD	24	F	8	6.4	2.7	1.0					
						M	16	5.8	3.3	0.8					
														no significant factors	used sqrt transformation

3	G	LIGHT/ DARK BOX	Percent Time in Light Side	WT	20	F	9	49.2	8.7	2.9	3 adult	ANOVA	CD x TG	Significant CD*TG Interaction: F(1,73)=424.277, p=0.011 Individual differences: WT - CD p=0.011, CD - TG/CD p=0.026	Using raw values; F TG and M TG were non-normal; variance was okay; no effect of sex so it was removed from the model to decrease mean square error
						M	11	51.3	7.1	2.1					
				TG	21	F	13	46.4	9.7	2.7					
						M	8	48.0	7.6	2.7					
				CD	17	F	9	43.5	10.0	3.3				Significant CD*TG Interaction: F(1,73)=7.460, p=0.008 Individual differences: WT - CD p=0.009, CD - TG/CD p=0.018	used sqrt_RelativePercent_LITTM to correct for two non-normal groups (M/F TG). No sex effect, removed from model.
						M	8	43.7	8.7	3.1					
				TG/CD	19	F	9	52.6	3.9	1.3					
						M	10	46.9	5.4	1.7					
	H	CONDITIONED FEAR	Percent Time Freezing - Training	WT	29	F	16	28.5	13.2	3.3	adult	ANOVA	CD x TG x Sex	Main effect of Sex: F(1,85)=5.944, p=0.017	raw data; minutes 3-5 of day 1, training pTGse; F freeze more tTGn M
						M	13	18.4	7.3	2.0					
				TG	21	F	11	31.0	11.0	3.3					
						M	10	18.9	11.2	3.5					
				CD	19	F	13	23.2	11.5	3.2	adult	ANOVA	CD x TG x Sex	Main effect of Sex: F(1,85)=5.606, p=0.02	sqrt transformation: FCD .032 SW,
						M	6	26.6	19.0	7.7					
				TG/CD	24	F	8	36.4	10.0	3.5					
						M	16	29.3	13.8	3.4					
	I	CONDITIONED FEAR	Percent Time Freezing - Context	WT	29	F	16	18.1	11.4	2.9	adult	ANOVA	CD x TG x Sex	Main effect of Sex: F(1,85)=4.599, p=0.035	raw data; minutes 1-8 of day 2; F freeze more tTGn M; underpowered to see Sex*CD*TG interaction?
						M	13	11.6	6.5	1.8					
				TG	21	F	11	19.2	9.9	3.0					
						M	10	15.2	10.2	3.2					
				CD	19	F	13	11.6	7.7	2.1	adult	ANOVA	CD x TG x Sex	Main effect of Sex: F(1,85)=5.650, p=0.02	used sqrt transformation; 0.016 WTF, MCD 0.01 SW;
						M	6	12.0	12.4	5.1					
				TG/CD	24	F	8	17.8	5.8	2.0					
						M	16	11.0	7.2	1.8					
	J/K	CONDITIONED FEAR	Percent Time Freezing - Cue	WT	29	F	16	29.5	14.5	3.6	adult	ANOVA	CD x TG x Sex	Main effect of TG: F(1,85)=29.860, p<0.001 TG*Sex Interaction: F(1,85)=11.861, p<0.001	raw data; minutes 2-10 of day 3; TG allele increases freezing, especially in F
						M	13	32.7	10.5	2.9					
				TG	21	F	11	51.1	17.9	5.4					
						M	10	36.8	11.5	3.6					
				CD	19	F	13	19.8	8.1	2.2	adult	ANOVA	CD x TG x Sex	Main effect of TG: F(1,85)=28.497, p<0.001 TG*Sex Interaction: F(1,85)=11.876, p<0.001	sqrt transformation; 0.056 MWT SW; likely unnecessary to use sqrt as SW not sig
						M	6	30.0	13.1	5.3					
				TG/CD	24	F	8	50.0	14.3	5.0					
						M	16	37.6	14.3	3.6					
	not shown	SHOCK SENSITIVITY	FLINCH	WT	29	F	16	0.1	0.0		adult	ANOVA	CD x TG x Sex	CD*Sex Interaction: F(85)=6.540, p=0.012	RAW DATA; CD F < CD M
						M	13	0.1	0.0						
				TG	21	F	11	0.2	0.0						
						M	10	0.2	0.0						
				CD	19	F	13	0.1	0.0		adult	Kruskal-Wallis	Group	H(7)=14.778, p=0.039	sqrt doesn't fix normality issues, but no individual group differences withstand Bonferonni correction.
						M	6	0.2	0.0						
				TG/CD	24	F	8	0.1	0.0						
						M	16	0.2	0.0						

Table S4. Statistical information for Figure 4 – Social Behavior Tasks

Figure	Panel	Task	Variable	Geno	n	Sex	n	Mean	SD	SEM	Age	Test	BY	Results	Notes
	C	USV	Number of Calls	WT	**111**	Age	5	95.89		13.7	1 P5/7/9	Linear mixed model	Geno x Age x Sex	Geno: F(3,125.549)=7.892, p<0.001 Geno*Age: F(8,217.406)=3.068, p=0.003	Sex was not significant (p=0.112) WT >> than CD & TG/CD at all ages.
							7	131.1		13.7					
							9	100.4		13.9					
				TG	69	Age	5	67.43		17.3					
							7	100.5		17.3					
							9	87.95		17.3					
				CD	72	Age	5	31.17		16.9		n here represented each animal by each day; divide by 3 for number of animals not sure if mean is affected by that distinction or not.			
							7	107.8		16.9					
							9	91.46		16.9					
				TG/CD	66	Age	5	28.6		17.7					
							7	50.1		17.7					
							9	25.97		17.7					
	D	USV	Weight	WT	35	Age	5	3.056	0.43	0.08	1 P5/7/9	rmANOVA	Geno	F(3,97)=7.375, p<0.001 Day and Day*Geno effects	weight increases over time (Day) WT - CD p=0.001 WT - TG/CD p=0.005
							7	3.987	0.53	0.1					
							9	4.897	0.62	0.12					
				TG	21	Age	5	3.075	0.5	0.1					
							7	3.924	0.61	0.12					
							9	4.7	0.72	0.15					
				CD	23	Age	5	2.651	0.39	0.1					
							7	3.412	0.5	0.12					
							9	4.129	0.65	0.14					
				TG/CD	22	Age	5	2.713	0.52	0.1					
							7	3.485	0.64	0.12					
							9	4.182	0.8	0.15					

4	E	Righting Reflex	Average Time to Right (s)	WT	18	F	8	1.11	0.73	0.26	1 P14	Kruskal-Wallis	Geno	H(3)=0.595, p=0.898	Normality and variance violations precluded parametric testing.
						M	10	0.98	0.53	0.17					
				TG	14	F	8	0.90	0.29	0.10					
						M	6	0.80	0.18	0.08					
				CD	17	F	9	0.93	0.26	0.09		Kruskal-Wallis	Group (Geno* Sex)	H(7)=4.848, p=0.679	No differences.
						M	8	0.76	0.14	0.05					
				TG/CD	18	F	10	0.81	0.17	0.05					
						M	8	1.04	0.35	0.12					
	not shown	Tube Test	Percent Wins	WT	20	F	9	63.0		14.1	3 adult	Kruskal-Wallis	Group	H(7)=19.356, p=0.007	F CD - F WT p=0.009 F TG - F WT p=0.006 F TG - M TG p=0.002 F TGCD - F WT p=0.014 and more
						M	11	24.2		6.5					
				TG	21	F	13	17.9		7.2					
						M	8	62.5		7.6					
				CD	17	F	9	18.5		11.3		Kruskal-Wallis	Geno	Female: H(3)=8.339, p=0.04	Female: CD-WT 0.016, TG-WT 0.012, TG/CD - WT 0.028 Male: TG-WT 0.007, TG/CD-WT 0.047
						M	8	45.8		15.3					
				TG/CD	19	F	9	18.5		8.1				Male: H(3)=7.797, p=0.05	
						M	10	33.3		8.1					
	F	Tube Test	TTDay 1 Wins Ratio	WT	20	F	9	0.7	0.5	0.2	3 adult	ANOVA	CD x TG x Sex	CD: F(1,69)=3.793, p=0.056 Sex*TG: F(1,69)=3.13, p=0.081 Sex*CD*HA: F(1,69)=2.914, p=0.092	violates levene's & normality so not best test but only way to see interaction effects?
						M	11	0.5	0.5	0.2					
				TG	21	F	13	0.3	0.5	0.1					
						M	8	0.9	0.4	0.1					
				CD	17	F	9	0.2	0.4	0.1		Kruskal-Wallis	Group	H(7)=11.102, p=0.134	not significant but perhaps underpowered due to potential sex*geno effects?
						M	8	0.4	0.5	0.2					
				TG/CD	19	F	9	0.3	0.5	0.2					
						M	10	0.5	0.5	0.2					
	G	Resident Intruder	Average Attacks by Resident	WT	11	F					3 adult	ANOVA	CD x TG	CD: F(1,32)=1.167, p=0.288 TG: F(1,32)=0.03, p=0.864 CD*HA: F(1,32)=0.202, p=0.656	Mutually exclusive numbers (i.e. either R or I attacks) no significant differences - likely due few animals actually engaging in attacks.
						M	11	12.0	13.7	4.1					
				TG	7	F									
						M	7	8.0	11.8	4.4					
				CD	8	F									remove non-attackers? Check to see how attackers fared in Tube Test
						M	8	14.4	16.7	5.9					
				TG/CD	10	F									
						M	10	16.5	17.7	5.6					
	H	Modified Social Approach	Social Investigation Time (%)	WT	29	F	16	36.9	8.5	2.1	~P33	ANOVA	CD x TG x Sex	No significant differences CD*HA: F(1,86)=2.970, p=0.088	Not significant but perhaps underpowered to find social effects, which are less straightforward; The insignificant CD*HA interaction is driven by females, which may be misled by low TG/CD female n.
						M	13	35.0	8.3	2.3					
				TG	21	F	11	41.6	9.2	2.8					
						M	10	42.4	12.0	3.8					
				CD	20	F	13	42.7	11.1	3.1		Kruskal-Wallis	Geno	H(3)=8.916, p=0.03 Individual differences from WT: TG: p=0.039, CD: p=0.013, TG/CD p=0.016	one group violated normality, so ran nonparametric analysis on Genotype alone since no sex effect observed in ANOVA; difference driven by 3rd timebin (10-15min)
						M	7	42.8	4.2	1.6					
				TG/CD	24	F	8	37.8	7.2	2.6					
						M	16	44.6	14.3	3.6					
	not shown	Modified Social Approach	Novel Investigation Time (%)	WT	28	F	15	31.4	12.4	3.2		ANOVA	CD x TG x Sex	Sex*CD*HA Interaction: F(1,85)=5.893, p=0.017	Just looking at Novel Investigation Time; WT M << CD M WT M << TG M HA F << HA M appears only Males have higher?
						M	13	22.3	11.7	3.2					
				TG	21	F	11	24.0	7.3	2.2					
						M	10	37.2	14.4	4.6					
				CD	20	F	13	31.5	10.9	3.0				H(7)=13.238, p=0.067	Perhaps Gtf2ird1 plays a larger role in M social behavior?? Though underpowered to see such effects clearly across all tasks.
						M	7	38.0	14.5	5.5					
				TG/CD	24	F	8	30.7	14.4	5.1					
						M	16	33.2	14.3	3.6					
	I	Modified Social Approach	Social Investigation at 15 min bin	WT	29	F	16	80.3	27.2	11.9		Kruskal-Wallis	Group	H(7)=14.366, p=0.045	incorporates sex
						M	13	87.7	44.0	12.8					
				TG	21	F	11	110.6	33.7	13.9					
						M	10	121.4	50.8	14.6					
				CD	20	F	13	122.4	67.4	12.8		Kruskal-Wallis	Geno	H(3)=13.160, p=0.004	eliminates sex; WT-TG p=0.009, WT-CD p=0.014, WT-TG/CD p=0.001. Looks like overexpression of TG allele has biggest effect in this case??? Why - MvF n probably...?
						M	7	115.3	28.5	17.4					
				TG/CD	24	F	8	111.6	37.7	16.3					
						M	16	131.5	53.6	11.5					
	J	Modified Social Approach	Mean Social Iinvestigation Bout (s)	WT	29	F	16	6.4	2.4	0.6		Kruskal-Wallis	Group	H(7)=13.670, p=0.057	
						M	13	6.5	3.0	0.8					
				TG	21	F	11	7.7	2.7	0.8					
						M	10	7.8	3.2	1.0					
				CD	20	F	13	10.6	6.2	1.7		Kruskal-Wallis	Geno	H(3)=12.574, p=0.006	WT << CD p=0.001 WT << TG/CD p=0.006
						M	7	8.5	2.7	1.0					
				TG/CD	24	F	8	7.2	1.0	0.3					
						M	16	9.6	4.5	1.1					
	not shown	Modified Social Approach	Mean Novel Iinvestigation Bout (s)	WT	28	F	15	5.1	1.7	0.4		Kruskal-Wallis	Group	H(7)=6.902, p=0.439	no differences
						M	13	4.4	1.6	0.5					
				TG	21	F	11	5.3	2.0	0.6					
						M	10	5.8	2.1	0.7					
				CD	20	F	13	6.4	4.1	1.1		Kruskal-Wallis	Geno	H(3)=3.072, P=0.381	no differences; similar visual trend but perhaps stress dampened towards the floor decreasing our ability to detect a difference. It almosts looks like TG allele rescues in F but not in M?? But hard to say because of sample size.
						M	7	7.2	4.3	1.6					
				TG/CD	24	F	8	4.4	1.8	0.7					
						M	16	6.1	4.2	1.1					

M	NOVEL OBJECT AVOIDANCE	Percent Time in Investigation Zone	WT	29	F	16	6.2	1.9	0.5	2 ~P35	ANOVA	CD x TG x Sex	Main effect of Sex: F(1,86)=11.069, p<0.001 Main effect of CD: F(1,86)=4.203, p=0.043	using raw values; some groups were slightly non-normal (p<0.05 but >0.01), so also ran sqrt transformation.
					M	13	7.3	1.4	0.4					
			TG	21	F	11	6.6	2.3	0.7					
					M	10	8.3	1.9	0.6					
			CD	20	F	13	5.2	1.2	0.3		ANOVA	CD x TG x Sex	Main effect of Sex: F(1,86)=11.901, p<0.001 Main effect of CD: F(1,86)=4.933, p=0.029	used sqrt transformation
					M	7	6.5	2.4	0.9					
			TG/CD	24	F	8	5.7	1.6	0.6					
					M	16	7.4	2.8	0.7					
N	NOVEL OBJECT AVOIDANCE	Mean Investigation Bout (s)	WT	29	F	16	0.9	0.2	0.0	2 ~P35	ANOVA	CD x TG x Sex	Main effect of CD: F(1,86)=11.013, p=.001	raw values; assumptions for normality are not met.
					M	13	0.9	0.2	0.1					
			TG	21	F	11	0.9	0.1	0.0					
					M	10	0.9	0.2	0.1					
			CD	20	F	13	1.2	0.4	0.1		Kruskal-Wallis	Geno	H(3)=11.271, p=0.010	on raw values; main effect of CD (higher mean bout than WT)
					M	7	1.1	0.3	0.1					
			TG/CD	24	F	8	1.1	0.3	0.1					
					M	16	1.1	0.4	0.1					
O/P	3-Chamber Social Approach	Social Cup Preference Index	WT	19	F	9	8.0	9.9	3.3	3 adult	ANOVA	CD x TG x Sex	Main effect of Sex: F(1,66)=5.414, p=0.023 Main effect of CD: F(1,66)=4.99, p=0.029	Males have higher pref index than females, 23.504 > 13.501 CD allele > WT allele pref index, 23.305 > 13.700
					M	10	19.7	19.1	6.0					
			TG	21	F	13	8.3	20.2	5.6					
					M	8	18.7	22.6	8.0					
			CD	15	F	9	16.5	13.7	4.6		ANOVA	CD x Sex	Sex: F(1,71)=6.061, p=0.016 CD: F(1,71)=5.527, p=0.022	removed TG allele, Mean square error is 306 lower than 327 is better fit
					M	6	28.2	21.0	8.6					
			TG/CD	19	F	9	21.2	21.2	7.1					
					M	10	27.4	14.1	4.5					
not shown	3-Chamber Social Approach	Novel Cup Preference Index	WT	19	F	9	20.8	25.1	8.4	3 adult	ANOVA	CD x TG x Sex	no significant differences CD: F(1,69)=3.299, p=0.074 TG: F(1,69)=2.959, p=0.09 Sex*TG: F(1,69)=3.607, p=0.062	Mean square error examining all interactions is 508.585; eliminate factors that aren't contributing; trimmed to Sex, CD, TG, Sex*TG; mean square error 487.886. Still high error, but better fit.
					M	10	37.9	11.8	3.7					
			TG	21	F	13	19.8	22.2	6.2					
					M	8	16.7	25.8	9.1					
			CD	15	F	9	27.5	24.0	8.0		ANOVA	CD x TG	Main effect of TG: F(1,30)=6.914, p=0.013	Split by Sex, showing Male results only - female results insignificant. MWT > MTG
					M	6	45.3	11.5	4.7					
			TG/CD	19	F	9	30.7	26.6	8.9					
					M	10	29.1	25.3	8.0					
S	Social Operant	Total Rewards (FR1 Mean)	WT		F	16	24.9		3.1	2 adult	Kruskal-Wallis	Geno	H(3)=7.961, p=0.047	INCLUDES LEARNERS & NONLEARNERS: TG - CD 0.005; TG - WT 0.069, so likely underpowered to see effects here, especially if effects are only in males vs females.
					M	13	23.2		1.3					
			TG		F	13	20.7		1.9					
					M	11	17.4		1.6					
			CD		F	13	27.2		2.8		ANOVA	CD x TG x Sex	Main effect of CD: F(1,47)=14.07, p<0.001	LEARNERS ONLY. Levene's is satisfied with sqrt of values; effect seems driven by males (only M sig when split, but not enough power to tell sex interaction effects).
					M	9	29.1		4.0					
			TG/CD		F	7	23.1		4.5					
					M	16	24.1		3.5					
T	Social Operant	Total Rewards (FR3)	WT	18	F	8	28.5		4.7	2 adult	ANOVA	CD x TG x Sex	CD*Sex interaction: F(1,44)=4.932, p=0.032	CD greater than WT alleles; specifically M WT vs MCD 0.015; and WT F vs M 0.006, so we're missing a sex diff with CD allele.
					M	10	21.2		2.5					
			TG	12	F	6	30.5		5.4					
					M	6	17.6		2.2					
			CD	13	F	9	24.1		3.0		ANOVA	CD x TG x Sex	Main effect of CD: F(1,22)=12.752, p=0.002	Males Only
					M	4	28.6		3.4					
			TG/CD	9	F	3	29.8		3.8					
					M	6	29.8		1.8					
U	Social Operant	Breakpoint (Learners)	WT	18	F	8	27.4		1.4	2 adult	ANOVA	CD x TG x Sex	F(1,46)=4.405, p = 0.041 (CD allele)	CD TG Sex CD*Sex CD*TG; removed all interactions with TG from model (TG*Sex and CD*TG*Sex). Better fit, lower MeanSquare error.
					M	10	24.3		1.7					
			TG	12	F	6	26.0		3.7					
					M	6	22.5		1.4					
			CD	13	F	9	27.7		2.6		Kruskal-Wallis	Geno	H(3)=8.092, p=0.044	LEARNERS - MALES TG - CD p=0.05 TG-TG/CD p=0.029 WT - TG/CD p=0.04 WT-CD p=0.074 (likely underpowered)
					M	4	29.3		1.4					
			TG/CD	9	F	3	28.0		4.0					
					M	6	29.5		1.4					

Chapter 4: CONCLUSION

4.1 Summary and Significance

In this thesis, I have contributed the first manipulation of the oxytocin system in the context of WS to directly test the oxytocin hypothesis in mice. An oxytocin antagonist had no significant effect on fear learning and recall, despite the differences present in contextual and cued recall in the CD mice. Interestingly, the exact phenotypes of the CD shown in Chapter 2 were not recapitulated during our deep characterization of the model in Chapter 3, suggesting that the Conditioned Fear task may be sensitive to other elements, making its reliability less than ideal. It is possible the baseline stress created by daily injections within the first study interfered with the expression of the phenotype. As the effect of oxytocin is known to be modulated by estrogen, and estrogen is sensitive to the stress response, it is not out of the question that stress prevented modulation of the phenotype by oxytocin. This outcome showed me it was important to fully characterize the CD model to better understand its phenotypic features so that future mechanistic studies will be more productive.

In Chapter 3, I completed this thorough characterization of the CD mouse modeling the most common WS deletion. Given the struggle to show a social phenotype in any WS mutants in our hands previously,[117,139] it was reassuring to find that social approach measures were consistently increased across tasks and correlated with social motivation increases. Compared to our prior work,[117,139] the expression of a hypersocial phenotype here but not in the previous papers are most likely due to strain differences; the relative hyperactivity of the FVB/AntJ x C57BL/6J hybrids we used previously could have interfered with our ability to see differences that were apparent on the typical C57BL/6J background. Indeed, such strain affects have been documented in other behavioral paradigms, such as those measuring activity and anxiety-like behaviors.[216]

Thus, it would be interesting to further explore the effects of strain within the context of the WS deletion, and this approach may yield even more information about how the rest of the genome influences social behaviors, or other unreproducible findings, and provide insight into the complexity and origins of the variability of the human phenotypes observed.

Within the context of the full deletion, we were also able to see clear roles for *Gtf2ird1* in sensorimotor and anxiety domains, though not in any social behaviors, which aligns with previous studies implicating *Gtf2i* and not *Gtf2ird1* in the social phenotype of WS. We were also able to highlight effects of imbalanced *Gtf2ird1* expression relative to the rest of the typical WS deletion in the high reactivity to sound stimuli apparent in the acoustic startle and cued recall elements of our suite of tasks. This work provides further evidence for *Gtf2ird1*'s role in sensory processing or integration, especially of sound, which may give insight into the specific phobias of WS that are often correlated with loud noises. In the future, it would also be interesting to study the impact of *Gtf2i* duplication more fully and assess the effect of a *Gtf2i* rescue on the phenotypes I show in the CD model; unfortunately, the BAC transgenic mice meant to overexpress *Gtf2i* did not produce protein (*data not shown*), and I was unable to test it here.

4.2 Future Directions

In addition to its utility for modeling the single gene duplication of *Gtf2ird1* and rescuing expression in CD mutants, this same transgenic mouse line might have use as a molecular tool to study the basic biology of the role of *Gtf2ird1* isoforms in gene regulation. Specifically, the novel *Gtf2ird1* line could be the ideal tool to tease apart potential differences in the alternatively spliced isoforms of the protein. This is because the HA tag on the exogenous *Gtf2ird1* is only found on a subset of the isoforms, due to alternative splicing of the last exons. Using antibodies specific to

the HA-tagged isoforms, ChIP-seq analysis could indicate targets that are specific to the different isoform variants, when compared to the overall binding of GTF2IRD1 (tagged or not).

In addition, further investigation of sensory processing and integration is implicated in the results of my comprehensive study of the WS-relevant models. Sensitivity to light and sound may be interfering with other measures and teasing apart their influence will help us better understand the mechanisms underlying behavioral differences, especially those that rely on external stimuli to successfully respond. In WS, balance deficits are present and are more apparent (compared to the non-WS population) when individuals use their eyes, which requires integrating the sensory information to coordinate movement.[43] In mice, it is possible sensory integration of information from the whiskers is altered, and could be used to model this sensory processing difference. The deficits we observed in CD mice in the Rotarod and Platform tasks could be the result of such altered sensory processing – locomotion, cliff detection, motor coordination, and even time in the center of an open field is affected by loss of whiskers.[204,217,218] To test whether abnormal processing is affecting this traits, it would be interesting to evaluate the effect of partial or total loss of whiskers, as previously published,[204] in both CD and WT animals to see if differences in behavior are eliminated. It would also be interesting to study possible differences in object localization by whiskers via a head-fixed paradigm adapted for mice,[219–221] or to employ sensory processing tasks that target visual processing and measure eye movements and pupil dilation in the context of learning related to integration of visual information.[222,223] While scent is never explicitly tested here, beyond inclusion in the Conditioned Fear Task, the olfactory system would also be interesting to investigate, perhaps initially via scent discrimination tasks; tactile and taste responsivity could also be integrated to create a broad sensory processing workflow to address possible differences in other sensory systems.[224] From there, an analysis of multisensory

processing could better inform broader sensory integration capabilities or deficits in CD mice or other WS-relevant models.[225]

Finally, a re-evaluation of the potential effects of oxytocin in WS is warranted. Given the expansive influence of the oxytocin system on systems clearly affected in WS, and its popularity as a treatment for atypical social behaviors, a targeted investigation of domains beyond conditioned fear learning is necessary to confirm or contradict oxytocin's role. By utilizing my work in Chapter 3, where I identify phenotypes of interest in the CD model, others should be able to determine whether oxytocin dysregulation in key to disruption in social or sensorimotor domains, which may provide more reliable support for the various roles oxytocin may play. When we initially chose to study fear learning, altered social behaviors were not evident in our hands; now that we have defined assays showing altered social approach and motivation in CD mice on the C57BL/6J background, there is a clear path forward to test the oxytocin dysregulation hypothesis. In addition, it may be worthwhile to study the vasopressin system as well, given that vasopressin and oxytocin have differential effect on anxiety-related measures, with vasopressin tending to increase anxiety, which is reflected in my measurements of open-field and light avoidance in the CD mice in Chapter 3. Alternatively, to best understand the potential implications of a modified oxytocin system on underlying mechanisms contributing to neurodevelopmental disabilities like WS, it may be required that we focus on an intermediary measure between the transiently expressed oxytocin and the altered outcomes observed.

4.3 Summary

Overall, this thesis provides critical information for future studies on the mechanisms underlying the effects of the CNVs in the WSCR. In Chapter 2, I showed overexpression of

oxytocin was not causing conditioned fear deficits in cued and contextual recall and no large effects of oxytocin receptor or serotonin transporter dysregulation existed. Despite the lack of critical findings, I did provide evidence that certain regions warrant follow-up studies to specifically examine with a more powerful study design. In Chapter 3, I comprehensively profiled the Complete Deletion mouse line with a suite of behavioral tasks that highlighted key phenotypes that can used in future studies focused on mechanistic discovery or pharmacological testing. I also presented a novel transgenic *Gtf2ird1* mouse line and demonstrated its use in modeling *Gtf2ird1* duplication or its rescue in the context of the most common WS deletion when crossed to the Complete Deletion line. This molecular rescue led to the discovery of three phenotypes where *Gtf2ird1* plays a significant role – Light/Dark Box, Sensorimotor Battery Platform, and Rotarod tasks – highlighting the importance of *Gtf2ird1* in sensorimotor processing. In addition to the work I have done to eradicate erroneous statistical evaluations of the commonly used social approach task, which is presented in the Appendix, I have specifically contributed to the field of WS genetics by presenting a novel *Gtf2ird1* transgenic line, new information about the role of the *Gtf2ird1* gene in the context of the complete WSCR deletion in a mouse model, and better characterization of that CD mouse model with the hopes that these findings can propel research towards meaningful discoveries to serve the WS population.

References

1. Kozel BA, Barak B, Kim CA, et al. Williams syndrome. *Nat Rev Dis Primer*. 2021;7(1):1-22. doi:10.1038/s41572-021-00276-z

2. Pérez Jurado LA, Peoples R, Kaplan P, Hamel BC, Francke U. Molecular definition of the chromosome 7 deletion in Williams syndrome and parent-of-origin effects on growth. *Am J Hum Genet*. 1996;59(4):781-792.

3. Bayés M, Magano LF, Rivera N, Flores R, Pérez Jurado LA. Mutational mechanisms of Williams-Beuren syndrome deletions. *Am J Hum Genet*. 2003;73(1):131-151. doi:10.1086/376565

4. Williams JCP, Barratt-Boyes BG, Lowe JB. Supravalvular Aortic Stenosis. *Circulation*. 1961;24(6):1311-1318. doi:10.1161/01.CIR.24.6.1311

5. Somerville MJ, Mervis CB, Young EJ, et al. Severe expressive-language delay related to duplication of the Williams-Beuren locus. *N Engl J Med*. 2005;353(16):1694-1701. doi:10.1056/NEJMoa051962

6. Beuren AJ, Apitz J, Harmjanz D. Supravalvular Aortic Stenosis in Association with Mental Retardation and a Certain Facial Appearance. *Circulation*. 1962;26(6):1235-1240. doi:10.1161/01.CIR.26.6.1235

7. Schubert C. *The Genomic Basis of the Williams–Beuren Syndrome*. Vol 66.; 2008. doi:10.1007/s00018-008-8401-y

8. Strømme P, Bjømstad PG, Ramstad K. Prevalence estimation of Williams syndrome. *J Child Neurol*. 2002;17(4):269-271.

9. Carrasco X, Castillo S, Aravena T, Rothhammer P, Aboitiz F. Williams syndrome: Pediatric, neurologic, and cognitive development. *Pediatr Neurol*. 2005;32(3):166-172. doi:10.1016/j.pediatrneurol.2004.09.013

10. Honjo RS, Dutra RL, Furusawa EA, et al. Williams-Beuren Syndrome: A Clinical Study of 55 Brazilian Patients and the Diagnostic Use of MLPA. *BioMed Res Int*. 2015;2015:903175. doi:10.1155/2015/903175

11. Grant J, Karmiloff-Smith A, Gathercole SA, et al. Phonological Short-term Memory and its Relationship to Language in Williams Syndrome. *Cognit Neuropsychiatry*. 1997;2(2):81-99. doi:10.1080/135468097396342

12. Bellugi U, Lichtenberger L, Jones W, Lai Z, St. George M. I. The Neurocognitive Profile of Williams Syndrome: A Complex Pattern of Strengths and Weaknesses. *J Cogn Neurosci*. 2000;12(supplement 1):7-29. doi:10.1162/089892900561959

13. Bellugi U, Bihrle A, Jernigan T, Trauner D, Doherty S. Neuropsychological, neurological, and neuroanatomical profile of Williams syndrome. *Am J Med Genet*. 1990;37(S6):115-125. doi:10.1002/ajmg.1320370621

14. D'Souza D, Cole V, Farran EK, et al. Face processing in Williams syndrome is already atypical in infancy. *Front Psychol*. 2015;6. Accessed July 8, 2022. https://www.frontiersin.org/articles/10.3389/fpsyg.2015.00760

15. Wang MS, Schinzel A, Kotzot D, et al. Molecular and clinical correlation study of Williams-Beuren syndrome: No evidence of molecular factors in the deletion region or imprinting affecting clinical outcome. *Am J Med Genet*. 1999;86(1):34-43. doi:10.1002/(SICI)1096-8628(19990903)86:1<34::AID-AJMG7>3.0.CO;2-4

16. Järvinen A, Korenberg JR, Bellugi U. The social phenotype of Williams syndrome. *Curr Opin Neurobiol*. 2013;23(3):414-422. doi:10.1016/j.conb.2012.12.006

17. Morris CA. The behavioral phenotype of Williams syndrome: A recognizable pattern of neurodevelopment. *Am J Med Genet C Semin Med Genet*. 2010;154C(4):427-431. doi:10.1002/ajmg.c.30286

18. Doyle TF, Bellugi U, Korenberg JR, Graham J. "Everybody in the world is my friend" hypersociability in young children with Williams syndrome. *Am J Med Genet A*. 2004;124A(3):263-273. doi:10.1002/ajmg.a.20416

19. Bellugi U, Adolphs R, Cassady C, Chiles M. Towards the neural basis for hypersociability in a genetic syndrome. *Neuroreport*. 1999;10(8):1653-1657.

20. Järvinen-Pasley A, Bellugi U, Reilly J, et al. Defining the Social Phenotype in Williams Syndrome: A Model for Linking Gene, the Brain, and Behavior. *Dev Psychopathol*. 2008;20(1):1-35. doi:10.1017/S0954579408000011

21. Jones W, Bellugi U, Lai Z, et al. II. Hypersociability in Williams Syndrome. *J Cogn Neurosci*. 2000;12 Suppl 1:30-46.

22. Constantino JN, Kennon-McGill S, Weichselbaum C, et al. Infant viewing of social scenes is under genetic control and is atypical in autism. *Nature*. 2017;547(7663):340. doi:10.1038/nature22999

23. Riby DM, Hancock PJB. Viewing it differently: Social scene perception in Williams syndrome and Autism. *Neuropsychologia*. 2008;46(11):2855-2860. doi:10.1016/j.neuropsychologia.2008.05.003

24. Riby D, Hancock PJB. Looking at movies and cartoons: eye-tracking evidence from Williams syndrome and autism. *J Intellect Disabil Res*. 2009;53(2):169-181. doi:10.1111/j.1365-2788.2008.01142.x

25. Porter MA, Shaw TA, Marsh PJ. An unusual attraction to the eyes in Williams-Beuren syndrome: a manipulation of facial affect while measuring face scanpaths. *Cognit Neuropsychiatry*. 2010;15(6):505-530. doi:10.1080/13546801003644486

26. Riby DM, Jones N, Brown PH, et al. Attention to Faces in Williams Syndrome. *J Autism Dev Disord*. 2011;41(9):1228-1239. doi:10.1007/s10803-010-1141-5

27. Kleberg JL, Riby D, Fawcett C, et al. Williams syndrome: reduced orienting to other's eyes in a hypersocial phenotype. *J Autism Dev Disord*. Published online April 20, 2022. doi:10.1007/s10803-022-05563-6

28. Haas BW, Hoeft F, Searcy YM, Mills D, Bellugi U, Reiss A. Individual differences in social behavior predict amygdala response to fearful facial expressions in Williams syndrome. *Neuropsychologia*. 2010;48(5):1283-1288. doi:10.1016/j.neuropsychologia.2009.12.030

29. von Arnim G, Engel P. Mental Retardation Related to Hypercalcaemia. *Dev Med Child Neurol*. 1964;6(4):366-377. doi:10.1111/j.1469-8749.1964.tb08138.x

30. Leyfer OT, Woodruff-Borden J, Klein-Tasman BP, Fricke JS, Mervis CB. Prevalence of Psychiatric Disorders in 4 - 16-Year-Olds with Williams Syndrome. *Am J Med Genet Part B Neuropsychiatr Genet Off Publ Int Soc Psychiatr Genet*. 2006;141B(6):615-622. doi:10.1002/ajmg.b.30344

31. Royston R, Howlin P, Waite J, Oliver C. Anxiety Disorders in Williams Syndrome Contrasted with Intellectual Disability and the General Population: A Systematic Review and Meta-Analysis. *J Autism Dev Disord*. 2017;47(12):3765-3777. doi:10.1007/s10803-016-2909-z

32. Ng-Cordell E, Hanley M, Kelly A, Riby DM. Anxiety in Williams Syndrome: The Role of Social Behaviour, Executive Functions and Change Over Time. *J Autism Dev Disord*. 2018;48(3):796. doi:10.1007/s10803-017-3357-0

33. Gosch A, Pankau R. Personality characteristics and behaviour problems in individuals of different ages with Williams syndrome. *Dev Med Child Neurol*. 1997;39(8):527-533. doi:10.1111/j.1469-8749.1997.tb07481.x

34. Blomberg S, Rosander M, Andersson G. Fears, hyperacusis and musicality in Williams syndrome. *Res Dev Disabil*. 2006;27(6):668-680. doi:10.1016/j.ridd.2005.09.002

35. Levitin DJ, Cole K, Lincoln A, Bellugi U. Aversion, awareness, and attraction: investigating claims of hyperacusis in the Williams syndrome phenotype. *J Child Psychol Psychiatry*. 2005;46(5):514-523. doi:10.1111/j.1469-7610.2004.00376.x

36. Dilts CV, Morris CA, Leonard CO. Hypothesis for development of a behavioral phenotype in Williams syndrome. *Am J Med Genet*. 1990;37(S6):126-131. doi:10.1002/ajmg.1320370622

37. Gothelf D, Farber N, Raveh E, Apter A, Attias J. Hyperacusis in Williams syndrome: characteristics and associated neuroaudiologic abnormalities. *Neurology*. 2006;66(3):390-395. doi:10.1212/01.wnl.0000196643.35395.5f

38. Paglialonga A, Barozzi S, Brambilla D, et al. Cochlear active mechanisms in young normal-hearing subjects affected by Williams syndrome: time-frequency analysis of otoacoustic emissions. *Hear Res*. 2011;272(1-2):157-167. doi:10.1016/j.heares.2010.10.004

39. Chapman CA, du Plessis A, Pober BR. Neurologic Findings in Children and Adults With Williams Syndrome. *J Child Neurol*. 1996;11(1):63-65. doi:10.1177/088307389601100116

40. Morris CA, Demsey SA, Leonard CO, Dilts C, Blackburn BL. Natural history of Williams syndrome: Physical characteristics. *J Pediatr*. 1988;113(2):318-326. doi:10.1016/S0022-3476(88)80272-5

41. Van der Geest JN, Lagers-van Haselen GC, van Hagen JM, et al. Visual depth processing in Williams–Beuren syndrome. *Exp Brain Res*. 2005;166(2):200-209. doi:10.1007/s00221-005-2355-1

42. Gagliardi C, Martelli S, Burt MD, Borgatti R. Evolution of Neurologic Features in Williams Syndrome. *Pediatr Neurol*. 2007;36(5):301-306. doi:10.1016/j.pediatrneurol.2007.01.001

43. Barozzi S, Soi D, Gagliardi C, et al. Balance function in patients with Williams syndrome. *Gait Posture*. 2013;38(2):221-225. doi:10.1016/j.gaitpost.2012.11.012

44. Atkinson J, Anker S, Braddick O, Nokes L, Mason A, Braddick F. Visual and visuospatial development in young children with Williams syndrome. *Dev Med Child Neurol*. 2001;43(5):330-337. doi:10.1017/s0012162201000615

45. Dutton GN. 'Dorsal stream dysfunction' and 'dorsal stream dysfunction plus': a potential classification for perceptual visual impairment in the context of cerebral visual impairment? *Dev Med Child Neurol*. 2009;51(3):170-172. doi:10.1111/j.1469-8749.2008.03257.x

46. Powell B, Van Herwegen J. Sensory Processing in Williams Syndrome: Individual differences and changes over time. *J Autism Dev Disord*. 2022;52(7):3129-3141. doi:10.1007/s10803-021-05197-0

47. Berg JS, Brunetti-Pierri N, Peters SU, et al. Speech delay and autism spectrum behaviors are frequently associated with duplication of the 7q11.23 Williams-Beuren syndrome region. *Genet Med*. 2007;9(7):427-441. doi:10.1097/GIM.0b013e3180986192

48. Morris CA, Mervis CB, Paciorkowski AP, et al. 7q11.23 Duplication Syndrome: Physical Characteristics and Natural History. *Am J Med Genet A*. 2015;167A(12):2916-2935. doi:10.1002/ajmg.a.37340

49. Zarate YA, Lepard T, Sellars E, et al. Cardiovascular and genitourinary anomalies in patients with duplications within the Williams syndrome critical region: phenotypic expansion and

review of the literature. *Am J Med Genet A*. 2014;164A(8):1998-2002. doi:10.1002/ajmg.a.36601

50. Parrott A, James J, Goldenberg P, et al. Aortopathy in the 7q11.23 microduplication syndrome. *Am J Med Genet A*. 2015;167A(2):363-370. doi:10.1002/ajmg.a.36859

51. Torniero C, Dalla Bernardina B, Novara F, et al. Dysmorphic features, simplified gyral pattern and 7q11.23 duplication reciprocal to the Williams-Beuren deletion. *Eur J Hum Genet*. 2008;16(8):880-887. doi:10.1038/ejhg.2008.42

52. Sanders SJ, Ercan-Sencicek AG, Hus V, et al. Multiple Recurrent De Novo CNVs, Including Duplications of the 7q11.23 Williams Syndrome Region, Are Strongly Associated with Autism. *Neuron*. 2011;70(5):863-885. doi:10.1016/j.neuron.2011.05.002

53. Lugo M, Wong ZC, Billington Jr CJ, et al. Social, neurodevelopmental, endocrine, and head size differences associated with atypical deletions in Williams–Beuren syndrome. *Am J Med Genet A*. 2020;182(5):1008-1020. doi:10.1002/ajmg.a.61522

54. Kopp ND, Parrish PCR, Lugo M, Dougherty JD, Kozel BA. Exome sequencing of 85 Williams–Beuren syndrome cases rules out coding variation as a major contributor to remaining variance in social behavior. *Mol Genet Genomic Med*. 2018;6(5):749-765. doi:10.1002/mgg3.429

55. Klein-Tasman BP, Yund BD, Mervis CB. The Behavioral Phenotype of 7q11.23 Duplication Syndrome Includes Risk for Oppositional Behavior and Aggression. *J Dev Behav Pediatr*. Published online May 13, 2022:10.1097/DBP.0000000000001068. doi:10.1097/DBP.0000000000001068

56. Torniero C, Bernardina B dalla, Novara F, et al. Cortical dysplasia of the left temporal lobe might explain severe expressive-language delay in patients with duplication of the Williams–Beuren locus. *Eur J Hum Genet*. 2007;15(1):62-67. doi:10.1038/sj.ejhg.5201730

57. Gregory MD, Kolachana B, Yao Y, et al. A method for determining haploid and triploid genotypes and their association with vascular phenotypes in Williams syndrome and 7q11.23 duplication syndrome. *BMC Med Genet*. 2018;19(1):53. doi:10.1186/s12881-018-0563-3

58. Ramocki MB, Bartnik M, Szafranski P, et al. Recurrent Distal 7q11.23 Deletion Including HIP1 and YWHAG Identified in Patients with Intellectual Disabilities, Epilepsy, and Neurobehavioral Problems. *Am J Hum Genet*. 2010;87(6):857-865. doi:10.1016/j.ajhg.2010.10.019

59. Jan S, Ochi A, Kagawa K, et al. Intractable apnoeic seizures in a child with a deletion typically associated with Williams syndrome. *Epileptic Disord*. 2018;20(6):530-534. doi:10.1684/epd.2018.1013

60. Marshall CR, Young EJ, Pani AM, et al. Infantile Spasms Is Associated with Deletion of the MAGI2 Gene on Chromosome 7q11.23-q21.11. *Am J Hum Genet*. 2008;83(1):106-111. doi:10.1016/j.ajhg.2008.06.001

61. Fusco C, Micale L, Augello B, et al. Smaller and larger deletions of the Williams Beuren syndrome region implicate genes involved in mild facial phenotype, epilepsy and autistic traits. *Eur J Hum Genet*. 2014;22(1):64-70. doi:10.1038/ejhg.2013.101

62. Karmiloff-Smith A, Grant J, Ewing S, et al. Using case study comparisons to explore genotype-phenotype correlations in Williams-Beuren syndrome. *J Med Genet*. 2003;40(2):136-140. doi:10.1136/jmg.40.2.136

63. Botta A, Novelli G, Mari A, et al. Detection of an atypical 7q11.23 deletion in Williams syndrome patients which does not include the STX1A and FZD3 genes. *J Med Genet*. 1999;36(6):478-480. doi:10.1136/jmg.36.6.478

64. Tassabehji M, Hammond P, Karmiloff-Smith A, et al. GTF2IRD1 in Craniofacial Development of Humans and Mice. *Science*. 2005;310(5751):1184-1187. doi:10.1126/science.1116142

65. Dai L, Bellugi U, Chen XN, et al. Is it Williams syndrome? GTF2IRD1 implicated in visual–spatial construction and GTF2I in sociability revealed by high resolution arrays. *Am J Med Genet A*. 2009;149A(3):302-314. doi:10.1002/ajmg.a.32652

66. Antonell A, Campo MD, Magano LF, et al. Partial 7q11.23 deletions further implicate GTF2I and GTF2IRD1 as the main genes responsible for the Williams–Beuren syndrome neurocognitive profile. *J Med Genet*. 2010;47(5):312-320. doi:10.1136/jmg.2009.071712

67. Gagliardi C, Bonaglia MC, Selicorni A, Borgatti R, Giorda R. Unusual cognitive and behavioural profile in a Williams syndrome patient with atypical 7q11.23 deletion. *J Med Genet*. 2003;40(7):526-530. doi:10.1136/jmg.40.7.526

68. Chailangkarn T, Noree C, Muotri AR. The contribution of GTF2I haploinsufficiency to Williams syndrome. *Mol Cell Probes*. Published online January 3, 2018. doi:10.1016/j.mcp.2017.12.005

69. Adamo A, Atashpaz S, Germain PL, et al. 7q11.23 dosage-dependent dysregulation in human pluripotent stem cells affects transcriptional programs in disease-relevant lineages. *Nat Genet*. 2014;47(2):ng.3169. doi:10.1038/ng.3169

70. Chailangkarn T, Trujillo CA, Freitas BC, et al. A human neurodevelopmental model for Williams syndrome. *Nature*. 2016;536(7616):338-343. doi:10.1038/nature19067

71. Oliver G, Schäfer EA. On the physiological action of extracts of pituitary body and certain other glandular organs: preliminary communication. *J Physiol*. 1895;18(3):277.

72. Dale HH. On some physiological actions of ergot. *J Physiol*. 1906;34(3):163.

73. Ott I, Scott JC. The action of infundibulin upon the mammary secretion. *Proc Soc Exp Biol Med*. 1910;8(2):48-49.

74. Sharpey-Schafer EA, Mackenzie K. The action of animal extracts on milk secretion. *Proc R Soc Lond Ser B Contain Pap Biol Character*. 1911;84(568):16-22.

75. Kamm O, Aldrich TB, Grote IW, Rowe LW, Bugbee EP. The active principles of the posterior lobe of the pituitary gland. 1 i. the demonstration of the presence of two active principles. ii. The separation of the two principles and their concentration in the form of potent solid preparations. *J Am Chem Soc*. 1928;50(2):573-601.

76. Tuppy H. The amino-acid sequence in oxytocin. *Biochim Biophys Acta*. 1953;11(3):449-450.

77. du Vigneaud V, Ressler C, Trippett S. THE SEQUENCE OF AMINO ACIDS IN OXYTOCIN, WITH A PROPOSAL FOR THE STRUCTURE OF OXYTOCIN. *J Biol Chem*. 1953;205(2):949-957. doi:10.1016/S0021-9258(18)49238-1

78. du Vigneaud V, Gish DT, Katsoyannis PG. A synthetic preparation possessing biological properties associated with argininevasopressin. *J Am Chem Soc*. 1954;76(18):4751-4752.

79. Vigneaud V du, Ressler C, Swan JM, Roberts CW, Katsoyannis PG. The Synthesis of Oxytocin1. ACS Publications. doi:10.1021/ja01641a004

80. Some Selected History of Oxytocin and Vasopressin. National Institute of Mental Health (NIMH). Accessed July 14, 2022. https://www.nimh.nih.gov/research/research-conducted-at-nimh/research-areas/clinics-and-labs/lcmr/snge/vpot/some-selected-history-of-oxytocin-and-vasopressin

81. Ivell R, Richter D. Structure and comparison of the oxytocin and vasopressin genes from rat. *Proc Natl Acad Sci*. 1984;81(7):2006-2010. doi:10.1073/pnas.81.7.2006

82. Kimura T, Tanizawa O, Mori K, Brownstein MJ, Okayama H. Structure and expression of a human oxytocin receptor. *Nature*. 1992;356(6369):526-529. doi:10.1038/356526a0

83. Einfeld SL, Smith E, McGregor IS, et al. A double-blind randomized controlled trial of oxytocin nasal spray in Prader Willi syndrome. *Am J Med Genet A*. 2014;164A(9):2232-2239. doi:10.1002/ajmg.a.36653

84. DeMayo MM, Song YJC, Hickie IB, Guastella AJ. A Review of the Safety, Efficacy and Mechanisms of Delivery of Nasal Oxytocin in Children: Therapeutic Potential for Autism and Prader-Willi Syndrome, and Recommendations for Future Research. *Pediatr Drugs*. 2017;19(5):391-410. doi:10.1007/s40272-017-0248-y

85. Lee HJ, Macbeth AH, Pagani JH, Young 3rd WS. Oxytocin: the great facilitator of life. *Prog Neurobiol*. 2009;88(2):127-151.

86. Gimpl G, Fahrenholz F. The Oxytocin Receptor System: Structure, Function, and Regulation. *Physiol Rev*. 2001;81(2):629-683. doi:10.1152/physrev.2001.81.2.629

87. Insel TR, Winslow JT, Wang ZX, Young L, Hulihan TJ. Oxytocin and the molecular basis of monogamy. *Adv Exp Med Biol*. 1995;395:227-234.

88. Insel TR, Winslow JT, Wang Z, Young LJ. Oxytocin, Vasopressin, and the Neuroendocrine Basis of Pair Bond Formation. In: Zingg HH, Bourque CW, Bichet DG, eds. *Vasopressin and Oxytocin: Molecular, Cellular, and Clinical Advances*. Advances in Experimental Medicine and Biology. Springer US; 1998:215-224. doi:10.1007/978-1-4615-4871-3_28

89. Young LJ, Huot B, Nilsen R, Wang Z, Insel TR. Species Differences in Central Oxytocin Receptor Gene Expression: Comparative Analysis of Promoter Sequences. *J Neuroendocrinol*. 1996;8(10):777-783. doi:10.1046/j.1365-2826.1996.05188.x

90. Insel TR, Shapiro LE. Oxytocin receptor distribution reflects social organization in monogamous and polygamous voles. *Proc Natl Acad Sci*. 1992;89(13):5981-5985. doi:10.1073/pnas.89.13.5981

91. Winslow JT, Insel TR. The social deficits of the oxytocin knockout mouse. *Neuropeptides*. 2002;36(2):221-229. doi:10.1054/npep.2002.0909

92. Guastella AJ, Mitchell PB, Dadds MR. Oxytocin Increases Gaze to the Eye Region of Human Faces. *Biol Psychiatry*. 2008;63(1):3-5. doi:10.1016/j.biopsych.2007.06.026

93. Kovács K, Kis A, Pogány Á, Koller D, Topál J. Differential effects of oxytocin on social sensitivity in two distinct breeds of dogs (Canis familiaris). *Psychoneuroendocrinology*. 2016;74(Supplement C):212-220. doi:10.1016/j.psyneuen.2016.09.010

94. Montag C, Sauer C, Reuter M, Kirsch P. An interaction between oxytocin and a genetic variation of the oxytocin receptor modulates amygdala activity toward direct gaze: evidence from a pharmacological imaging genetics study. *Eur Arch Psychiatry Clin Neurosci*. 2013;263(2):169-175. doi:10.1007/s00406-013-0452-x

95. Janeček M, Dabrowska J. Oxytocin facilitates adaptive fear and attenuates anxiety responses in animal models and human studies - potential interaction with the corticotropin releasing factor (CRF) system in the bed nucleus of the stria terminalis (BNST). *Cell Tissue Res*. 2019;375(1):143-172. doi:10.1007/s00441-018-2889-8

96. Kirsch P, Esslinger C, Chen Q, et al. Oxytocin Modulates Neural Circuitry for Social Cognition and Fear in Humans. *J Neurosci*. 2005;25(49):11489-11493. doi:10.1523/JNEUROSCI.3984-05.2005

97. Knobloch HS, Charlet A, Hoffmann LC, et al. Evoked Axonal Oxytocin Release in the Central Amygdala Attenuates Fear Response. *Neuron*. 2012;73(3):553-566. doi:10.1016/j.neuron.2011.11.030

98. Moaddab M, Dabrowska J. Oxytocin receptor neurotransmission in the dorsolateral bed nucleus of the stria terminalis facilitates the acquisition of cued fear in the fear-potentiated startle paradigm in rats. *Neuropharmacology*. 2017;121:130-139.

99. Petersson M, Eklund M, Uvnäs-Moberg K. Oxytocin decreases corticosterone and nociception and increases motor activity in OVX rats. *Maturitas*. 2005;51(4):426-433. doi:10.1016/j.maturitas.2004.10.005

100. De Coster L, Mueller SC, T'Sjoen G, De Saedeleer L, Brass M. The influence of Oxytocin on automatic motor simulation. *Psychoneuroendocrinology*. 2014;50:220-226. doi:10.1016/j.psyneuen.2014.08.021

101. Bowen MT, Peters ST, Absalom N, Chebib M, Neumann ID, McGregor IS. Oxytocin prevents ethanol actions at δ subunit-containing GABAA receptors and attenuates ethanol-induced motor impairment in rats. *Proc Natl Acad Sci*. 2015;112(10):3104-3109. doi:10.1073/pnas.1416900112

102. Quattrocki E, Friston K. Autism, oxytocin and interoception. *Neurosci Biobehav Rev*. 2014;47:410-430. doi:10.1016/j.neubiorev.2014.09.012

103. Josef L, Goldstein P, Mayseless N, Ayalon L, Shamay-Tsoory SG. The oxytocinergic system mediates synchronized interpersonal movement during dance. *Sci Rep*. 2019;9(1):1894. doi:10.1038/s41598-018-37141-1

104. Alcorn III JL, Green CE, Schmitz J, Lane SD. Effects of oxytocin on aggressive responding in healthy adult males. *Behav Pharmacol*. 2015;26(8 0 0):798.

105. Dai L, Carter CS, Ying J, Bellugi U, Pournajafi-Nazarloo H, Korenberg JR. Oxytocin and Vasopressin Are Dysregulated in Williams Syndrome, a Genetic Disorder Affecting Social Behavior. *PLOS ONE*. 2012;7(6):e38513. doi:10.1371/journal.pone.0038513

106. Procyshyn TL, Spence J, Read S, Watson NV, Crespi BJ. The Williams syndrome prosociality gene GTF2I mediates oxytocin reactivity and social anxiety in a healthy population. *Biol Lett*. 2017;13(4):20170051. doi:10.1098/rsbl.2017.0051

107. Haas BW, Smith AK. Oxytocin, vasopressin, and Williams syndrome: epigenetic effects on abnormal social behavior. *Front Genet*. 2015;6. doi:10.3389/fgene.2015.00028

108. Haas BW, Filkowski MM, Cochran RN, et al. Epigenetic modification of OXT and human sociability. *Proc Natl Acad Sci*. 2016;113(27):E3816-E3823. doi:10.1073/pnas.1602809113

109. Henrichsen CN, Csárdi G, Zabot MT, et al. Using Transcription Modules to Identify Expression Clusters Perturbed in Williams-Beuren Syndrome. *PLOS Comput Biol*. 2011;7(1):e1001054. doi:10.1371/journal.pcbi.1001054

110. Strong E, Butcher DT, Singhania R, et al. Symmetrical Dose-Dependent DNA-Methylation Profiles in Children with Deletion or Duplication of 7q11.23. *Am J Hum Genet*. 2015;97(2):216-227. doi:10.1016/j.ajhg.2015.05.019

111. Crespi BJ, Hurd PL. Cognitive-behavioral phenotypes of Williams syndrome are associated with genetic variation in the GTF2I gene, in a healthy population. *BMC Neurosci*. 2014;15(1):127. doi:10.1186/s12868-014-0127-1

112. Swartz JR, Waller R, Bogdan R, et al. A Common Polymorphism in a Williams Syndrome Gene Predicts Amygdala Reactivity and Extraversion in Healthy Adults. *Biol Psychiatry*. 2017;81(3):203-210. doi:10.1016/j.biopsych.2015.12.007

113. Jabbi M, Chen Q, Turner N, et al. Variation in the Williams syndrome GTF2I gene and anxiety proneness interactively affect prefrontal cortical response to aversive stimuli. *Transl Psychiatry*. 2015;5(8):e622-e622. doi:10.1038/tp.2015.98

114. Kimura R, Tomiwa K, Inoue R, et al. Dysregulation of the oxytocin receptor gene in Williams syndrome. *Psychoneuroendocrinology*. 2020;115:104631. doi:10.1016/j.psyneuen.2020.104631

115. Nygaard KR, Swift RG, Glick RM, et al. Oxytocin receptor activation does not mediate associative fear deficits in a Williams Syndrome model. *Genes Brain Behav*. 2022;21(1):e12750. doi:10.1111/gbb.12750

116. Çalışkan E, Şahin MN, Güldağ MA. Oxytocin and Oxytocin Receptor Gene Regulation in Williams Syndrome: A Systematic Review. *Yale J Biol Med*. 2021;94(4):623-635.

117. Kopp N, McCullough K, Maloney SE, Dougherty JD. Gtf2i and Gtf2ird1 mutation do not account for the full phenotypic effect of the Williams syndrome critical region in mouse models. *Hum Mol Genet*. Published online 2019. doi:10.1093/hmg/ddz176

118. DeSilva U, Massa H, Trask BJ, Green ED. Comparative Mapping of the Region of Human Chromosome 7 Deleted in Williams Syndrome. *Genome Res*. 1999;9(5):428-436.

119. Osborne LR. Animal models of Williams syndrome. *Am J Med Genet C Semin Med Genet*. 2010;154C(2):209-219. doi:10.1002/ajmg.c.30257

120. Crawley JN. Designing mouse behavioral tasks relevant to autistic-like behaviors. *Ment Retard Dev Disabil Res Rev*. 2004;10(4):248-258. doi:10.1002/mrdd.20039

121. Ramos A, Berton O, Mormède P, Chaouloff F. A multiple-test study of anxiety-related behaviours in six inbred rat strains. *Behav Brain Res*. 1997;85(1):57-69. doi:10.1016/S0166-4328(96)00164-7

122. Ramos A. Animal models of anxiety: do I need multiple tests? *Trends Pharmacol Sci*. 2008;29(10):493-498. doi:10.1016/j.tips.2008.07.005

123. Ramos A, Pereira E, Martins GC, Wehrmeister TD, Izídio GS. Integrating the open field, elevated plus maze and light/dark box to assess different types of emotional behaviors in one single trial. *Behav Brain Res*. 2008;193(2):277-288. doi:10.1016/j.bbr.2008.06.007

124. Nygaard KR, Maloney SE, Dougherty JD. Erroneous inference based on a lack of preference within one group: Autism, mice, and the social approach task. *Autism Res*. 2019;12(8):1171-1183. doi:10.1002/aur.2154

125. Segura Puimedon M. *Use of Mouse Models to Establish Genotype-Phenotype Correlations in Williams-Beuren Syndrome*. Ph.D. Thesis. Universitat Pompeu Fabra; 2012. Accessed June 27, 2022. http://www.tdx.cat/handle/10803/101408

126. Tadenev ALD, Burgess RW. Model validity for preclinical studies in precision medicine: precisely how precise do we need to be? *Mamm Genome*. 2019;30(5):111-122. doi:10.1007/s00335-019-09798-0

127. Borralleras C, Sahun I, Pérez-Jurado LA, Campuzano V. Intracisternal Gtf2i Gene Therapy Ameliorates Deficits in Cognition and Synaptic Plasticity of a Mouse Model of Williams–Beuren Syndrome. *Mol Ther*. 2015;23(11):1691-1699. doi:10.1038/mt.2015.130

128. Ortiz-Romero P, Borralleras C, Bosch-Morató M, et al. Epigallocatechin-3-gallate improves cardiac hypertrophy and short-term memory deficits in a Williams-Beuren syndrome mouse model. *PLOS ONE*. 2018;13(3):e0194476. doi:10.1371/journal.pone.0194476

129. Li HH, Roy M, Kuscuoglu U, et al. Induced chromosome deletions cause hypersociability and other features of Williams–Beuren syndrome in mice. *EMBO Mol Med*. 2009;1(1):50-65. doi:10.1002/emmm.200900003

130. Roy AL. Biochemistry and biology of the inducible multifunctional transcription factor TFII-I. *Gene*. 2001;274(1):1-13. doi:10.1016/S0378-1119(01)00625-4

131. Young EJ, Lipina T, Tam E, et al. Reduced fear and aggression and altered serotonin metabolism in Gtf2ird1-targeted mice. *Genes Brain Behav*. 2008;7(2):224-234. doi:10.1111/j.1601-183X.2007.00343.x

132. Palmer SJ, Santucci N, Widagdo J, et al. Negative Autoregulation of GTF2IRD1 in Williams-Beuren Syndrome via a Novel DNA Binding Mechanism. *J Biol Chem*. 2010;285(7):4715-4724. doi:10.1074/jbc.M109.086660

133. Mervis CB, Dida J, Lam E, et al. Duplication of GTF2I Results in Separation Anxiety in Mice and Humans. *Am J Hum Genet*. 2012;90(6):1064-1070. doi:10.1016/j.ajhg.2012.04.012

134. Sakurai T, Dorr NP, Takahashi N, McInnes LA, Elder GA, Buxbaum JD. Haploinsufficiency of Gtf2i, a gene deleted in Williams Syndrome, leads to increases in social interactions. *Autism Res*. 2011;4(1):28-39. doi:10.1002/aur.169

135. Howard ML, Palmer SJ, Taylor KM, et al. Mutation of Gtf2ird1 from the Williams–Beuren syndrome critical region results in facial dysplasia, motor dysfunction, and altered vocalisations. *Neurobiol Dis*. 2012;45(3):913-922. doi:10.1016/j.nbd.2011.12.010

136. O'Leary J, Osborne LR. Global Analysis of Gene Expression in the Developing Brain of Gtf2ird1 Knockout Mice. *PLOS ONE*. 2011;6(8):e23868. doi:10.1371/journal.pone.0023868

137. Schneider T, Skitt Z, Liu Y, et al. Anxious, hypoactive phenotype combined with motor deficits in Gtf2ird1 null mouse model relevant to Williams syndrome. *Behav Brain Res*. 2012;233(2):458-473. doi:10.1016/j.bbr.2012.05.014

138. Enkhmandakh B, Stoddard C, Mack K, et al. Generation of a mouse model for a conditional inactivation of Gtf2i allele. *genesis*. 2016;54(7):407-412. doi:10.1002/dvg.22948

139. Kopp ND, Nygaard KR, Liu Y, et al. Functions of Gtf2i and Gtf2ird1 in the developing brain: transcription, DNA binding and long-term behavioral consequences. *Hum Mol Genet*. 2020;29(9):1498-1519. doi:10.1093/hmg/ddaa070

140. Franke Y, Peoples RJ, Francke U. Identification of GTF2IRD1, a putative transcription factor within the Williams-Beuren syndrome deletion at 7q11.23. *Cytogenet Genome Res*. 1999;86(3-4):296-304. doi:10.1159/000015322

141. Bayarsaihan D, Dunai J, Greally JM, et al. 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*. 2002;79(1):137-143. doi:10.1006/geno.2001.6674

142. Hirota H, Matsuoka R, Chen XN, et al. Williams syndrome deficits in visual spatial processing linked to GTF2IRD1 and GTF2I on Chromosome 7q11.23. *Genet Med*. 2003;5(4):311-321. doi:10.1097/00125817-200307000-00007

143. Mervis CB, Robinson BF, Pani JR. Visuospatial Construction. *Am J Hum Genet*. 1999;65(5):1222-1229.

144. Donnai D, Karmiloff-Smith A. Williams syndrome: From genotype through to the cognitive phenotype. *Am J Med Genet*. 2000;97(2):164-171. doi:10.1002/1096-8628(200022)97:2<164::AID-AJMG8>3.0.CO;2-F

145. Mervis CB, Robinson BF, Bertrand J, Morris CA, Klein-Tasman BP, Armstrong SC. The Williams Syndrome Cognitive Profile. *Brain Cogn*. 2000;44(3):604-628. doi:10.1006/brcg.2000.1232

146. Martens MA, Wilson SJ, Reutens DC. Research Review: Williams syndrome: a critical review of the cognitive, behavioral, and neuroanatomical phenotype. *J Child Psychol Psychiatry*. 2008;49(6):576-608. doi:10.1111/j.1469-7610.2008.01887.x

147. Doherty-Sneddon G, Riby DM, Calderwood L, Ainsworth L. Stuck on you: Face-to-face arousal and gaze aversion in Williams syndrome. *Cognit Neuropsychiatry*. 2009;14(6):510-523. doi:10.1080/13546800903043336

148. Barak B, Feng G. Neurobiology of social behavior abnormalities in autism and Williams syndrome. *Nat Neurosci*. 2016;19(5):647-655. doi:10.1038/nn.4276

149. Fishman I, Yam A, Bellugi U, Mills D. Language and sociability: insights from Williams syndrome. *J Neurodev Disord*. 2011;3(3):185-192. doi:10.1007/s11689-011-9086-3

150. Levitin DJ, Menon V, Schmitt JE, et al. Neural Correlates of Auditory Perception in Williams Syndrome: An fMRI Study. *NeuroImage*. 2003;18(1):74-82. doi:10.1006/nimg.2002.1297

151. Korenberg JR, Chen XN, Hirota H, et al. VI. Genome Structure and Cognitive Map of Williams Syndrome. *J Cogn Neurosci*. 2000;12(supplement 1):89-107. doi:10.1162/089892900562002

152. Meyer-Lindenberg A, Mervis CB, Berman KF. Neural mechanisms in Williams syndrome: a unique window to genetic influences on cognition and behaviour. *Nat Rev Neurosci*. 2006;7(5):380-393. doi:10.1038/nrn1906

153. Klein-Tasman BP, Li-Barber KT, Magargee ET. Honing in on the Social Phenotype in Williams Syndrome Using Multiple Measures and Multiple Raters. *J Autism Dev Disord*. 2011;41(3):341-351. doi:10.1007/s10803-010-1060-5

154. Pisansky MT, Hanson LR, Gottesman II, Gewirtz JC. Oxytocin enhances observational fear in mice. *Nat Commun*. 2017;8(1):2102. doi:10.1038/s41467-017-02279-5

155. Guzmán YF, Tronson NC, Sato K, et al. Role of oxytocin receptors in modulation of fear by social memory. *Psychopharmacology (Berl)*. 2014;231(10):2097-2105. doi:10.1007/s00213-013-3356-6

156. Heinrichs M, von Dawans B, Domes G. Oxytocin, vasopressin, and human social behavior. *Front Neuroendocrinol*. 2009;30(4):548-557. doi:10.1016/j.yfrne.2009.05.005

157. Meyer-Lindenberg A, Domes G, Kirsch P, Heinrichs M. Oxytocin and vasopressin in the human brain: social neuropeptides for translational medicine. *Nat Rev Neurosci*. 2011;12(9):524-538. doi:10.1038/nrn3044

158. Dölen G, Darvishzadeh A, Huang KW, Malenka RC. Social reward requires coordinated activity of nucleus accumbens oxytocin and serotonin. *Nature*. 2013;501(7466):179-184. doi:10.1038/nature12518

159. Segura-Puimedon M, Sahún I, Velot E, et al. Heterozygous deletion of the Williams–Beuren syndrome critical interval in mice recapitulates most features of the human disorder. *Hum Mol Genet*. 2014;23(24):6481-6494. doi:10.1093/hmg/ddu368

160. Zoicas I, Slattery DA, Neumann ID. Brain Oxytocin in Social Fear Conditioning and Its Extinction: Involvement of the Lateral Septum. *Neuropsychopharmacology*. 2014;39(13):3027-3035. doi:10.1038/npp.2014.156

161. Hasan MT, Althammer F, Silva da Gouveia M, et al. A Fear Memory Engram and Its Plasticity in the Hypothalamic Oxytocin System. *Neuron*. 2019;103(1):133-146.e8. doi:10.1016/j.neuron.2019.04.029

162. Modi ME, Majchrzak MJ, Fonseca KR, et al. Peripheral Administration of a Long-Acting Peptide Oxytocin Receptor Agonist Inhibits Fear-Induced Freezing. *J Pharmacol Exp Ther*. 2016;358(2):164-172. doi:10.1124/jpet.116.232702

163. Mottolese R, Redouté J, Costes N, Le Bars D, Sirigu A. Switching brain serotonin with oxytocin. *Proc Natl Acad Sci U S A*. 2014;111(23):8637-8642. doi:10.1073/pnas.1319810111

164. Viviani D, Charlet A, Burg E van den, et al. Oxytocin Selectively Gates Fear Responses Through Distinct Outputs from the Central Amygdala. *Science*. 2011;333(6038):104-107. doi:10.1126/science.1201043

165. Rydén G, Sjöholm I. Half-Life of Oxytocin in Blood of Pregnant and Non-Pregnant Woman. *Acta Obstet Gynecol Scand*. 1969;48(S3):139-140. doi:https://doi.org/10.3109/00016346909157733

166. Goode TD, Leong KC, Goodman J, Maren S, Packard MG. Enhancement of striatum-dependent memory by conditioned fear is mediated by beta-adrenergic receptors in the basolateral amygdala. *Neurobiol Stress*. 2016;3:74-82. doi:10.1016/j.ynstr.2016.02.004

167. Campbell-Smith EJ, Holmes NM, Lingawi NW, Panayi MC, Westbrook RF. Oxytocin signaling in basolateral and central amygdala nuclei differentially regulates the acquisition, expression, and extinction of context-conditioned fear in rats. *Learn Mem*. 2015;22(5):247-257. doi:10.1101/lm.036962.114

168. Oettl LL, Ravi N, Schneider M, et al. Oxytocin Enhances Social Recognition by Modulating Cortical Control of Early Olfactory Processing. *Neuron*. 2016;90(3):609-621. doi:10.1016/j.neuron.2016.03.033

169. Lukas M, Toth I, Veenema AH, Neumann ID. Oxytocin mediates rodent social memory within the lateral septum and the medial amygdala depending on the relevance of the social stimulus: Male juvenile versus female adult conspecifics. *Psychoneuroendocrinology*. 2013;38(6):916-926. doi:10.1016/j.psyneuen.2012.09.018

170. Jiang Y, Platt ML. Oxytocin and vasopressin flatten dominance hierarchy and enhance behavioral synchrony in part via anterior cingulate cortex. *Sci Rep*. 2018;8(1):8201. doi:10.1038/s41598-018-25607-1

171. Yamagishi A, Lee J, Sato N. Oxytocin in the anterior cingulate cortex is involved in helping behaviour. *Behav Brain Res*. 2020;393:112790. doi:10.1016/j.bbr.2020.112790

172. Raam T, McAvoy KM, Besnard A, Veenema AH, Sahay A. Hippocampal oxytocin receptors are necessary for discrimination of social stimuli. *Nat Commun*. 2017;8(1):2001. doi:10.1038/s41467-017-02173-0

173. Lin YT, Hsieh TY, Tsai TC, Chen CC, Huang CC, Hsu KS. Conditional Deletion of Hippocampal CA2/CA3a Oxytocin Receptors Impairs the Persistence of Long-Term Social Recognition Memory in Mice. *J Neurosci*. 2018;38(5):1218-1231. doi:10.1523/JNEUROSCI.1896-17.2017

174. Choe HK, Reed MD, Benavidez N, et al. Oxytocin Mediates Entrainment of Sensory Stimuli to Social Cues of Opposing Valence. *Neuron*. 2015;87(1):152-163. doi:10.1016/j.neuron.2015.06.022

175. Resendez SL, Namboodiri VMK, Otis JM, et al. Social Stimuli Induce Activation of Oxytocin Neurons Within the Paraventricular Nucleus of the Hypothalamus to Promote

Social Behavior in Male Mice. *J Neurosci*. 2020;40(11):2282-2295. doi:10.1523/JNEUROSCI.1515-18.2020

176. Proulx É, Young EJ, Osborne LR, Lambe EK. Enhanced prefrontal serotonin 5-HT 1A currents in a mouse model of Williams-Beuren syndrome with low innate anxiety. *J Neurodev Disord*. 2010;2(2):99. doi:10.1007/s11689-010-9044-5

177. Lew CH, Groeniger KM, Hanson KL, et al. Serotonergic innervation of the amygdala is increased in autism spectrum disorder and decreased in Williams syndrome. *Mol Autism*. 2020;11(1):12. doi:10.1186/s13229-019-0302-4

178. Campeau S, Davis M. Involvement of the central nucleus and basolateral complex of the amygdala in fear conditioning measured with fear-potentiated startle in rats trained concurrently with auditory and visual conditioned stimuli. *J Neurosci*. 1995;15(3):2301-2311. doi:10.1523/JNEUROSCI.15-03-02301.1995

179. Lehmann K, Lesting J, Polascheck D, Teuchert-Noodt G. Serotonin fibre densities in subcortical areas: differential effects of isolated rearing and methamphetamine. *Dev Brain Res*. 2003;147(1):143-152. doi:10.1016/S0165-3806(03)00130-5

180. Avery SN, Clauss JA, Blackford JU. The Human BNST: Functional Role in Anxiety and Addiction. *Neuropsychopharmacology*. 2016;41(1):126-141. doi:10.1038/npp.2015.185

181. Moore RY, Halaris AE, Jones BE. Serotonin neurons of the midbrain raphe: Ascending projections. *J Comp Neurol*. 1978;180(3):417-438. doi:https://doi.org/10.1002/cne.901800302

182. Toth I, Neumann ID, Slattery DA. Central administration of oxytocin receptor ligands affects cued fear extinction in rats and mice in a timepoint-dependent manner. *Psychopharmacology (Berl)*. 2012;223(2):149-158. doi:10.1007/s00213-012-2702-4

183. Gunduz-Cinar O, Brockway ET, Castillo LI, Pollack GA, Erguven T, Holmes A. Selective sub-nucleus effects of intra-amygdala oxytocin on fear extinction. *Behav Brain Res*. 2020;393:112798. doi:10.1016/j.bbr.2020.112798

184. Menon R, Grund T, Zoicas I, et al. Oxytocin Signaling in the Lateral Septum Prevents Social Fear during Lactation. *Curr Biol*. 2018;28(7):1066-1078.e6. doi:10.1016/j.cub.2018.02.044

185. Guzmán YF, Tronson NC, Jovasevic V, et al. Fear-enhancing effects of septal oxytocin receptors. *Nat Neurosci*. 2013;16(9):1185-1187. doi:10.1038/nn.3465

186. El-Ghundi M, O'Dowd BF, George SR. Prolonged fear responses in mice lacking dopamine D1 receptor. *Brain Res*. 2001;892(1):86-93. doi:10.1016/S0006-8993(00)03234-0

187. Darvas M, Fadok JP, Palmiter RD. Requirement of dopamine signaling in the amygdala and striatum for learning and maintenance of a conditioned avoidance response. *Learn Mem*. 2011;18(3):136-143. doi:10.1101/lm.2041211

188. Gainetdinov RR, Mohn AR, Bohn LM, Caron MG. Glutamatergic modulation of hyperactivity in mice lacking the dopamine transporter. *Proc Natl Acad Sci*. 2001;98(20):11047-11054. doi:10.1073/pnas.191353298

189. Maloney SE, Rieger MA, Al-Hasani R, Bruchas MR, Wozniak DF, Dougherty JD. Loss of CELF6 RNA binding protein impairs cocaine conditioned place preference and contextual fear conditioning. *Genes Brain Behav*. 2019;18(7):e12593. doi:https://doi.org/10.1111/gbb.12593

190. Khuchua Z, Wozniak DF, Bardgett ME, et al. Deletion of the n-terminus of murine map2 by gene targeting disrupts hippocampal ca1 neuron architecture and alters contextual memory. *Neuroscience*. 2003;119(1):101-111. doi:10.1016/S0306-4522(03)00094-0

191. Raposinho PD, Pierroz DD, Broqua P, White RB, Pedrazzini T, Aubert ML. Chronic administration of neuropeptide Y into the lateral ventricle of C57BL/6J male mice produces an obesity syndrome including hyperphagia, hyperleptinemia, insulin resistance, and hypogonadism. *Mol Cell Endocrinol*. 2001;185(1):195-204. doi:10.1016/S0303-7207(01)00620-7

192. Radulovic J, Rühmann A, Liepold T, Spiess J. Modulation of Learning and Anxiety by Corticotropin-Releasing Factor (CRF) and Stress: Differential Roles of CRF Receptors 1 and 2. *J Neurosci*. 1999;19(12):5016-5025.

193. Mantella RC, Vollmer RR, Li X, Amico JA. Female Oxytocin-Deficient Mice Display Enhanced Anxiety-Related Behavior. *Endocrinology*. 2003;144(6):2291-2296. doi:10.1210/en.2002-0197

194. Ferguson JN, Young LJ, Hearn EF, Matzuk MM, Insel TR, Winslow JT. Social amnesia in mice lacking the oxytocin gene. *Nat Genet*. 2000;25(3):284-288. doi:10.1038/77040

195. Ferguson JN, Aldag JM, Insel TR, Young LJ. Oxytocin in the Medial Amygdala is Essential for Social Recognition in the Mouse. *J Neurosci*. 2001;21(20):8278-8285. doi:10.1523/JNEUROSCI.21-20-08278.2001

196. Burkett JP, Andari E, Johnson ZV, Curry DC, Waal FBM de, Young LJ. Oxytocin-dependent consolation behavior in rodents. *Science*. 2016;351(6271):375-378. doi:10.1126/science.aac4785

197. Insel TR, Winslow JT, Witt DM. Homologous regulation of brain oxytocin receptors. *Endocrinology*. 1992;130(5):2602-2608. doi:10.1210/endo.130.5.1315251

198. Artymyshyn R, Smith A, Wolfe BB. The use of 3H standards in 125I autoradiography. *J Neurosci Methods*. 1990;32(3):185-192. doi:10.1016/0165-0270(90)90139-7

199. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods*. 2012;9(7):671-675. doi:10.1038/nmeth.2089

200. Paxinos G, Franklin KBJ. *Paxinos and Franklin's the Mouse Brain in Stereotaxic Coordinates*. Academic Press; 2019.

201. Ritz C, Baty F, Streibig JC, Gerhard D. Dose-Response Analysis Using R. *PLOS ONE*. 2015;10(12):e0146021. doi:10.1371/journal.pone.0146021

202. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. *J Stat Softw*. 2015;67(1):1-48. doi:10.18637/jss.v067.i01

203. Zanella M, Vitriolo A, Andirko A, et al. Dosage analysis of the 7q11.23 Williams region identifies BAZ1B as a major human gene patterning the modern human face and underlying self-domestication. *Sci Adv*. 2019;5(12):eaaw7908. doi:10.1126/sciadv.aaw7908

204. Haridas S, Ganapathi R, Kumar M, Manda K. Whisker dependent responsiveness of C57BL/6J mice to different behavioral test paradigms. *Behav Brain Res*. 2018;336:51-58. doi:10.1016/j.bbr.2017.08.004

205. Moy SS, Nadler JJ, Perez A, et al. Sociability and preference for social novelty in five inbred strains: an approach to assess autistic-like behavior in mice. *Genes Brain Behav*. 2004;3(5):287-302. doi:10.1111/j.1601-1848.2004.00076.x

206. Chen J, Lambo ME, Ge X, et al. A MYT1L syndrome mouse model recapitulates patient phenotypes and reveals altered brain development due to disrupted neuronal maturation. *Neuron*. 2021;109(23):3775-3792.e14. doi:10.1016/j.neuron.2021.09.009

207. Maloney SE, Sarafinovska S, Weichselbaum C, et al. A comprehensive assay of social motivation reveals sex-differential roles of ASC-associated genes and oxytocin. Published online May 21, 2022:2022.05.21.492918. doi:10.1101/2022.05.21.492918

208. TAY ESE, GUVEN KL, SUBRAMANIAM N, et al. Regulation of alternative splicing of Gtf2ird1 and its impact on slow muscle promoter activity. *Biochem J*. 2003;374(2):359-367. doi:10.1042/bj20030189

209. Ebert PJ, Campbell DB, Levitt P. Bacterial artificial chromosome transgenic analysis of dynamic expression patterns of regulator of G-protein signaling 4 during development. I. Cerebral cortex. *Neuroscience*. 2006;142(4):1145-1161. doi:10.1016/j.neuroscience.2006.08.011

210. Maloney SE, Chandler KC, Anastasaki C, Rieger MA, Gutmann DH, Dougherty JD. Characterization of early communicative behavior in mouse models of neurofibromatosis type 1. *Autism Res*. 2018;11(1):44-58.

211. Chevallier C, Kohls G, Troiani V, Brodkin ES, Schultz RT. The social motivation theory of autism. *Trends Cogn Sci*. 2012;16(4):231-239. doi:10.1016/j.tics.2012.02.007

212. Maloney SE, Yuede CM, Creeley CE, et al. Repeated neonatal isoflurane exposures in the mouse induce apoptotic degenerative changes in the brain and relatively mild long-term behavioral deficits. *Sci Rep*. 2019;9(1):2779. doi:10.1038/s41598-019-39174-6

213. Yuede CM, Olney JW, Creeley CE. Developmental Neurotoxicity of Alcohol and Anesthetic Drugs Is Augmented by Co-Exposure to Caffeine. *Brain Sci*. 2013;3(3):1128-1152. doi:10.3390/brainsci3031128

214. Nilsson SR, Goodwin NL, Choong JJ, et al. Simple Behavioral Analysis (SimBA) – an open source toolkit for computer classification of complex social behaviors in experimental animals. Published online April 21, 2020:2020.04.19.049452. doi:10.1101/2020.04.19.049452

215. Nath T, Mathis A, Chen AC, Patel A, Bethge M, Mathis MW. Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nat Protoc*. 2019;14(7):2152-2176. doi:10.1038/s41596-019-0176-0

216. van Gaalen MM, Steckler T. Behavioural analysis of four mouse strains in an anxiety test battery. *Behav Brain Res*. 2000;115(1):95-106. doi:10.1016/S0166-4328(00)00240-0

217. Arakawa H, Erzurumlu RS. Role of whiskers in sensorimotor development of C57BL/6 mice. *Behav Brain Res*. 2015;287:146-155. doi:10.1016/j.bbr.2015.03.040

218. Warren RA, Zhang Q, Hoffman JR, et al. A rapid whisker-based decision underlying skilled locomotion in mice. Goldberg JH, Calabrese RL, Goldberg JH, eds. *eLife*. 2021;10:e63596. doi:10.7554/eLife.63596

219. O'Connor DH, Clack NG, Huber D, Komiyama T, Myers EW, Svoboda K. Vibrissa-Based Object Localization in Head-Fixed Mice. *J Neurosci*. 2010;30(5):1947-1967. doi:10.1523/JNEUROSCI.3762-09.2010

220. Stüttgen MC. Toward Behavioral Benchmarks for Whisker-Related Sensory Processing. *J Neurosci*. 2010;30(14):4827-4829. doi:10.1523/JNEUROSCI.0713-10.2010

221. Dominiak SE, Nashaat MA, Sehara K, Oraby H, Larkum ME, Sachdev RNS. Whisking Asymmetry Signals Motor Preparation and the Behavioral State of Mice. *J Neurosci*. 2019;39(49):9818-9830. doi:10.1523/JNEUROSCI.1809-19.2019

222. Erisken S, Vaiceliunaite A, Jurjut O, Fiorini M, Katzner S, Busse L. Effects of Locomotion Extend throughout the Mouse Early Visual System. *Curr Biol*. 2014;24(24):2899-2907. doi:10.1016/j.cub.2014.10.045

223. Jurjut O, Georgieva P, Busse L, Katzner S. Learning Enhances Sensory Processing in Mouse V1 before Improving Behavior. *J Neurosci*. 2017;37(27):6460-6474. doi:10.1523/JNEUROSCI.3485-16.2017

224. Tillerson JL, Caudle WM, Parent JM, Gong C, Schallert T, Miller GW. Olfactory discrimination deficits in mice lacking the dopamine transporter or the D2 dopamine receptor. *Behav Brain Res*. 2006;172(1):97-105. doi:10.1016/j.bbr.2006.04.025

225. Siemann JK, Muller CL, Bamberger G, Allison JD, Veenstra-VanderWeele J, Wallace MT. A novel behavioral paradigm to assess multisensory processing in mice. *Front Behav*

Neurosci. 2015;8. Accessed July 18, 2022. https://www.frontiersin.org/articles/10.3389/fnbeh.2014.00456

226. Bader PL, Faizi M, Kim LH, et al. Mouse model of Timothy syndrome recapitulates triad of autistic traits. *Proc Natl Acad Sci*. 2011;108(37):15432-15437.

227. Chadman KK, Gong S, Scattoni ML, et al. Minimal aberrant behavioral phenotypes of neuroligin-3 R451C knockin mice. *Autism Res*. 2008;1(3):147-158.

228. Copping NA, Berg EL, Foley GM, et al. Touchscreen learning deficits and normal social approach behavior in the Shank3B model of Phelan–McDermid syndrome and autism. *Neuroscience*. 2017;345:155-165.

229. Dougherty JD, Maloney SE, Wozniak DF, et al. The disruption of Celf6, a gene identified by translational profiling of serotonergic neurons, results in autism-related behaviors. *J Neurosci*. 2013;33(7):2732-2753.

230. Feyder M, Karlsson RM, Mathur P, et al. Association of mouse Dlg4 (PSD-95) gene deletion and human DLG4 gene variation with phenotypes relevant to autism spectrum disorders and Williams' syndrome. *Am J Psychiatry*. 2010;167(12):1508-1517.

231. Grabrucker S, Boeckers TM, Grabrucker AM. Gender dependent evaluation of autism like behavior in mice exposed to prenatal zinc deficiency. *Front Behav Neurosci*. 2016;10:37.

232. Lugo JN, Swann JW, Anderson AE. Early-life seizures result in deficits in social behavior and learning. *Exp Neurol*. 2014;256:74-80.

233. Maloney SE, Akula S, Rieger MA, et al. Examining the reversibility of long-term behavioral disruptions in progeny of maternal SSRI exposure. *Eneuro*. 2018;5(4).

234. Page DT, Kuti OJ, Prestia C, Sur M. Haploinsufficiency for Pten and Serotonin transporter cooperatively influences brain size and social behavior. *Proc Natl Acad Sci*. 2009;106(6):1989-1994.

235. Peñagarikano O, Lázaro MT, Lu XH, et al. Exogenous and evoked oxytocin restores social behavior in the Cntnap2 mouse model of autism. *Sci Transl Med*. 2015;7(271):271ra8-271ra8. doi:10.1126/scitranslmed.3010257

236. Samaco RC, Mandel-Brehm C, McGraw CM, Shaw CA, McGill BE, Zoghbi HY. Crh and Oprm1 mediate anxiety-related behavior and social approach in a mouse model of MECP2 duplication syndrome. *Nat Genet*. 2012;44(2):206-211.

237. Schwartzer JJ, Careaga M, Onore CE, Rushakoff JA, Berman RF, Ashwood P. Maternal immune activation and strain specific interactions in the development of autism-like behaviors in mice. *Transl Psychiatry*. 2013;3(3):e240-e240.

238. Stoppel LJ, Kazdoba TM, Schaffler MD, et al. R-baclofen reverses cognitive deficits and improves social interactions in two lines of 16p11. 2 deletion mice. *Neuropsychopharmacology*. 2018;43(3):513-524.

239. Won H, Lee HR, Gee HY, et al. Autistic-like social behaviour in Shank2-mutant mice improved by restoring NMDA receptor function. *Nature*. 2012;486(7402):261-265.

240. Walsh JJ, Christoffel DJ, Heifets BD, et al. 5-HT release in nucleus accumbens rescues social deficits in mouse autism model. *Nature*. 2018;560(7720):589-594.

241. Lee EJ, Lee H, Huang TN, et al. Trans-synaptic zinc mobilization improves social interaction in two mouse models of autism through NMDAR activation. *Nat Commun*. 2015;6(1):1-13.

242. Nadler JJ, Moy SS, Dold G, et al. Automated apparatus for quantitation of social approach behaviors in mice. *Genes Brain Behav*. 2004;3(5):303-314. doi:10.1111/j.1601-183X.2004.00071.x

243. Molosh AI, Johnson PL, Spence JP, et al. Social learning and amygdala disruptions in Nf1 mice are rescued by blocking p21-activated kinase. *Nat Neurosci*. 2014;17(11):1583-1590.

244. Smith GD, White J, Lugo JN. Superimposing Status Epilepticus on Neuron Subset-Specific PTEN Haploinsufficient and Wild Type Mice Results in Long-term Changes in Behavior. *Sci Rep*. 2016;6. doi:10.1038/srep36559

245. Zhou Y, Kaiser T, Monteiro P, et al. Mice with Shank3 Mutations Associated with ASD and Schizophrenia Display Both Shared and Distinct Defects. *Neuron*. 2016;89(1):147-162. doi:10.1016/j.neuron.2015.11.023

246. Nieuwenhuis S, Forstmann BU, Wagenmakers EJ. Erroneous analyses of interactions in neuroscience: a problem of significance. *Nat Neurosci*. 2011;14(9):1105-1107.

247. Kafkafi N, Agassi J, Chesler EJ, et al. Reproducibility and replicability of rodent phenotyping in preclinical studies. *Neurosci Biobehav Rev*. 2018;87:218-232.

248. Matosin N, Frank E, Engel M, Lum JS, Newell KA. Negativity towards negative results: a discussion of the disconnect between scientific worth and scientific culture. *Dis Model Mech*. 2014;7(2):171-173.

249. Filipello F, Morini R, Corradini I, et al. The microglial innate immune receptor TREM2 is required for synapse elimination and normal brain connectivity. *Immunity*. 2018;48(5):979-991.

250. Shin S, Pribiag H, Lilascharoen V, Knowland D, Wang XY, Lim BK. Drd3 signaling in the lateral septum mediates early life stress-induced social dysfunction. *Neuron*. 2018;97(1):195-208.

251. Minakova E, Lang J, Medel-Matus JS, et al. Melanotan-II reverses autistic features in a maternal immune activation mouse model of autism. *PloS One*. 2019;14(1):e0210389.

252. Beery AK, Zucker I. Oxytocin and same-sex social behavior in female meadow voles. *Neuroscience*. 2010;169(2):665-673.

253. Kopp N, McCullough K, Maloney SE, Dougherty JD. Gtf2i and Gtf2ird1 mutation are not sufficient to reproduce mouse phenotypes caused by the Williams Syndrome critical region. Published online February 24, 2019:558544. doi:10.1101/558544

254. Kumar A, Wadhawan R, Swanwick CC, Kollu R, Basu SN, Banerjee-Basu S. Animal model integration to AutDB, a genetic database for autism. *BMC Med Genomics*. 2011;4(1):1-9.

Appendix: Erroneous inference based on a lack of preference within one group: Autism, mice, and the social approach task

Kayla R. Nygaard, Susan E. Maloney, and Joseph D. Dougherty

From:
Erroneous inference based on a lack of preference within one group: Autism, mice, and the social approach task.

Nygaard KR, Maloney SE, Dougherty JD. Erroneous inference based on a lack of preference within one group: autism, mice, and the social approach task. Autism Res. 2019: 12(8):1171-1183. https://doi.org/10.1002/aur.2154

A.1 Abstract

The Social Approach Task is commonly used to identify sociability deficits when modeling liability factors for autism spectrum disorder (ASD) in mice. It was developed to expand upon existing assays to examine distinct aspects of social behavior in rodents and has become a standard component of mouse ASD-relevant phenotyping pipelines. However, there is variability in the statistical analysis and interpretation of results from this task. A common analytical approach is to conduct within-group comparisons only, and then interpret a difference in *significance* levels as if it were a group difference, without any direct comparison. As an efficient shorthand, we named this approach EWOCs: Erroneous Within-group Only Comparisons. Here, we examined the prevalence of EWOCs and used simulations to test whether this approach could produce misleading inferences. Our review of Social Approach studies of high-confidence ASD genes revealed 45% of papers sampled used only this analytical approach. Through simulations, we then demonstrate how a lack of significant difference within one group often does not correspond to a significant difference between groups, and show this erroneous interpretation increases the rate of false positives up to 25%. Finally, we define a simple solution: use an index, like a social preference score, with direct statistical comparisons between groups to identify significant differences. We also provide power calculations to guide sample size in future studies. Overall, elimination of EWOCs and adoption of direct comparisons should result in more accurate, reliable, and reproducible data interpretations from the Social Approach Task across ASD liability models.

Lay Summary

The Social Approach Task is widely used to assess social behavior in mice and is frequently used in studies modeling autism. However, reviewing published studies showed nearly half do not use correct comparisons to interpret these data. Using simulated and original data, we argue the correct statistical approach is a direct comparison of scores between groups. This simple solution should reduce false positives and improve consistency of results across studies.

A.2 Introduction

The Social Approach Task is one of the most widely used behavioral assays for investigation of mouse models of liability factors associated with autism spectrum disorder

(ASD).[226–239] The use of the Social Approach Task has both helped identify ASD liability models with good face validity and advanced our understanding of the circuitry underlying social approach deficits. Unlike reciprocal social interaction assays requiring manual scoring, the Social Approach Task is automated, making it ideal for mechanistic studies that require several experiments with different interventions or genetic models. For example, a recent study showed the role of dorsal raphe serotonergic connections to the nucleus accumbens in social approach behavior, and how stimulation of this pathway can correct social deficits in the 16p11.2 deletion model associated with ASD.[240] Another group showed NMDAR activation rescued social approach behavior in *Shank2*$^{-/-}$ and *Tbr1*$^{+/-}$ mutants.[239,241] Together, these studies highlight the value in using this task to identify pathways that contribute to social approach behavior and targets that can be further interrogated as potential pharmacotherapy candidates.

The motivation behind the development of the Social Approach Task was to improve face validity of murine social behavioral assays with regard to specific social impairments that characterize ASD.[205,242] Abnormal social approach is one such attribute of the ASD social phenotype. This task was unique in the field because it required the sociability be initiated by the test mouse. Thus, it was, and is, meant to help identify a lack of social interest in mice that may be reminiscent of the social approach deficits in humans with ASD. The typical version of this task comprises two test trials: the *sociability* trial and the *preference for social novelty* trial,[205,242] along with two preceding habituation trials. During the *sociability* trial, the test mouse can freely explore the three-chambered apparatus to investigate either a novel conspecific stimulus in a restraining container (inverted wire cup) or an empty but otherwise identical wire cup (**Figure 1A**). Likewise, in the *preference for social novelty* trial, a new stimulus mouse is added to the empty cup, and the same test mouse is then assayed to examine preference for the novel mouse over the familiar mouse. The *preference for social novelty* trial is optional and some study designs require only an investigation of sociability. In addition, there are various deviations often used, including habituation to the wire cups, a novel object placed inside the nonsocial cup, and a 24-hr intertrial interval for further social memory assessment, among others.[243–245]

The original Social Approach Task studies examined sociability and preference for social novelty in inbred mouse strains.[120,205,242] The main purpose of these studies was to establish that most mouse strains exhibit sociability, and thus comparisons in these original studies were the within-group comparisons of time spent in the chamber or time spent sniffing the stimulus mice. As these were not comparisons of two groups, the experimental design of many of these original experiments did not allow for between-subjects comparisons during analysis. Subsequently, as researchers adapted this task to compare across groups, likely consulting these original studies for experimental and statistical design, many failed to incorporate the appropriate between-subjects comparisons needed for their own experimental designs. Thus, since the task was first developed, the within-group only analysis has also been perpetuated across studies of between-group factors, such as mutation of ASD candidate genes.

While it is relatively straightforward to test significance in the Social Approach Task with only one group, where the null hypothesis is "the mouse will spend equal time with both stimuli," there is no gold-standard approach for comparing two different groups. One commonly used approach is to separately test the null hypothesis within each group, and then compare those results between groups. However, the accurate null hypothesis when comparing multiple groups is "the social preference of one group equals the other." Therefore, only considering the within-group null hypothesis would result in a flawed interpretation because the accurate null hypothesis is no longer tested. In other words, the lack of a statistically significant preference in one group is interpreted

as a statistically significant difference between groups. We labeled this approach Erroneous Within-group Only Comparisons (EWOCs).

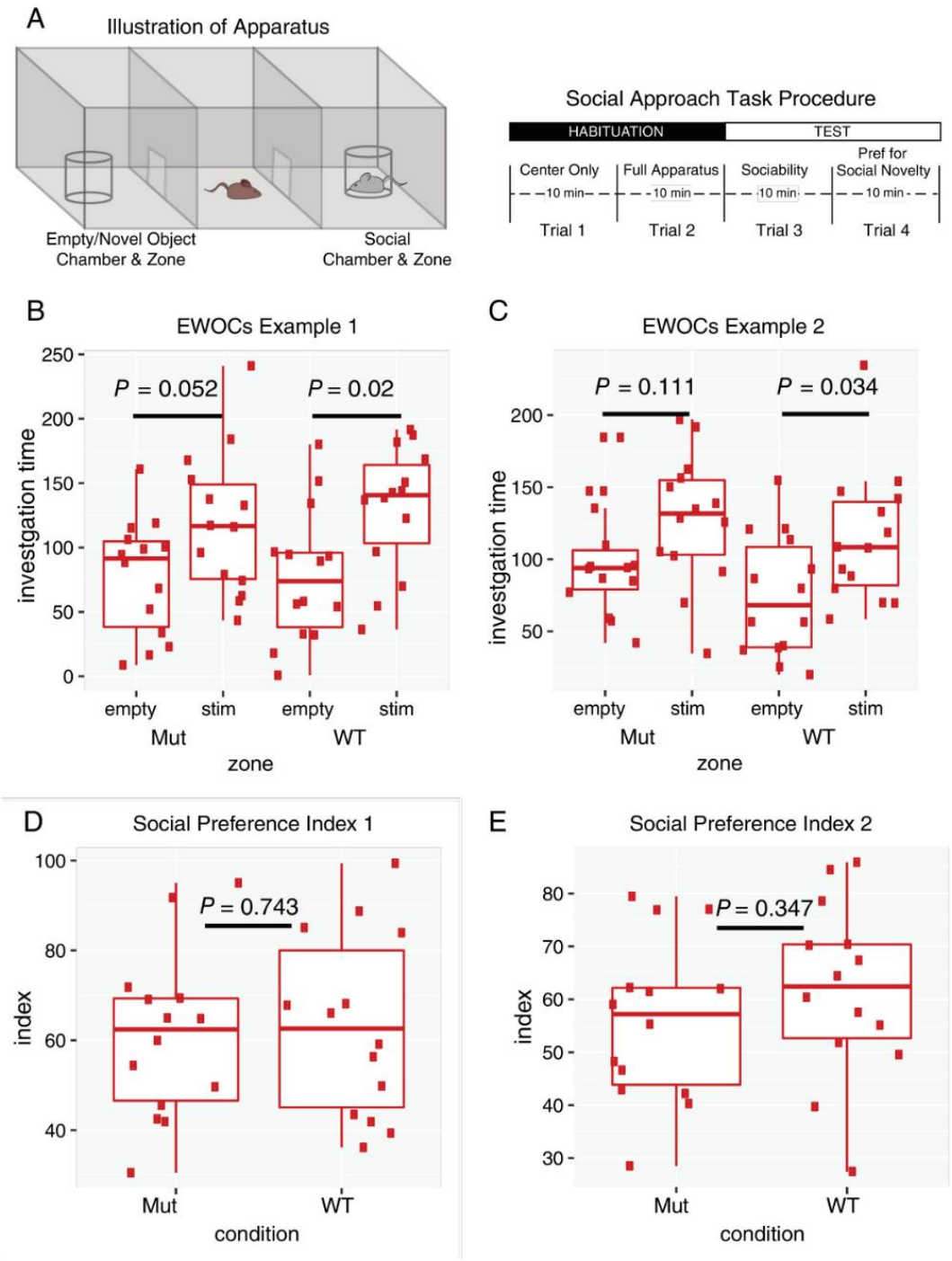

Figure 1. Illustration of the Social Approach Task and two different analytical approaches. A) Schematic of Social Approach Task apparatus and typical procedure. **B,C)** Example plots from simulated data using EWOCs. Two arbitrary groups ("Mut" and "WT") were tested for a within-group difference between the time spent with the social stimulus (stim) compared to the empty cup (empty). Only the WT group showed significant preference ($p < 0.05$), while the Mut mice did not ($p = 0.052$, or $p = 0.111$). **D,E)** Example of these same data plotted as a social preference index: $time_{[stim]} / (time_{[stim]} + time_{[empty]}) \times 100$. Direct statistical comparison of Mut to WT indices shows no significant difference ($p = 0.743$, 0.347).

To directly test between groups, a commonly used and more statistically appropriate approach is a repeated measures ANOVA with appropriate between-subjects factors to examine stimulus interaction times. A related between-groups approach is to calculate a single value summarizing social preference for each mouse for downstream statistical testing. A commonly used social preference index is $time_{[stim]} / (time_{[stim]} + time_{[empty]}) \times 100$, which results in a value from 0 (all time with the empty cup) to 100 (all time with the stimulus mouse), where 50 represents equal time with both. Indices for each mouse can then be compared across groups with a t-test, ANOVA, or appropriate nonparametric test for non-normal data. However, neither of these two approaches alone is complete. Examination of the original data is still imperative in this situation to confirm that the control group demonstrates a preference for time spent with the social stimulus cup versus the empty/novel object cup.

Here, we further demonstrate why EWOCs should not be applied to identify a difference between groups in the Social Approach Task by using data simulations to show how EWOCs can be misleading. We also review recent mouse literature to characterize the widespread use of EWOCs. We further show how direct comparison of an index, like a social preference score, between groups may reduce false positives and improve consistency of results across studies, and provide power estimates, parameterized in data from more than 400 mice, to guide future studies. Finally, we present a standardized rubric for the analysis of the Social Approach Task between groups. We believe that elimination of EWOCs from practice, and adoption of a standardized approach, will result in more robust and reproducible social approach findings when modeling ASD liability factors in mice.

A.3 Results

A.3.1 Interpreting EWOCs as a Difference Between Groups Is Fundamentally Flawed Logic

We first present a simple illustration from simulated data to demonstrate how a within-group only comparative approach to analysis could lead to erroneous inference (**Figure 1B**). In these simulated data of a *sociability* trial, the mutant mice do not show a statistically significant social preference, with $p = 0.052$. As this exceeds the critical alpha cutoff of 0.05, it does not result in a rejection of the null hypothesis. The WT mice, however, reach $p = 0.02$, which passes the cutoff. The null hypothesis is rejected, and the WT mice are considered to have shown a statistically significant preference for the social stimulus. Even though the outcome of the tests within the groups are different for mutant and WT mice, does this mean there is a significant difference in the social preference between these groups or is it a false positive? In this example, where p-values are just on either side of the threshold, it becomes obvious that a separate statistical test is necessary to determine if the groups themselves are statistically different. Indeed, calculating a social preference index and comparing them directly for these same data reveal that there is no difference between the groups (**Figure 1D**). However, in an alternate scenario, where WT mice exceed the critical alpha with $p = 0.034$ but mutants only reach $p = 0.111$ (**Figure 1C**), it may not be obvious, despite the appropriate statistical test revealing there is no significant difference in this case either (**Figure 1E**). To reiterate, a lack of difference in time spent with each stimulus within one group does not indicate a significant difference in sociability between the groups.

Unfortunately, this simple statistical misinterpretation exists widely in the neuroscience literature and is applied to many kinds of experiments.[246] It also exists in key papers evaluating genetic mouse models of ASD liability. In the studies reviewed from the SFARI database, EWOCs

were employed in 13 of 29 (44.8%) studies showing a phenotype in the *sociability* trial and 11 of the 25 (44.0%) studies that also included the *preference for social novelty* trial. Thus, the use of EWOCs is widespread.

This raises important questions: To what extent might these represent false positive results? Could widespread use of EWOCs account for why there are such challenges in finding reproducible phenotypes in behavioral models?[247] In order to determine how vulnerable this approach is to false positive interpretations, we conducted extensive simulation studies as detailed below.

A.3.2 Simulations Demonstrate EWOCs Result in an Elevated Rate of False Positives, Dependent on Sample Number

We first modeled how likely false positive results would be when using EWOCs. To base the simulation on real parameters, we examined social approach data from all mice previously tested in our lab to identify typical mean interaction times and SDs. We also extracted the data examined in all 29 data sets from the reviewed papers for comparison. We found the median group size was n = 16 across the 29 papers (**Figure 2A**), with studies ranging from 6 to 30. We then generated random data for two groups with no true difference in their social preference (drawing from the same normal distribution) such that both "WT" and "Mut" groups should have a 1.5-fold preference for the social stimulus over the empty cup (social preference index = 60; **Figure 2B**). We then systematically varied the n in each group from 5 to 30 and conducted 10 simulations of 1000 studies at each n. When we simulated n at the median of published studies (i.e., 16 per group), we observed a false positive rate of 25% using EWOCs (**Figure 2C**). Specifically, a false positive result is when the conclusion is that the two groups are different (**Figure 1B,C**), since in these simulated data the two groups were drawn from the same distribution. Even extending n to 25, we still observed a false positive rate of 10%, which is two times higher than the false positive rate of 0.05 that is the standard accepted critical alpha in the field. Note that a solution for controlling the false positive rate is quite simple: a t-test assessing the social preference index, with $p < 0.05$ critical alpha cutoff, results in the false positive rate of 5%, regardless of n (**Figure 2D**). Similar results are also achieved if one analyzes the stimulus interaction times across groups using a mixed ANOVA with between- and within-subjects simple main effects following significant interaction terms (*not shown*). Importantly, if n is imbalanced, then statistical power is also imbalanced. For example, sometimes mutants are harder to generate than WTs (indeed, one-third of the reviewed studies had smaller mutant than WT groups). This might further inflate false positive rates when EWOCs are used. By varying n for "Mut" but keeping n = 12 for "WT," we show this is the case (**Figure 2E**). Again, this can be corrected by directly comparing groups statistically (**Figure 2F**).

It is worth noting that even with equal n, other results can also occur. For example, if WT and mutant mice are truly not different, there is an equal chance that the "Mut" mice will show a significant preference for the social stimulus in the same trial that the "WT" mice do not (**Figure 2B,C**; purple lines). There is also a chance, especially at low n, that neither group will show a significant within-group result (**Figure 2B,C**; green lines). Given the known bias in published literature for positive over negative results,[248] it is likely that either of these possibilities are underreported in the literature. For example, they may simply be considered failed trials by the experimenters and repeated, since the positive control (i.e., a preference for the stimulus mouse in the WT group) did not work. One danger of this repeated EWOCs approach is that it could further increase the possibility of a false positive, as the experiment would be repeated until the outcome is either both groups are social, or only the mutants have a deficit. Overall, even with a single

experiment of simulated data at n = 16, there is only a <70% chance of correctly identifying both groups as social.

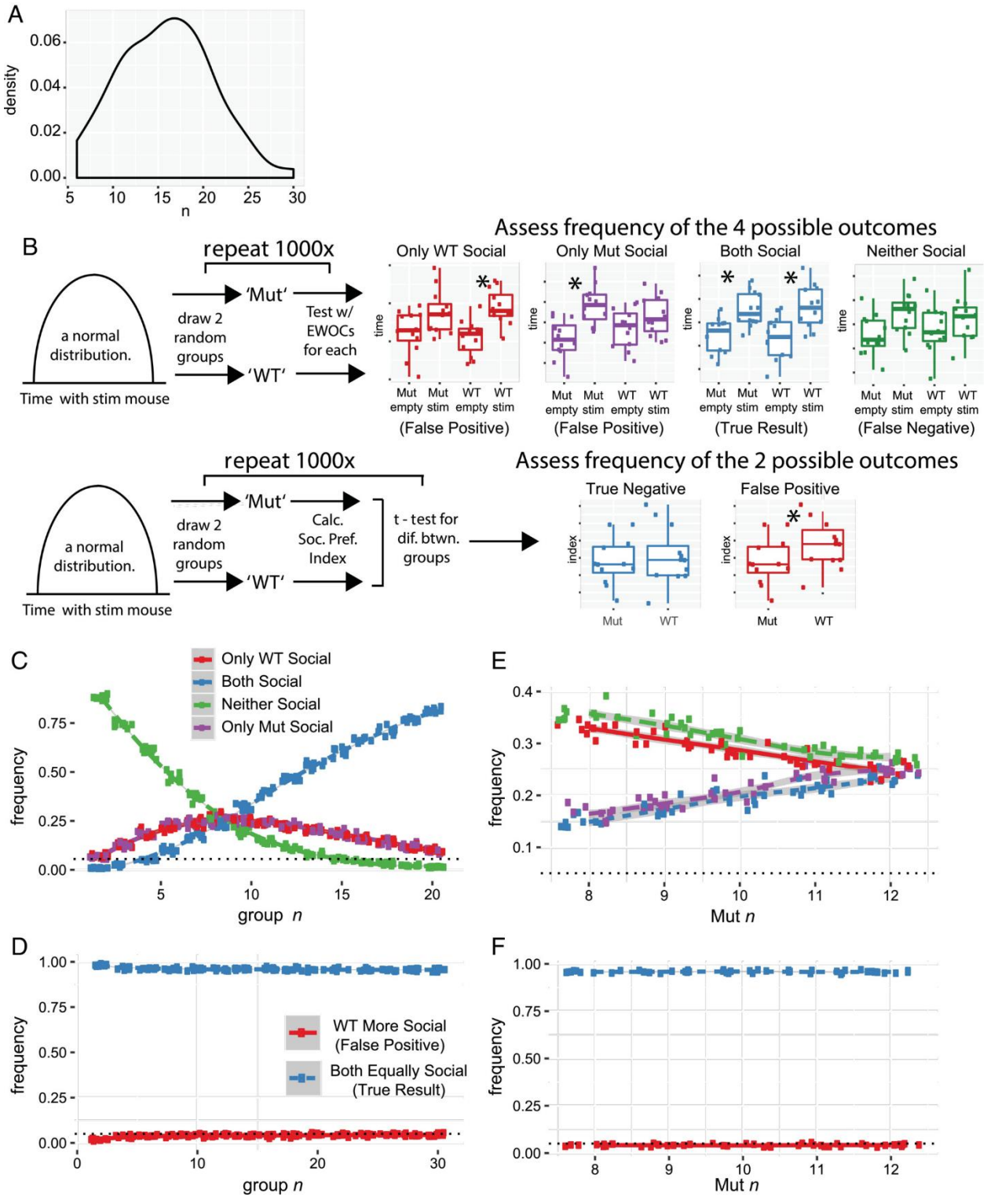

Figure 2. Using EWOCs can result in substantially elevated false positive rates, especially at low sample sizes. A) Distribution of group sizes (combined for genotype) across 77 groups in 29 papers. **B)** Cartoon of simulations and possible outcomes. Two

groups ("Mut" and "WT") are drawn from the same distribution with identical social preference magnitude, and then tested with EWOCs (*upper panel*) or a social preference index (*lower panel*). **C)** Plot of simulations results as function of n, after 10 × 1000 simulated experiments for each n, drawing two groups from the same distribution and analyzing with EWOCs. The true result is both groups are social (*blue*), so incorrect conclusions were drawn a substantial proportion of the time. **D)** Plot of t-test on social preference index, showing false positive rate as a function of n. **E)** Simulation plot as a function of imbalanced n with "WT" n = 12, and "Mut" n varied from 8 to 12, using EWOCs. **F)** Simulation plot as a function of imbalanced n, using t-test on the social preference index.

A.3.3 Simulation Demonstrates EWOCs False Positive Rates Are Also Influenced by Magnitude of Social Preference

Of course, statistical power is also a function of effect size—in this case, the magnitude of the social preference. In our first model, we assumed a 1.5-fold preference for the stimulus mouse over the empty cage, modeling a normal distribution with a mean interaction time of 126 sec with the stimulus mouse and 86 sec with the empty cup (giving a social preference index of 60). While this is a plausible social preference magnitude, and slightly higher than the mean we saw in our reanalyzed mice (124.06), it is a bit below the median social preference index of published groups (64.41 [58.96–69.70 interquartile range (IQR)]; across all 77 groups of extractable data from the 29 studies; **Figure 3A**). Therefore, we also fixed n at ten and varied the simulated preference of all mice for the social stimulus. This showed a high rate of erroneous inference resulting from EWOCs. Interestingly, a social preference index around 64 was particularly vulnerable to EWOCs false positive interpretation (**Figure 3B**), with rates at nearly 25%. Note, differences in effect size are also readily controlled by appropriately comparing the two groups statistically (**Figure 3C**).

Also worth discussion is the possibility that the published median social preference magnitude is slightly inflated compared to the actual social preference, again, because of the bias toward publication of positive results. Indeed, if we plot the social preference index of the last 421 mice analyzed in our lab (Fig. 3D), published or not, we see a median preference of 58.95 (48.95–68.48 IQR) for the *sociability* trial, and 63.49 (51.69–71.64 IQR) for the mice that were also tested in the *preference for social novelty* trial (n = 325, not shown). We also noticed a commonly used inbred strain (FVB/AntJ, e.g., the standard background strain of FMRP mutants) showed a marginally lower social preference index than the more ubiquitous inbred C57BL/6J strain (54.8 vs. 60.1, Welch's t-test t = 2.3128, p = 0.023, df = 107.03), and, generally, males showed a higher social preference index than females across strains (60.98 vs. 55.04, t = 3.9615, p = 8.7E-5, df = 418.72).

Thus, the expected magnitude of social preference in this task may vary by sex and strain, and may be low enough to warrant increased n when using both sexes for experiments, which is an important practice, and currently required by National Institutes of Health funding, for many reasons, including the sexually dimorphic nature of various diseases.

Therefore, as a resource, we have estimated the number of animals required to have well-powered studies detecting an absence of social preference (i.e., social preference index of 50 or less) in a mutant group compared to a variety of potential WT group preference index levels. Our estimates show that to have 80% power to detect a significant effect requires approximately 30 animals per group using both sexes of C57BL/6J mice, and possibly substantially more with other strains (**Figure 3E–G**), though such strains may be better when assaying manipulations that increase sociability. Further, *social novelty* trials, where the effect size is typically somewhat larger, would require fewer animals. Finally, these power calculations highlight the nuance of interpreting a negative result even with correct between-group comparisons (especially reanalyzing historic data with smaller n): a $p > 0.05$ can always mean the effect of the mutation could simply be too small to see reliably given the group sizes used in a particular study.

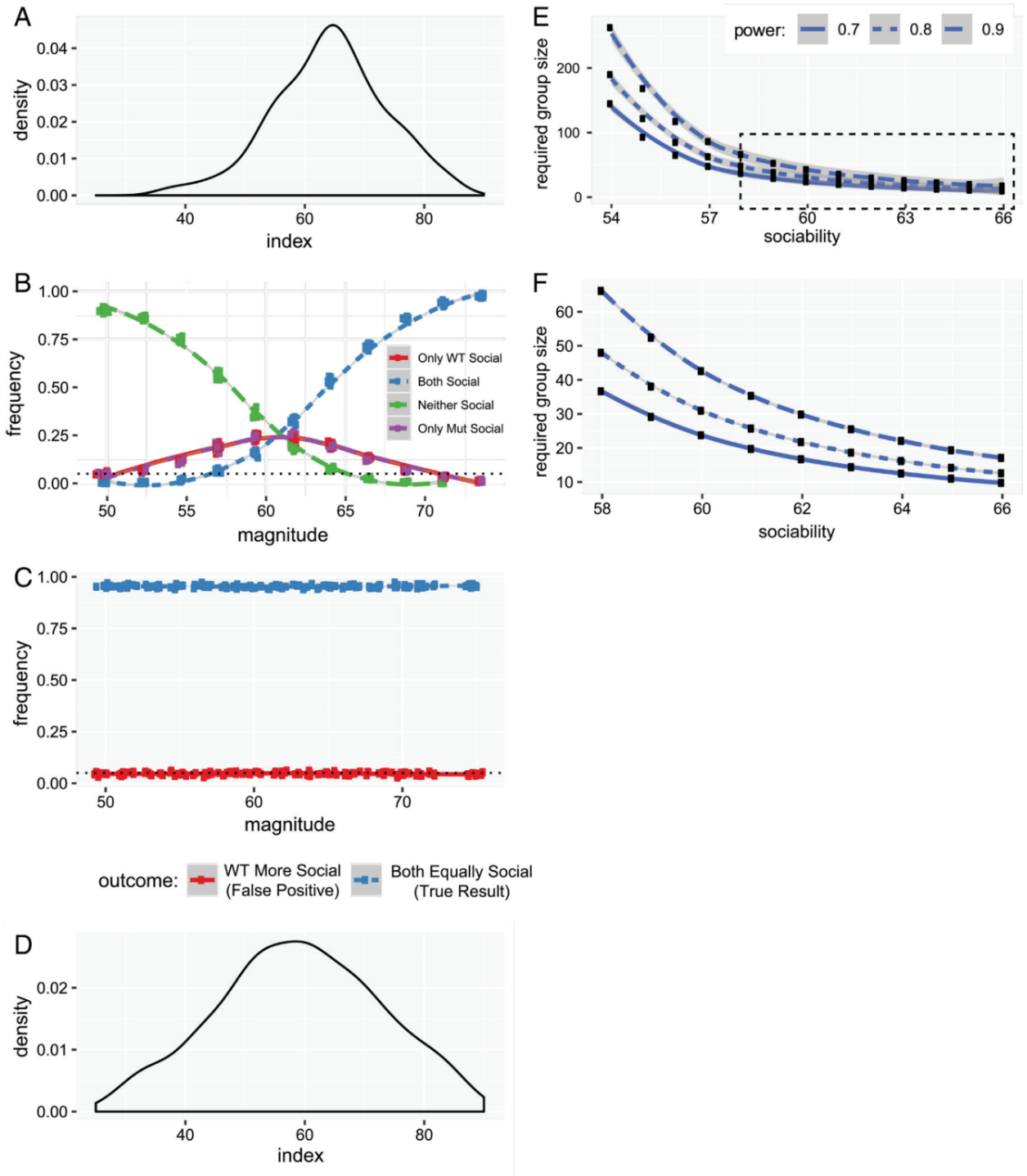

Figure 3. Elevation of false positive rates depends on the magnitude of the social preference when EWOCs are used. A) Distributions of average magnitudes of social preference indices across groups from the 29 reviewed studies. **B)** Plot of outcomes as a function of social preference magnitude when using EWOCs. **C)** Plot of false positive rate as a function of social preference magnitude when using t-test on social preference index. **D)** Distributions of magnitudes of social preference indices from all mice run in our lab (n = 421). **E)** Power calculations showing required n per group as a function of the WT social preference index, to have 70%, 80%, or 90% power to detect a difference at 0.05. **F)** Same, replotting boxed region from **E).**

A.3.4 Simulation Demonstrates that Behavioral Disruptions that Increase Variance in Mutants Will Also Lead to Higher False Positive Rates with EWOCs

Finally, there are even more subtle features of mouse behavior that might lead to inflated false positive rates with EWOCs. This is because commonly used test statistics are defined as the difference in the means divided by a measure of variance. Thus, if one group is significantly *more* variable than another, it is *less* likely to have a large test statistic and thus *less* likely to achieve a significant p-value.

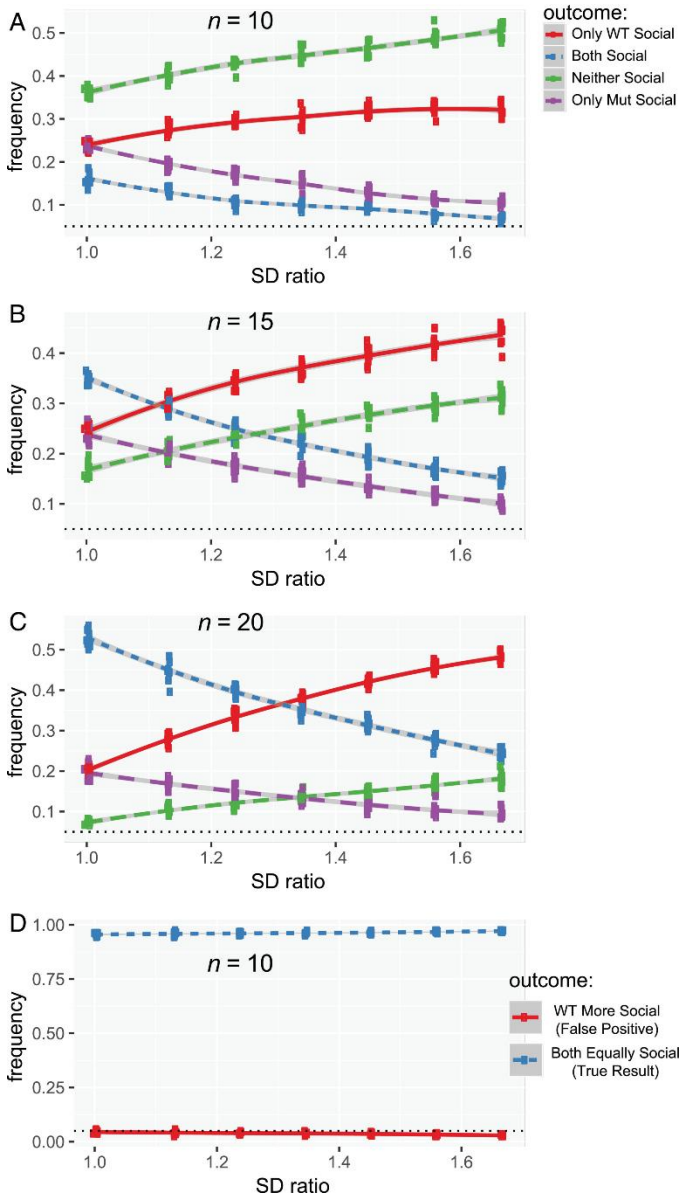

Figure 4. Increased variance in mutants can also lead to inflated false positive rates when EWOCs are used. A) Plot of false positive results when using EWOCs as a function of increased variance in only Mut at n = 10, **B)** at n = 15, **C)** at n = 20. **D)** Plot of t-test false positive rate as a function of increased variance at n = 10. SD ratio: the ratio of the Mut to the WT standard deviation (SD; varied from 1 to 1.5).

For example, if mutant mice tend to have a compulsive grooming phenotype making their movement in the task more stochastic (i.e., they might spontaneously enter a long bout of compulsive grooming), then their variance might simply be higher in this task compared to controls. It is hard to determine how frequently such a thing might be occurring in the literature, but it is straightforward to model—holding a constant n (10) and social preference index (60), we altered the variance of the distribution from which we drew the "Mut," but not the "WT," group. This profoundly decreased the ability to detect a significant social preference in the "Mut" group (**Figure 4A**), and, interestingly, this phenomenon could not be readily rescued by increasing n (**Figure 4B,C**). Thus, mutations that increase variability in mouse behavior, when using EWOCs, can mask true social preference. Again, when you directly compare groups statistically, the false positive rate stays at a well-controlled 5% (**Figure 4D**).

To demonstrate that the flawed logic of EWOCs extend to chamber time data, as well, we duplicated all our above analyses using simulations based on means and SDs extracted from a published paper using chamber time instead of investigation zone time.[249] The results were substantially similar (*data not shown*). This further indicates the results of our simulations were robust across parameters derived from multiple groups.

A.4 Discussion

The Social Approach Task has been heavily relied on to assess social behavior phenotypes in genetic liability factors for ASD. Thus, it is essential to use appropriate statistical approaches to ensure proper interpretation of the results. Only this will allow for correct conclusions to be drawn about the influence of ASD candidate genes and other liability factors on social approach circuits.

In almost half of published papers based on our sampling, the interpretation of results of this task was based on within-group only comparisons without a direct comparison between the experimental and control groups. Thus, EWOCs are frequently interpreted as a difference between groups. The problem with using this approach, essentially concluding that "if the result is not significant, sociability is absent," is that statistical tests are designed only to identify significant differences. They are *not* designed to identify a significant *lack* of differences. In other words, the correct interpretation when $p > 0.05$ is not "We are 95% confident there is no difference in preference between the mouse and the cup." It is "We are *not* 95% confident that there *is* a difference between the mouse and the cup." Statistical tests would have to be completely redesigned to be able to state with 95% confidence that there is no preference, and it is far simpler to directly compare the relevant groups with standard tests. We refer the reader back to the example in Figure 1B illustrating how EWOCs do not hold up against a direct comparison between groups. Of course, when the p-value of the mutant group is presented and shown to be very close to 0.05, the logical flaw becomes more evident and many scientists would interpret their own findings with caution, even if using EWOCs. But consider alternate scenarios where WT mice were perhaps $p < 0.04$ and mutants were $p < 0.12$ (**Figure 1C**). Often a result of $p < 0.12$ would not be considered approaching significance and would not be shown. Yet this result could equally fairly be stated as "We are 96% certain that the wild-type mice are social, and 88% certain that the mutant mice are social." Expressed this way, few scientists would be confident that the mutant mice have a significant social deficit.

It could be argued that sociability in this task should be considered a binary outcome measure rather than a quantitative trait. Yet, evidence suggests this is not a categorical phenotype and these data are indeed continuous. Multiple studies have now shown that typical sociability can be heightened following stimulation of different pathways in the brain.[240,250] For example,

optogenetic stimulation of the dorsal raphe neurons or their fibers in the nucleus accumbens increased the social preference index in WT mice.[240] Pharmacological agents have also shown promise as a means to ameliorating abnormal social approach behaviors. It was recently shown that Melanotan-II, a melanocortin receptor 4 agonist that stimulates oxytocin activity, corrected the social approach deficits in male mice of the maternal immune activation model.[251] Thus, to better screen for treatment effects in this task, which are likely to be quantitative and not qualitative, it is valuable to analyze social approach as continuous. Clearly, this phenotype has a range that can be altered and deserves appropriate quantification. We have tried to make the argument here that directly comparing groups using an index, such as a social preference score, creates a suitably quantitative design, provided sufficient n is used, to overcome variability inherent in mouse behavior.

Furthermore, we have included power analyses to help guide the selection of sample sizes that will be needed to confidently overcome this variability. These sample sizes also assume a complete loss of sociability in the mutants. If the phenotype is only partial, sample size would have to be correspondingly higher. Nonetheless, while the sample size required in C57BL/6J is substantially higher than often used (**Figure 2A**), it is still reasonably achievable. However, the very high sample size required in some combinations of sex and strain suggests that considering new variations of the method that further automate the task, or that collect more repeated measures of the same mice to reduce the per mouse variance, could offer pragmatic solutions to improving power. Indeed, it is interesting that the *social novelty* trial is better powered (because of its larger effect size) than the *sociability* trial. Since the *preference for social novelty* trial is typically run with the same mice after they have experienced the *sociability* trial, it might be that further exposing the same mice to the Social Approach Task over multiple days allows for better estimates of the social preference of each, enabling studies that do not require as large of a sample size.

In our review of studies investigating high confidence ASD genes, almost half of studies we examined used a flawed statistical logic to interpret the Social Approach Task results. Of these studies, 85% (11/13) concluded that the mutation impaired social behavior, and it is worrying that a substantial fraction of these might be false positives. Yet, despite the flawed statistical approach, it is possible these studies would truly show a difference between mutant and controls if these data were analyzed with an appropriate between-subjects design. For the authors with primary data, it may be worth assessing whether this is the case. For example, in one of our prior publications, along with the standard paradigm, we employed a variation of the task we hypothesized might be more sensitive to measure preference for social novelty (cagemate vs. novel conspecific).[229] We also examined time spent investigating a cagemate vs. an empty cup. We encountered an odd situation in which the mutant mice showed a significant preference for the cagemate, whereas the control mice did not. We interpreted these within-subject differences as no deficits in sociability toward a cagemate in the mutant mice given that there were no between-subjects differences in time with the cagemate or empty cups. However, while we conducted a full repeated measures ANOVA design that included between-group simple main effects, we did not provide those results and explicitly state that the between-subjects comparisons were nonsignificant, thus creating ambiguity in the interpretation of our results. Therefore, here we conducted a reanalysis of these data using the preference score. This provides clear evidence that there was no difference between genotypes for sociability toward a cagemate (control: $M = 55.48$, $SD = 9.96$; Mutant: $M = 62.72$, $SD = 13.38$; $t(16) = 1.226$, $p = 0.238$). We provide this example of our own data to demonstrate how ambiguous studies can be quickly reanalyzed for clarity. Similarly, another published study, from which we drew simulation parameters,[249] was able to rapidly

analyze their data and confirm a between-group difference in their mutants (Matteloi, *personal communication*). Other key studies that used EWOCs may benefit from corrigendums or preprint postings clarifying the results when these data are reanalyzed using direct statistical comparisons between groups. If prior studies were actually not significant, it could have important implications on future studies involving these ASD liability genes.

It is worth noting that the use of a social preference index is only valid if used in combination with some analysis of original data as well. Exclusive use of a preference score could also lead to flawed conclusions under some circumstances. For example, without confirmation of a preference for time spent with the social stimulus cup vs. the empty/novel object cup in the control group, a direct comparison of a social preference index between controls and the experimental group is meaningless; if there is not a within-group preference detected with a reasonable n of control animals, this may indicate some problem in the execution of the task. Likewise, the absolute time values of both groups are also important to examine during data analysis. There may be an instance in which the social preference index is not different between groups, but the absolute time spent with the stimuli is greatly reduced or increased in the experimental group. A clear example of this can be found in Lee *et al.* (2015), in which the greatly reduced absolute investigation times in *Shank2* homozygous mutants was found to be due to motor stereotypies.[241] This interesting phenotype may not have been detected if only the social preference index was examined. Visual investigations of absolute time plots and additional analysis with a repeated measure ANOVA should always be part of the analytical pipeline of these data.

To provide a standardized rubric, we have included a decision tree (**Figure 5**) that schematizes what we think is the best approach to analyze data from the Social Approach Task. This includes a repeated measures ANOVA at the apex of the tree. The preference index should be in *addition* to a full factorial repeated measures (mixed model) ANOVA as a substitution for erroneous interpretation of multiple within-subjects comparisons but not as a substitution for examination of the original data. We have provided a sample script (https://bitbucket.org/jdlabteam/ewocs/src/master/social_approach_analysis_files/) for SPSS code implementing such an analysis to facilitate adoption by the field.

While we have highlighted the occurrence of EWOCs with regard to this one assay, this flaw certainly has been seen in a variety of other experiments in the past,[246] and the same erroneous logic could easily be applied to a variety of other experiments in behavior (e.g., novel object recognition task) and beyond. A very similar paradigm in voles, the partner preference task, is easily susceptible to a similarly flawed approach to analysis, and preference indices are being used more frequently in this field, as well.[252] We have been very deliberate in developing a novel term as we hope that providing a simple name for the phenomenon ("EWOCs") will aid in rapid recognition of this flaw when it occurs. More importantly, we hope the presentation of a simple solution (direct statistical comparisons) will encourage authors, editors, and reviewers to root out this kind of inference from the literature generally, and from this assay specifically.

Excellent standardized behavioral assays are essential for assessing face validity of mouse models of ASD liability and discovering new therapeutic options. A vital aspect of the validity and reliability of an assay is appropriate interpretation of its data, which requires the correct statistical approaches. The Social Approach Task is a valuable tool to assess mouse social approach behavior, one domain that could be related to the abnormal social phenotype in ASD. As such, it has been used extensively over the last 14 years and will likely continue to be frequently applied to various mouse models. Our hope, moving forward, is to begin to apply more appropriate statistical analyses to Social Approach Task data so that accurate, reliable, and reproducible conclusions are

drawn across ASD liability models. This will allow the ASD research community to move forward confidently with studies of new therapeutic strategies based on convincing and concrete results.

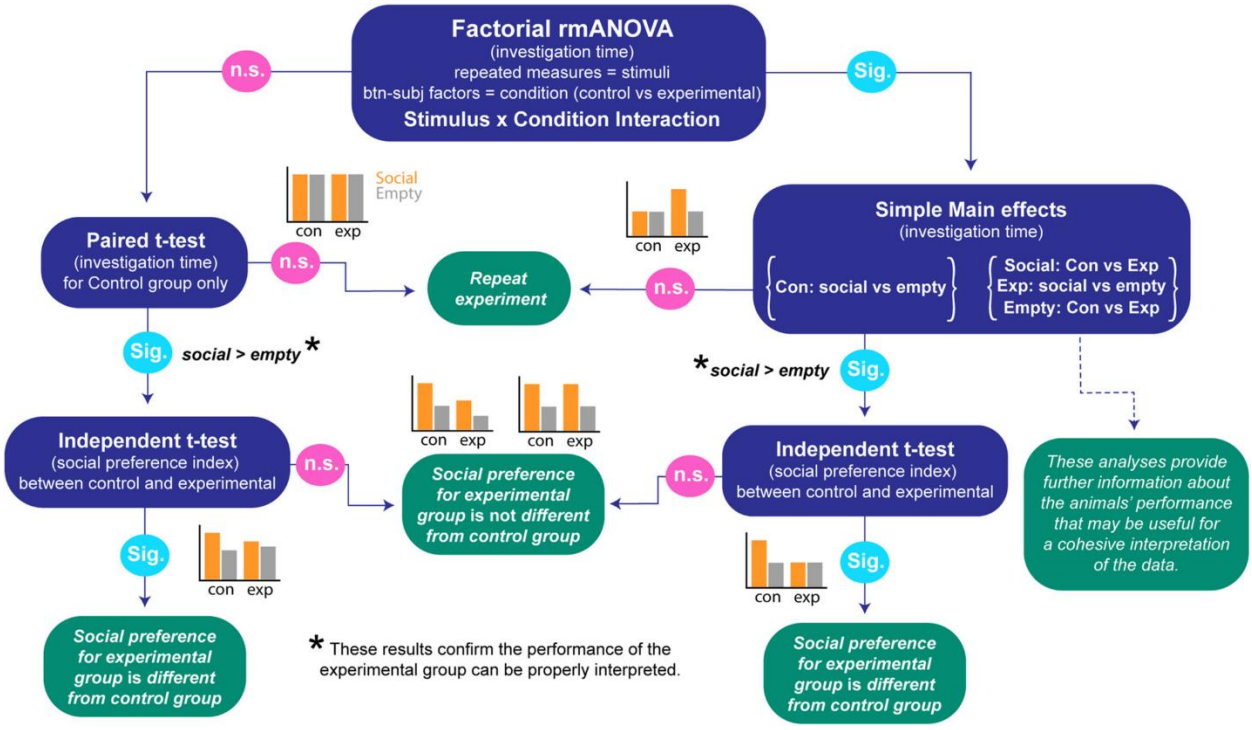

Figure 5. Social Approach Task data analysis decision tree. A decision tree schematizing a statistical pathway for Social Approach Task data analysis, provided data are normal and meet the other assumptions of univariate analysis. The blue bubbles present statistical tests with dependent variable of interest in parentheses. The green bubbles present interpretations of the test results. Sig., significant; n.s., nonsignificant. Example graphs provide representations of possible data for each outcome (con, control group; exp, experimental group).

A.5 Methods

A.5.1 Simulation Studies

We conducted multiple analyses of simulated data to explore the frequency of erroneous inferences when using only EWOCs to determine a difference between groups. First, we collected all Social Approach data previously generated in the lab, which includes 217 mice previously published and an additional 204 mice subsequently tested (see **Table 1** for descriptive data).[229,233] Using these data, we calculated the mean interaction time in seconds with the stimulus mouse ($time_{[stim]}$; 124.06 ± 52.90 [standard deviation, SD]), and the mean time with the empty cup ($time_{[empty]}$; 87.51 ± 40.59 [SD]). We then wrote a simple function in R to generate 1000 random experiments with a sample size of 10 per group using the function rnorm to sample two arbitrary groups ("Mut" and "WT") from the same normal distribution with parameters derived from data above ($time_{[stim]} = 124$ sec, $time_{[empty]} = 88$ sec, $SD = 47$ sec). Using this function, we calculated the frequency of incorrect interpretations when using EWOCs (conducting separate t-tests comparing $time_{[stim]}$ to $time_{[empty]}$ for Mut and WT groups and comparing the results) and repeated the thousand-experiment simulation ten times. Incorrect interpretations are any results that do not reveal both groups to have a social preference (e.g., both groups are not social, only Mut is social, or only WT is social). Second, we repeated this method and systematically varied the group sample size (n) from 2 to 30 to illustrate the vulnerability of EWOCs to false positives across n, and what

happens when n is mismatched between groups. In this case, a false positive is the conclusion that the experimental group (Mut) is significantly different from the control group (WT), despite the fact that preference data for both groups were drawn from the same distribution and, thus, an appropriate statistical test would reveal they do not significantly differ 95% of the time. Third, we modeled the consequences of varying the magnitude of social preference by changing the mean of the sampled normal distributions across a range of values. We set indices for a range of social preference values from 50 (no preference) to 75 (a threefold preference for the stimulus mouse) by setting values of $time_{[stim]}$ from 106 to 159 sec (and correspondingly adjusted the mean for $time_{[empty]}$). Fourth, we modeled the effect of differential group variability by increasing the *SD* of only the Mut group from 47 to 78 but keeping the mean preferences the same for both groups.

Table 1. Descriptive Statistics for Simulation Analyses Data Collected in the Dougherty Laboratory

Total sample size	Sex distribution		Grouping distribution		Background strain	Reference
	Females	Males	Experimental	Control		
20	0	20	11	9	C57BL/6J	Dougherty *et al.*[229]
197	99	98	113	84	C57BL/6J	Maloney *et al.*[233]
121	69	52	75	46	Hybrid C57BL/6J x FVB	Kopp *et al.*[253]
69	38	31	51	18	FVB	Unpublished
14	7	7	0	14	C57BL/6J	Unpublished
Total: 421	213	208	250	171	--	--

We then repeated all the above analyses but calculated the frequency of erroneous inferences when transforming the time into a social preference index, defined as $time_{[stim]} / (time_{[stim]} + time_{[empty]}) \times 100$, and then conducted a t-test comparing indices of the two groups. In addition, we duplicated all our analyses using simulations based on parameters extracted from a published paper,[249] using chamber time instead of investigation zone time, which yielded substantially similar conclusions.

A.5.2 Systematic Review of the Literature

To assess the potential impact of EWOCs in ASD-related research, we systematically reviewed the literature referenced in the SFARI Animal Models database[254] (accessed July 18, 2018) for genes with a score of 1, classified as high confidence. We further limited this to the 29 papers that used the Social Approach Task, including both the *sociability* (all 29 papers) and *preference for social novelty* (a subset of 25 papers) trials. From these papers, we extracted the results for the *sociability* and *preference for social novelty* trials, sample size, and whether EWOCs were used. If a study used both within-group and between-group comparisons, it was not counted as an EWOCs study. Finally, an independent researcher reread all studies to confirm only this interpretation was used.

A.5.3 Power Calculations

We estimated the required group sizes with the pwr.t.test function in R, using the settings for two samples with a one-tailed hypothesis, where the direction of the effect is predicted prior to the study. We ran the algorithm for three magnitudes of power (0.7, 0.8, and 0.9) and systematically varied the effect size across a range of plausible values. We parameterized our calculation of effect size (Cohen's d) using values based on the 421 mice from our lab. Specifically, we set the pooled SD for the social preference index at 15.64 (the SD of our mice), calculated effect sizes assuming a mutant group would have no social preference (a group mean of 50), and varied the corresponding wild-type (WT) preference to between 54 and 66. These preference values range from below the group mean of our least social group (54.81) to slightly above our most social group (63.3) and the mean of the reviewed published studies (64.17). Resulting group sizes were then plotted as a function of effect size and desired power.

A.6 Acknowledgments

This work was supported by 1R01MH107515 (J.D.D.), National Institutes of Health training grant 5T32GM007067-43 (K.R.N.), and National Science Foundation Graduate Research Fellowship DGE-1745038 (K.R.N.). The authors thank John Constantino, Chris Gunter, Carla Yuede, and Joshua Dearborn for valuable discussion during the preparation of this manuscript.