Published in final edited form as:

Hum Genet. 2018 October ; 137(10): 807–815. doi:10.1007/s00439-018-1939-3.

Replication of a Rare Risk Haplotype on 1p36.33 for Autism Spectrum Disorder

N.H. Chapman¹, R.A. Bernier^{2,3}, S.J. Webb^{2,3}, J. Munson^{2,3}, E.M. Blue¹, D.-H. Chen⁴, E. Heigham⁴, W.H. Raskind^{1,2,5}, and E.M. Wijsman, Ph.D.^{1,5,6}

¹Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, WA, 98195, USA

²Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA, 98195, USA

³Center on Human Development and Disability, University of Washington, Seattle, WA, 98195, USA

⁴Department of Neurology, University of Washington, Seattle, WA, 98195, USA

⁵Department of Genome Sciences, University of Washington, Seattle, WA, 98195, USA

⁶Department of Biostatistics, University of Washington, Seattle, WA, 98195, USA

Abstract

Hundreds of genes have been implicated in autism spectrum disorders (ASDs). In genetically heterogeneous conditions, large families with multiple affected individuals provide strong evidence implicating a rare variant, and replication of the same variant in multiple families is unusual. We previously published linkage analyses and follow-up exome sequencing in 7 large families with ASDs, implicating 14 rare exome variants. These included rs200195897, which was transmitted to 4 affected individuals in 1 family. We attempted replication of those variants in the MSSNG database. MSSNG is a unique resource for replication of ASD risk loci, containing whole genome sequence (WGS) on thousands of individuals diagnosed with ASDs and family members. For each exome variant, we obtained all carriers and their relatives in MSSNG, using a TDT test to quantify evidence for transmission and association. We replicated the transmission of rs200195897 to 4 affected individuals in 3 additional families. rs200195897 was also present in 3 singleton affected individuals, and no unaffected individuals other than transmitting parents. We identified 2 additional rare variants (rs566472488 and rs185038034) transmitted with rs200195897 on 1p36.33. Sanger sequencing confirmed the presence of these variants in the original family segregating rs200195897. To our knowledge, this is the first example of a rare haplotype being transmitted with ASD in multiple families. The candidate risk variants include a missense

Direct correspondence to: Ellen Wijsman, Ph.D. wijsman@uw.edu, Phone: (206) 543-8987 Fax: (206) 616-1973, Regular Mail: Division of Medical Genetics, Box 359460, University of Washington, Seattle, WA 98195-7720, Courier Address: Statistical Genetics Lab, T15, 4333 Brooklyn Ave NE, Seattle, WA 98195-9460.

Ethical approval: "All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards."

Conflict of Interest: On behalf of all authors, the corresponding author states that there is no conflict of interest.

mutation in *SAMD11*, an intronic variant in *NOC2L*, and a regulatory region variant close to both genes. *NOC2L* is a transcription repressor, and several genes involved in transcription regulation have been previously associated with ASDs.

Keywords

autism spectrum disorder; replication; whole genome sequence; MSSNG; transcription regulation; *SAMD11*; *NOC2L*

INTRODUCTION

Autism spectrum disorders (ASDs) are behaviorally defined neurodevelopmental disorders. They are characterized by deficits in communication and social interactions, and by the presence of repetitive behaviors or restricted interests (American Psychiatric Association 2013). Individuals with ASD also have elevated risk of epilepsy (Bolton et al. 2011) and intellectual disability (Dykens and Lense 2011) compared to the general population. Including the most mildly affected individuals, ASDs affect 1 in 59 children at 8 years of age (Baio et al. 2018). Individuals with ASDs require significant support in childhood and often into adulthood, including special education, occupational therapy, and supportive living accommodation (Buescher et al. 2014). The cost of these supports over a lifetime for an individual with an ASD without intellectual disability was recently estimated at US\$1.4 million (Buescher et al. 2014).

ASDs are substantially heritable, with concordance rates between monozygotic twins of 50–90%, and between siblings of 3–26% (Berg and Geschwind 2012). The risk of ASD to subsequent siblings of children with autism is estimated at 18.7% (Ozonoff et al. 2011). Unfortunately, this relatively high heritability is not the result of a simple genetic architecture involving a small number of genes. Because of the complex biology underpinning the development of the human brain, it is likely that there are many genes that can be involved in the development of ASD. The number of genes estimated to be involved in ASDs ranges from hundreds (O’Roak et al. 2012; Sanders et al. 2011) to as high as a thousand (Sanders et al. 2012). Studies using data sets composed of singleton cases (Neale et al. 2012; O’Roak et al. 2012; Sanders et al. 2012) yield long lists of genes where potentially deleterious changes (including single nucleotide variation, structural variation, and copy number variation) are observed in individuals with ASD. Without large and well-matched control populations, it is difficult to identify which are false positives, and which are genuine risk variants (Gratten et al. 2013). Nonetheless, these findings have allowed description of protein networks and pathways that are likely related to ASDs (De Rubeis et al. 2014; Krumm et al. 2014; O’Roak et al. 2012). Three categories of genes that are consistently implicated in autism include those that (i) affect synapse formation (Gilman et al. 2011) and function (Krumm et al. 2014), or are involved in (ii) transcriptional regulation (Ben-David and Shifman 2013), or (iii) chromatin remodeling (Krumm et al. 2014).

In other highly genetically heterogeneous traits, analysis of large extended families with multiple affected subjects has been instrumental in identification of risk genes (e.g. retinitis pigmentosa (Kozma et al. 2005; Villanueva et al. 2014; Yu et al. 2016)). This approach

provides solid evidence for single genes in such a disorder. In ASDs, identification of large families is difficult, and it is unusual to find affected relatives more distantly related than siblings. This may be due to reduced reproductive fitness of individuals with ASD (Power et al. 2013), reduced penetrance or variable expressivity of genotypes containing ASD risk alleles, or the challenges inherent in measuring a behavioral phenotype. In nuclear families with two or more children affected by ASD, unaffected relatives often show the broader autism phenotype (BAP) – that is, they show some characteristics associated with ASD but not enough to merit an ASD diagnosis (Bernier et al. 2012; Piven et al. 1997). This phenotypic variation may be the result of incomplete penetrance or variable expressivity of ASD alleles. Rare variants that segregate to individuals with ASD and BAP within a family are excellent candidates for autism risk variants, and identify autism risk genes (Chapman et al. 2015; Cukier et al. 2014). Because of the extreme genetic heterogeneity underlying ASDs, finding two families in which the same gene is implicated is rare, and typically involves different variants within the gene (Cukier et al. 2014). Finding two unrelated families in which the *same* rare variant or haplotype segregates with ASD is exceedingly rare. In a resequencing study of CHD8 (a strongly supported ASD risk gene) in 3,730 children with developmental delay or ASD, 17 rare potentially disruptive single nucleotide variants were identified, and none were carried by more than one affected individual (Bernier et al. 2014).

In 2015, we published a linkage analysis of ASD in 7 extended families, each with 3 or more affected individuals (Chapman et al. 2015). We performed linkage analyses using dense SNP panels and identified multiple regions of interest within each family. Whole exome sequence (WES) was available for select individuals in each family, resulting in a list of 28 candidate exome variants (Chapman et al. 2015, Table 4). At that time, no other ASD data sets were publicly available in which to attempt replication of our findings.

Here, by interrogation of the MSSNG database (Yuen et al. 2017), we provide replication of some of our earlier findings. The MSSNG database contains whole genome sequence (WGS) on thousands of individuals diagnosed with ASD and their immediate family members. Remarkably, in this database we found the transmission of one candidate exome variant identified in our earlier study (Chapman et al. 2015), to multiple ASD individuals in multiple families. Furthermore, we identified 2 other rare variants transmitted on the same background haplotype, which in turn implicate two genes. Satisfyingly, one of these genes is a transcription repressor, and thus is a good candidate to be the cause of ASD in the families identified.

METHODS

MSSNG database

The version of the MSSNG database (<https://research.mss.ng/db5>, October 20 2017) (Yuen et al. 2017) that we used contains WGS for 7,191 people (3,427 with ASD and 3,764 without), representing families ascertained in nine different studies. Platforms used for WGS include Complete Genomics, Illumina HiSeq 2000 and HiSeq X. Detailed phenotypic information is available for most individuals with ASD, although the depth of phenotyping and the exact measures used vary extensively across individuals in the database. Individuals

from 2,757 families are represented, most of whom are in trios (1 affected child, 2 parents) or quads (2 affected children, 2 parents). Details of ascertainment, phenotyping, and sequencing have been published separately (Yuen et al. 2017).

TDT analysis of candidate exome variants

Our original list of candidate variants (Chapman et al. 2015) was based on filtering for alternate allele frequency <5% in 1,000 Genomes Europeans (Altshuler et al. 2015), because the vast majority of our families were of European descent. Here we tightened our frequency requirement to <1% in the ExAC database (Lek et al. 2016), further restricting the number of variants to consider to 14. For each variant, we identified all carriers in MSSNG, then extracted available genotype and phenotype (affected/unaffected) information for them and all related individuals. We examined carrier parents who were members of trios (parents and 1 affected child) or nuclear families (parents and 2 or more affected children) and tabulated the number of transmissions (T) and non-transmissions (U) of the alternate allele from carrier parents to affected children. A standard TDT test (Spielman et al. 1993) then offers a test of the null hypothesis that the variant is neither linked to nor associated with ASD. Because many of the variants are exceedingly rare and ASD is so genetically heterogeneous, we expect to see few copies of the variants and therefore power is not large, even using the large MSSNG database. Therefore we did not require strictly significant p-values, but rather used them to prioritize variants for further consideration. When studying rare variants and complex disease, rigorously assessing significance is difficult, and correcting for multiple testing compromises power. Identification of three families segregating a rare variant is considered strong evidence of causality (Brunham and Hayden 2013). Only 1 of the 14 variants under consideration showed evidence of linkage and/or association with ASD, and subsequent analyses focused on this candidate exome variant alone.

Identification of rare cosegregating whole genome variants

Because the candidate exome variant was ascertained using WES in local family AU071 (Chapman et al. 2015), it is possible that the cosegregation with ASD observed in MSSNG and AU071 actually reflects a risk variant outside of the exome on the underlying haplotype. If this is the case, the exome variant is merely tagging a risk haplotype rather than itself being a risk variant. The availability of WGS in MSSNG allows us to pursue this question. Most affected individuals in MSSNG are part of a trio design, where both the individual and their parents have WGS. This allows unambiguous determination of the haplotype transmitted from the parent carrying the candidate exome variant to the affected child at all locations where only one parent is heterozygous (likely when the alternate allele is either very common or very rare). We restricted consideration to cosegregating variants that were rare (alternate allele frequency in 1,000 Genomes Europeans < 1%).

TDT testing of rare cosegregating whole genome variants

Variants that cosegregate with the candidate exome variant are not necessarily associated with ASD, because they may also appear independently. To test for linkage and association with ASD, we extracted carriers of each rare cosegregating variant and their family members from MSSNG and performed the TDT test, as previously described for the exome variants.

Annotation of candidate risk variants

The candidate exome variant and rare cosegregating variants not ruled out by the TDT test were annotated using Variant Effect Predictor and Ensembl(Zerbino et al. 2018) for functional annotation and gnomAD(Lek et al. 2016) for allele frequencies. GERP++(Davydov et al. 2010) and phastCons scores(Felsenstein and Churchill 1996) were also obtained, using CADD v1.3(Kircher et al. 2014). Tissue-specific protein expression levels were obtained from Protein Atlas(Uhlen et al. 2015).

Sanger sequencing of candidate risk variants

Given the expense of WGS, we used Sanger sequencing to provide genotypes in individuals in local family AU071 for the small number of candidate risk variants. Genomic DNA was PCR amplified using customized primers designed by the online tool Primer3Plus(Untergasser et al. 2007) with HotStartTaq Plus DNA Polymerase (QIAGEN) with annealing temperatures of 58C. GenScript Biotech (Piscataway, NJ) performed Sanger sequencing. Biosamples are not available for individuals in the MSSNG database, so there was no opportunity to validate variants in them. The study involving AU071 was approved by the University of Washington Institutional Review Board, and informed consent was obtained from all participants or their parents.

RESULTS

Support for rs200195897 in MSSNG

Table 1 shows the results of analyses using the MSSNG data and the 14 previously identified exome variants. All but 1 of the 14 are observed in the MSSNG database, but only one stands out as having a transmission pattern consistent with being a risk variant for ASD. Variant rs200195897(1:877523 C→G, GRCh37) is a missense variant in SAMD11 and is present in 10 copies in MSSNG. Table 2 describes the MSSNG families in which this variant occurs. Figure 1a shows our family AU071, and Figure 1b depicts the MSSNG families. In the MSSNG families, 7 of the 10 copies are involved in transmissions to 4 affected children. This is scored as 4 transmissions and 0 non-transmissions (exact TDT p-value=0.0625). Three copies of the variant appear in singleton affected individuals whose parents are not sequenced, and one of these also has an affected half-sibling who does not carry the variant.

Identification of rare variants cosegregating with rs200195897 in MSSNG

Because rs200195897 was ascertained using WES in AU071, it is possible that cosegregation of this variant with ASD reflects a risk variant outside of the exome, on the underlying haplotype. If this is the case, rs200196897 merely tags a risk haplotype, rather than itself being a risk variant. The availability of WGS in MSSNG yields an *exhaustive* list of rare variants cosegregating with rs200195897, on all 4 transmitted haplotypes (see Table 3). One copy of each of the 5 cosegregating variants is also found in each of the 3 singletons, which is consistent with the singletons also carrying the same haplotype.

TDT testing of cosegregating variants

The rare variants that cosegregate with rs200195897 are not necessarily associated with ASD because they may also appear independently of rs200195897. To test for linkage and association with ASD, we performed the TDT test on genotype data from all carriers of each of the 5 cosegregating variants and their family members in MSSNG. Table 3 shows that 2 of the 5 are present in many additional individuals and are not associated with ASD. The remaining 3 cosegregating variants are present in exactly the same individuals as rs200195897, and no additional individuals, and therefore have the same borderline significant p-value. Thus we are left with a total of 4 SNPs that are candidate risk variants for ASD in these families, marked with a superscript in Table 3.

Sanger sequencing of candidate risk variants in AU071

By Sanger sequencing in AU071 we confirmed i) the genotypes of the exome variant rs200195897 as indicated in Figure 1a, and ii) the presence of rs56647288 and rs185038034 alternate alleles in all carriers of rs200195897, and no one else. This confirms that the background haplotype containing the rs200195897 alternate allele is shared amongst AU071 and all MSSNG families. rs544092417 was not observed in AU071. The most likely explanation is that this represents sequencing artifacts in the original WES data and the WGS data reported in MSSNG, as the variant is next to a LINE element (Kent et al. 2002; Rosenbloom et al. 2013), and this may interfere with accurate alignment.

Annotation of candidate risk variants

Table 4 shows available functional annotation and allele frequencies for the 4 candidate risk variants. All 4 variants are in chromosomal band 1p36.33, at the very p-terminus. They are very rare, with a maximum frequency in Non-Finnish Europeans of approximately 0.00067. They are also present in a similar frequency in Finnish Europeans, and in a lower frequency in African American samples. rs200195897 is the variant whose predicted consequences are the most severe. It is a missense variant in SAMD11, and also part of a promoter binding site in regulatory region ENSR00000000109, which may influence transcription of NOC2L, a gene about 2kb distant. rs200195897 is in a conserved element, as identified by GERP++ (GerpRS=789.8 $p=7 \times 10^{-71}$). Additionally, phastCons scores indicate high probabilities of conservation within vertebrates (0.998) and within the subset of mammals (0.955).

Phenotypes of affected carriers

Examination of phenotype data suggests some common characteristics of carriers. Tables 5 and 6 summarize phenotypes relating to language delay (age at first word and phrase, Autism Diagnostic Interview-Revised (ADI-R, Rutter 2003)), IQ (Wechsler Intelligence Scales (Wechsler 1981, 1989, 1992), Peabody Picture Vocabulary Test (Dunn and Dunn 2007), Stanford-Binet Intelligence Scales 5 (Roid 2005)), autism diagnosis (ADI-R (Rutter 2003), Autism Diagnostic Observation Schedule (ADOS, Lord et al. 2000; Lord et al. 1999; Lord 2003)), and social functioning (Social Responsiveness Scale-2 (SRS2, Constantino 2012)) in all available affected carriers of the candidate risk variants. AU071-05 does not meet ADOS criteria for an ASD diagnosis, but shows characteristics of BAP. The data in Table 5 indicate that there is no evidence of language delay, and (with one exception)

individuals are of average or higher intelligence. Table 6 shows consistent impairments in social communication and interactions for most carriers. Online Resource 1 contains tables that summarize other available phenotypes, including adaptive measures (Vineland Adaptive Behavior Scales II (Sparrow 2005), Adaptive Behavior Assessment System II (Harrison and Oakland 2008)) and measures of social functioning (SRS2- Constantino 2012). Because the individuals were recruited in different studies, there are few phenotypic measures that overlap. Also, no phenotype data were available for AU4148–301 beyond the information that he is affected.

DISCUSSION

We replicated the transmission of one candidate exome variant identified in our earlier study (Chapman et al. 2015), to multiple ASD individuals in multiple families in MSSNG. Non-replication of transmission of the other 13 variants has several possible explanations. First, 4 of the variants were simply too rare for meaningful testing in the MSSNG sample. Second, the original variants were identified by linkage analysis, using suggestive evidence of linkage because of limited family sizes, followed by filtering of WES variants based on both frequency and function. The lower threshold used for linkage evidence would be expected to introduce some false positives. A third possibility is that some of these exome variants are on a background haplotype where a non-exome variant increases ASD risk – such variants are invisible to exome sequencing.

In addition to observation of rs200195897 in MSSNG, we identified 2 other rare variants transmitted on the same background haplotype. The two genes near the variants are sterile- α -motif domain containing gene 11 (*SAMD11*) and a NOC2-like nucleolar-associated transcriptional repressor (*NOC2L*). *SAMD11* is evolutionarily conserved from zebrafish to humans. It is expressed primarily in developing retinal photoreceptors and in the adult mouse pineal gland (Jin et al. 2013), and shows very low levels of RNA transcript in human brain (Uhlen et al. 2015). It is therefore not an obvious candidate gene for ASD, although there are other genes containing a SAM domain that have been implicated in ASD (e.g. *SHANK1*–3 (Shi et al. 2017)). *NOC2L* is ubiquitously expressed throughout embryonic development and adulthood (Hublitz et al. 2005). It is expressed at high levels in Purkinje cells of the cerebellum, and medium levels in neuronal cells in both the hippocampus and the caudate nucleus (Uhlen et al. 2015). Several genes that are involved in transcriptional regulation play a major role in ASD (De Rubeis et al. 2014) including *CHD8*, which is the biggest contributor of de-novo variation linked to ASD (Cotney et al. 2015), and *MECP2*, which is responsible for Rett Syndrome (Amir et al. 1999). *NOC2L* is therefore an excellent candidate gene, and transmission of a promoter region variant in multiple families is suggestive of a role in ASD.

The evidence presented here implicates this haplotype on 1p36.33 in susceptibility to ASD. In light of the extreme genetic heterogeneity underlying ASD, replication of transmission in 4 distinct families provides a high degree of confidence in the validity of this finding. Functional studies are the natural next step to understand the importance and consequences of these variants. One approach would be to compare *SAMD11* and *NOC2L* expression in neurons derived from carrier-specific induced pluripotent stem cells (Lin et al. 2011) (iPSCs).

to that in iPSCs derived from non-carrier siblings. Another approach would be to generate knock-in mice carrying these variants (Singh et al. 2015) and evaluate the phenotypic consequences (Silverman et al. 2010). Functional experiments such as these are beyond the scope of the current project, but are needed to determine which of the variants convey risk. In the meantime, *SAMD11*, *NOC2L*, and nearby regulatory regions should be considered as genomic regions where variants can increase the risk of ASD.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

This study was supported by National Institute of Mental Health grant R01MH094293. We greatly appreciate the time and effort given by the family members who participated in both our local study and the MSSNG database. The authors also wish to acknowledge the resources of MSSNG (www.mss.ng), Autism Speaks and The Centre for Applied Genomics at The Hospital for Sick Children, Toronto, Canada. We thank the donors who supported this program for their generosity.

REFERENCES

- Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, Chakravarti A, et al. (2015) A global reference for human genetic variation. *Nature* 526: 68 [PubMed: 26432245]
- American Psychiatric Association (2013) Diagnostic and Statistical Manual (Fifth Edition). American Psychiatric Association, Washington, DC
- Amir RE, Van den Veyver IB, Wan M, Tran CQ, Francke U, et al. (1999) Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nature Genetics* 23: 185–188 [PubMed: 10508514]
- Baio J, Wiggins L, Christensen DL, Maenner MJ, Daniels J, et al. (2018) Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years - Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2014 (vol 67, pg 1, 2018). *Mmwr-Morbidity and Mortality Weekly Report* 67: 564–564 [PubMed: 29771874]
- Ben-David E, Shifman S (2013) Combined analysis of exome sequencing points toward a major role for transcription regulation during brain development in autism. *Molecular Psychiatry* 18: 1054–1056 [PubMed: 23147383]
- Berg JM, Geschwind DH (2012) Autism genetics: searching for specificity and convergence. *Genome Biol* 13: 247 [PubMed: 22849751]
- Bernier R, Gerdtz J, Munson J, Dawson G, Estes A (2012) Evidence for broader autism phenotype characteristics in parents from multiple-incidence autism families. *Autism Res* 5: 13–20 [PubMed: 21905246]
- Bernier R, Golzio C, Xiong B, Stessman HA, Coe BP, et al. (2014) Disruptive CHD8 Mutations Define a Subtype of Autism Early in Development. *Cell* 158: 263–276 [PubMed: 24998929]
- Bolton PF, Carcani-Rathwell I, Hutton J, Goode S, Howlin P, et al. (2011) Epilepsy in autism: features and correlates. *Br J Psychiatry* 198: 289–94 [PubMed: 21972278]
- Brunham LR, Hayden MR (2013) Hunting human disease genes: lessons from the past, challenges for the future. *Human Genetics* 132: 603–617 [PubMed: 23504071]
- Buescher AV, Cidav Z, Knapp M, Mandell DS (2014) Costs of autism spectrum disorders in the United Kingdom and the United States. *JAMA Pediatr* 168: 721–8 [PubMed: 24911948]
- Chapman NH, Nato AQ, Jr., Bernier R, Ankenman K, Sohi H, et al. (2015) Whole exome sequencing in extended families with autism spectrum disorder implicates four candidate genes. *Hum Genet* 134: 1055–68 [PubMed: 26204995]
- Constantino JN (2012) Social Responsiveness Scale 2nd Edition. Western Psychological Services, Los Angeles

- Cotney J, Muhle RA, Sanders SJ, Liu L, Willsey AJ, et al. (2015) The autism-associated chromatin modifier CHD8 regulates other autism risk genes during human neurodevelopment. *Nature Communications* 6
- Cukier HN, Dueker ND, Slifer SH, Lee JM, Whitehead PL, et al. (2014) Exome sequencing of extended families with autism reveals genes shared across neurodevelopmental and neuropsychiatric disorders. *Mol Autism* 5: 1 [PubMed: 24410847]
- Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, et al. (2010) Identifying a High Fraction of the Human Genome to be under Selective Constraint Using GERP plus. *Plos Computational Biology* 6
- De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, et al. (2014) Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* 515: 209–215 [PubMed: 25363760]
- Dunn LM, Dunn DM (2007) Peabody Picture Vocabulary Test, Fourth Edition
- Dykens EM, Lense M (2011) Intellectual disabilities and autism spectrum disorder: a cautionary note In: Amaral D, Dawson G, Geschwind DH (eds) *Autism spectrum disorders* Oxford University Press, New York, pp 263–270
- Felsenstein J, Churchill GA (1996) A hidden Markov Model approach to variation among sites in rate of evolution. *Molecular Biology and Evolution* 13: 93–104 [PubMed: 8583911]
- Gilman SR, Iossifov I, Levy D, Ronemus M, Wigler M, et al. (2011) Rare De Novo Variants Associated with Autism Implicate a Large Functional Network of Genes Involved in Formation and Function of Synapses. *Neuron* 70: 898–907 [PubMed: 21658583]
- Gratten J, Visscher PM, Mowry BJ, Wray NR (2013) Interpreting the role of de novo protein-coding mutations in neuropsychiatric disease. *Nature Genetics* 45: 234–238 [PubMed: 23438595]
- Harrison P, Oakland T (2008) *Adaptive Behavior Assessment System® Second Edition*. MHS Assessments
- Hublitz P, Kunowska N, Mayer UP, Muller JM, Heyne K, et al. (2005) NIR is a novel INHAT repressor that modulates the transcriptional activity of p53. *Genes & Development* 19: 2912–2924 [PubMed: 16322561]
- Jin GR, Long CD, Liu WW, Tang Y, Zhu YJ, et al. (2013) Identification and characterization of novel alternative splice variants of human SAMD11. *Gene* 530: 215–221 [PubMed: 23978614]
- Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, et al. (2002) The human genome browser at UCSC. *Genome Research* 12: 996–1006 [PubMed: 12045153]
- Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, et al. (2014) A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics* 46: 310 [PubMed: 24487276]
- Kozma P, Hughbanks-Wheaton DK, Locke KG, Fish GE, Gire AI, et al. (2005) Phenotypic characterization of a large family with RP10 autosomal-dominant retinitis pigmentosa: An Asp226Asn mutation in the IMPDH1 gene. *American Journal of Ophthalmology* 140: 858–867 [PubMed: 16214101]
- Krumm N, O’Roak BJ, Shendure J, Eichler EE (2014) A de novo convergence of autism genetics and molecular neuroscience. *Trends in Neurosciences* 37: 95–105 [PubMed: 24387789]
- Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, et al. (2016) Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536: 285 [PubMed: 27535533]
- Lin MY, Pedrosa E, Shah A, Hrabovsky A, Maqbool S, et al. (2011) RNA-Seq of Human Neurons Derived from iPS Cells Reveals Candidate Long Non-Coding RNAs Involved in Neurogenesis and Neuropsychiatric Disorders. *Plos One* 6: e23356 [PubMed: 21915259]
- Lord C, Risi S, Lambrecht L, Cook EH, Leventhal BL, et al. (2000) The Autism Diagnostic Observation Schedule-Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders* 30: 205–223 [PubMed: 11055457]
- Lord C, Rutter M, DiLavore P, Risi S (1999) *Autism Diagnostic Observation Schedule Western* Psychological Services Los Angeles, CA
- Lord C, Rutter M, DiLavore PC, & Risi S (2003) *Autism Diagnostic Observation Schedule Manual* Western Psychological Services, Los Angeles, CA

- Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, et al. (2012) Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* 485: 242–5 [PubMed: 22495311]
- O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, et al. (2012) Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* 485: 246–50 [PubMed: 22495309]
- Ozonoff S, Young GS, Carter A, Messinger D, Yirmiya N, et al. (2011) Recurrence risk for autism spectrum disorders: a Baby Siblings Research Consortium study. *Pediatrics* 128: e488–95 [PubMed: 21844053]
- Piven J, Palmer P, Jacobi D, Childress D, Arndt S (1997) Broader autism phenotype: evidence from a family history study of multiple-incidence autism families. *Am J Psychiatry* 154: 185–90 [PubMed: 9016266]
- Power RA, Kyaga S, Uher R, MacCabe JH, Langstrom N, et al. (2013) Fecundity of Patients With Schizophrenia, Autism, Bipolar Disorder, Depression, Anorexia Nervosa, or Substance Abuse vs Their Unaffected Siblings. *Jama Psychiatry* 70: 22–30 [PubMed: 23147713]
- Roid GH (2005) Stanford-Binet Intelligence Scales, Fifth Edition. Western Psychological Services
- Rosenbloom KR, Sloan CA, Malladi VS, Dreszer TR, Learned K, et al. (2013) ENCODE data in the UCSC Genome Browser: year 5 update. *Nucleic Acids Res* 41: D56–63 [PubMed: 23193274]
- Rutter M, LeCouteur A, Lord C (2003) Autism Diagnostic Interview Revised: WPS Edition Manual Western Psychological Services, Los Angeles
- Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, et al. (2011) Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron* 70: 863–85 [PubMed: 21658581]
- Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, et al. (2012) De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* 485: 237–41 [PubMed: 22495306]
- Shi R, Redman P, Ghose D, Liu Y, Ren XB, et al. (2017) Shank Proteins Differentially Regulate Synaptic Transmission. *Eneuro* 4
- Silverman JL, Yang M, Lord C, Crawley JN (2010) Behavioural phenotyping assays for mouse models of autism. *Nature Reviews Neuroscience* 11: 490–502 [PubMed: 20559336]
- Singh P, Schimenti JC, Bolcun-Filas E (2015) A Mouse Geneticist's Practical Guide to CRISPR Applications. *Genetics* 199: 1–15 [PubMed: 25271304]
- Sparrow S, Cichetti D, Balla D (2005) Vineland Adaptive Behavior Scales: 2nd Edition. Pearson, Bloomington, IN
- Spielman RS, McGinnis RE, Ewens WJ (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *American Journal of Human Genetics* 52: 506–516 [PubMed: 8447318]
- Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, et al. (2015) Tissue-based map of the human proteome. *Science* 347
- Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, et al. (2007) Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Research* 35: W71–W74 [PubMed: 17485472]
- Villanueva A, Willer JR, Bryois J, Dermitzakis ET, Katsanis N, et al. (2014) Whole Exome Sequencing of a Dominant Retinitis Pigmentosa Family Identifies a Novel Deletion in PRPF31. *Investigative Ophthalmology & Visual Science* 55: 2121–2129 [PubMed: 24595387]
- Wechsler D (1981) Wechsler adult intelligence scale - revised (WAIS-R).
- Wechsler D (1989) WPPSI-R manual: Wechsler preschool and primary scale of intelligence, revised
- Wechsler D (1992) Wechsler intelligence scale for children - third edition (WISC-III)
- Yu XP, Shi W, Cheng LL, Wang YF, Chen D, et al. (2016) Identification of a rhodopsin gene mutation in a large family with autosomal dominant retinitis pigmentosa. *Scientific Reports* 6
- Yuen RKC, Merico D, Bookman M, Howe JL, Thiruvahindrapuram B, et al. (2017) Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nature Neuroscience* 20: 602 [PubMed: 28263302]
- Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, et al. (2018) Ensembl 2018. *Nucleic Acids Research* 46: D754–D761 [PubMed: 29155950]

Fig. 1.

Pedigrees carrying rs200195897. Slashes indicate individuals without genotype or sequence information. ♦ indicates a heterozygous carrier of rs200195897. (a) Family AU071. Some sexes have been obscured to protect privacy. Carrier status was confirmed with Sanger sequencing. (b) Families found in the MSSNG database, carriers identified by WGS.

Table 1:

MSSNG database counts and TDT p-values for 14 exome variants previously identified.

rsname	Position(GRCh37)	ref	alt	ExAC freq.	MSSNG count	T	U	TDT p-value	Gene
rs200195897	1:877,523	C	G	0.0013	10	4	0	0.046	SAMD11
rs756479452	1:1,844,777	T	G	-	10	2	4	0.414	CALM6
rs189748047	1:216,062,273	C	T	0.00029	2	0	2	0.157	USH2A
rs201269761	4:520,973	G	A	0.00079	2	0	1	0.317	PIGG
rs146433325	4:1,656,767	G	A	0.00063	0				FAM53A
rs35671301	5:102,490,411	T	G	0.0078	125	39	41	0.823	PIIP5K2
rs140072759	6:90,024,862	A	G	0.00014	9	4	2	0.414	GABRR2
rs145467198	7:121,944,239	T	C	0.0063	174	22	29	0.327	FEZF1
rs75076193	7:142,836,646	C	T	0.0075	90	26	33	0.362	PIP
rs117136255	7:149,522,165	G	A	0.0049	90	30	28	0.793	SSPO
rs144177217	7:150,389,751	C	A	0.00026	4	1	3	0.317	GIMAP2
rs117403721	10:102,089,635	C	A	0.0048	100	36	25	0.159	PKD2L1
rs71317487	21:28,210,457	T	C	0.0025	55	23	14	0.139	ADAMTS1
rs61735785	21:40,190,408	C	A	0.0072	142	44	41	0.745	ETS2

Table 2:

Carriers of rs200195897 in MSSNG, by family.

Family	# copies	Comments on carriers
AU3839	3	Father transmits to one affected daughter ^a , one affected son ^a
AU3857	2	Father transmits to affected son ^a
1-1040	2	Mother transmits to affected son ^a
1-0891	1	Affected male, parents not sequenced
AU4148	1	Affected male, parents not sequenced
AU4094	1	Affected female, parents not sequenced. Affected half brother does not carry.

^a marks haplotype transmissions that can be directly examined because both parents and child are sequenced.

Table 3:

Rare variants cosegregating with rs200195897 in MSSNG.

rsname	Position (GRCh37)	ref	alt	IKGP-EUR	Present in all 3 singletons	T	U	TDT p-value	# carrier parents	# singleton affected carriers
rs564913738	1:789,465	C	T	0.007	Yes	23	30	0.34	39	17
rs566472488 ^a	1:845,956	C	T	0.002	Yes	4	0	0.046	3	3
rs544092417 ^a	1:848,029	T	C	0.002	Yes	4	0	0.046	3	3
rs200195897 ^a	1:877,523	C	G	NA	Yes	4	0	0.046	3	3
rs142806645	1:882,109	C	G	0.004	Yes	16	13	0.58	20	3
rs185038034 ^a	1:890,252	A	G	0.002	Yes	4	0	0.046	3	3

^a indicates candidate risk variants based on TDT

IKGP-EUR: frequency of the alternate allele in 1,000 Genomes Europeans.

Table 4:

Frequency and function annotation for candidate risk variants.

rsname	Present in in AU071	Annotation ^a	Gene	European Non-Finnish	European Finnish	African/ African-American
rs566472488	yes	Intronic, promoter flanking	lincRNA	6.7×10 ⁻⁴	5.7×10 ⁻⁴	3.45×10 ⁻⁴
rs544092417	no	Upstream/downstream	lincRNA	6.7×10 ⁻⁴	5.7×10 ⁻⁴	2.29×10 ⁻⁴
rs200195897	yes	Missense, promoter	SAMD11	6.0×10 ⁻⁴	5.7×10 ⁻⁴	2.30×10 ⁻⁴
rs185038034	yes	Intronic	NOC2L	6.1×10 ⁻⁴	5.8×10 ⁻⁴	2.31×10 ⁻⁴

^aVariant Effect Predictor and Ensembl, both GRCh38.p10

^bin gnomAD r2.0.2 WGS. None of the variants were observed in Ashkenazi, Latino or East Asian populations.

Phenotypes in carriers: ages at first word and phrase, IQ scores. No language development or IQ phenotypes are available for AU4148-301, 1-1040-003, or 1-0891-003.

Table 5:

	Sex	Age(yrs)		Age (mo)		IQ assessment							
		At testing	First word	First phrase	Wechsler			Peabody		Stanford-Binet			
					FSIQ	PIQ	VIQ	Receptive Vocabulary	FSIQ	VIQ	NVIQ		
AU071-03	M	11.8	20	33		79	67	97					
AU071-04	M	10.5	18	30		89	78	103					
AU071-05	M	7.6				114	122	103					
AU071-06 ^a	F	12.4				97	89	106					
AU3839-302	F	12.5	11	18					114	99	98	100	
AU3839-303	M	9.4	12	15					99	92	93	92	
AU3857-301	M	15.0	13	30					111				
AU4094-302	F	5.0	16	18					105	100	102	97	

^a Ages at first word and phrase are not available for AU071-06, but mother reports no language delay.

Table 6:

Autism phenotypes in carriers: ADOS (Module 3), ADI, and SRS-2, where available. No ADOS, ADI, or SRS-2 available for AU4148-301, 1-1040-003, or 1-0891-003.

	ADOS					ADI			SRS-2	
	Comm. – A Total	Social – B Total	Behav. – D Total	Social Total	Comm. Total	Behavior Total	Regression Yes/No	Seizures Yes/No	Total T-score	
AU071-03	3	7	1	24	20	6	No	-	-	
AU071-04	5	7	2	10	7	10	Yes	-	-	
AU071-05	2	1	0	-	-	-	-	-	-	
AU071-06	2	5	0	-	-	-	-	-	66	
AU3839-302	3	7	3	19	18	6	No	No	60,70 ^a	
AU3839-303	3	11	1	19	18	6	No	No	85,71 ^a	
AU3857-301	-	-	-	20	14	7	No	No	93	
AU4094-302	2	11	4	16	11	6	No	No	-	

^a indicates individuals where there are 2 entries for the SRS-2, about 3 months apart.