

High-risk Autism Spectrum Disorder Utah pedigrees: a novel Shared Genomic Segments analysis.

Todd M Darlington, PhD<sup>1</sup>  
Deborah Bilder, MD<sup>1</sup>  
Jubel Morgan, MS<sup>1</sup>  
Leslie Jerominski, MS<sup>1</sup>  
Venkatesh Rajamanickam, MS<sup>2</sup>  
Rob Sargent, BS<sup>2</sup>  
Nicola J Camp, PhD<sup>2</sup>  
Hilary H Coon, PhD<sup>1</sup>

<sup>1</sup>Department of Psychiatry, University of Utah, Salt Lake City, Utah 84108

<sup>2</sup>Department of Internal Medicine, Division of Cancer Control and Population Sciences, University of Utah, Salt Lake City, Utah 84108

Corresponding author:

Todd Darlington, PhD  
Huntsman Cancer Institute  
2000 Circle of Hope, Room 4121  
Salt Lake City, UT 84112  
(o) (801) 213-5763  
(f) (801) 581-5604  
todd.darlington@utah.edu

## Abstract

Progress in gene discovery for Autism Spectrum Disorder (ASD) has been rapid over the past decade, with major successes in validation of risk of predominantly rare, penetrant, *de novo* and inherited mutations in over 100 genes (de Rubies et al., 2015; Sanders et al., 2015). However, the majority of individuals with ASD diagnoses do not carry a rare, penetrant genetic risk factor. In fact, recent estimates suggest that most of the genetic liability of ASD is due to as yet undiscovered common, less penetrant inherited variation (Gaugler et al., 2014) which is much more difficult to detect. The study of extended, high-risk families adds significant information in our search for these common inherited risk factors. Here, we present results of a new, powerful pedigree analysis method (Shared Genomic Segments--SGS) on three large families from the Utah Autism Research Program. The method improves upon previous methods by allowing for within-family heterogeneity, and identifying exact region boundaries and subsets of cases who share for targeted follow-up analyses. Our SGS analyses identified one genome-wide significant shared segment on chromosome 17 (q21.32,  $p=1.47 \times 10^{-8}$ ). Additional regions with suggestive evidence were identified on chromosomes 3, 4, 6, 8, 11, 13, 14, 15, and 18. Several of these segments showed evidence of sharing across families. Genes of interest in these regions include *ATP8A1*, *DOCK3*, *CACNA2D2*, *ITGB3*, *AMBRA1*, *FOLH1*, *DGKZ*, *MTHFS*, *ARNT2*, *BTN2A2*, *BTN3A1*, *BTN3A3*, *BTN2A1*, and *BTN1A1*. We are exploring multiple other lines of evidence to follow up these implicated regions and genes.

## Introduction

Autism Spectrum Disorder (ASD) affects an estimated 1 in 68 children (Centers for Disease Control and Prevention, 2016), and imposes an enormous psychological and economic burden on affected individuals, their families, and society [1]. Genetic risk factors are important in the etiology of ASD, a conclusion strongly supported by recent population-based twin studies [2,3], and an additional recent twin meta-analysis [4]. These results suggest not only strong genetic effects, but also complex

inheritance, with heterogeneity between individuals, in addition to the likely involvement of multiple interacting genes within individuals. Indeed, recent genetic discovery studies have identified numerous inherited and *de novo* variants, and place the estimates for numbers of causal variants anywhere from several hundred to over a thousand [5–7].

Given this complex landscape, studies of large extended pedigrees provide an opportunity to add value. Allelic and locus heterogeneity are reduced and the ability to detect familial variants is enhanced [8,9]. Familial sharing of segments of the genome have the potential to provide important prioritization for the analysis of sequence variation, and may also add essential information about transmission, penetrance, and specificity provided by family relatives. Importantly, extended pedigrees offer sufficient statistical power to detect familial variants, often in a single pedigree, through the efficient study of relatively few individuals, although care must be taken to determine variants present in the pedigree relatives contributing to familial sharing to focus subsequent follow-up work. In addition, evidence is emerging that some variants contribute to phenotypes related, but not always specific to autism, such as intellectual disability [10–12]. The analysis of family members will also allow investigation of potential associations with phenotypes related to ASD in clinically affected and unaffected relatives.

The Utah Autism Research Program has ascertained a resource of high-risk extended families using the Utah Population Database (UPDB), a rich resource of health data and genealogical information for over eight million individuals who include descendants of the nineteenth century Utah pioneers [13]. Previous studies have shown low rates of inbreeding within the UPDB [14,15], and the population is primarily of Northern European ancestry [16]. We used the genealogical data within the UPDB resources to identify extended pedigrees with significantly increased incidence of ASD. This analysis focuses on the identification of segments of the genome shared among affected individuals in three of these extended families, and shared by few of their unaffected relatives, followed by a presentation of possible causal exome sequence variants within these segments. While we have previously explored these extended families using linkage methods [17], here we have employed the

novel method Shared Genomic Segments (SGS) [18]. This method has the advantage of identifying which pedigree members share which genomic segments, allowing for more accurate subsequent follow-up identification of possible variation explaining the pedigree sharing in sequence data. An ability to focus on shared regions in specific individuals will be of particular importance as we map more common variants that may exist outside the exome.

## **Materials and Methods**

### *Subjects*

Subjects were selected from three of the largest high-risk extended families in a previously described research sample [17,19]. Extended families in our study were identified or confirmed using the UPDB. Many additional distant family relationships between individuals with ASD that were not known to the subjects or their immediate families were identified. Each of the three large families in this study has multiple generations connecting many related nuclear families (see Figure 1). We analyzed data from a total of 24 individuals with ASD (8 each from families 10001, 25002, and 8002). In family 10001, there is one monozygotic affected twin pair; one member of the pair (64049) was randomly selected for this analysis. Additionally, at the data QC phase of the project, one individual with ASD in family 25002 was found to have a chromosome anomaly, and was therefore omitted from further analyses (83681), bringing the number of affected cases in 25002 in the final analyses to 7. We also studied 23 unaffected siblings in these families (8 in 10001, 8 in 25002, and 7 in 8002). Table 1 shows the characteristics of these individuals, and they are pictured in Figure 1. This study has ongoing approval from the University of Utah Institutional Review Board (approval IRB\_00006042) and the Resource for Genetic and Epidemiologic (RGE) committee which governs the use of UPDB data. Written consent was obtained for all study subjects.

### *Phenotyping*

Family members were requested to complete questionnaires and in-person testing. ASD was diagnosed with the Autism Diagnostic Interview-Revised (ADI-R) [20] and the Autism Diagnostic Observation Schedule (ADOS) [21], for all but two affected participants. One of these was given the ADOS only; the other received a lifetime DSM-IV diagnosis of autism by a clinical expert. Participants with ASD had no other reported major medical conditions or evidence of brain injury. One affected individual in family 8002 and one in family 10001 were non-verbal; all other affected individuals were verbal.

All but two affected participants had data available for the Social Responsiveness Scale, a quantitative measure of ASD characteristics ranging from significant impairment to above average social abilities [22,23]. SRS data were also obtained for all unaffected siblings with the exception of three adult unaffected siblings of case 73930 in family 8002. The parents reported no behavioral concerns for these siblings.

IQ was assessed using measurements appropriate for age and verbal ability, as described in Coon et al (2010) [17]. All affected cases had IQ estimates except two cases in family 10001; these two subjects were not testable due to intellectual disability. For the two unaffected siblings missing IQ in this study, parents reported average or above average school performance. Additionally, other self-reported medical history data were collected on most participants.

### *Genotyping*

Subjects were genotyped on the Illumina OmniExpress v1.0 platform using protocols and standards as set by Illumina. Raw data were output as Illumina forward orientation. To accurately compare our data with the publicly available control set from the 1000 Genomes Project (1KGP) [24], our data was converted to 1KGP forward orientation. A reference panel of 152 subjects with data from both the Illumina OmniExpress v1.0 platform and sequencing in 1KGP were used to determine a translation between the two platforms. A total of 25,166 variants were discarded due to inconsistency between the platforms. Only bi-allelic, single nucleotide variants (SNVs) were used.

### *Quality control - Plink*

Quality control was performed using PLINK [25] on the whole set of pedigrees combined with 168 control subjects of European descent from the 1KGP. First, 49 variants were removed with less than 95% call rate. No subjects were removed for genotype rates less than 90%. Next, sex checks were performed, resulting in one subject removed for a chromosomal anomaly. Relatedness estimates confirmed relationship status between pedigree members. A total of 12,688 variants were discarded due to deviation from Hardy-Weinberg equilibrium. Finally, 8,402 variants were removed for likely genotype errors, identified by Mendelian misinheritance, and 8,100 uninformative, monomorphic markers were removed prior to analysis.

### *Shared Genomic Segments*

Plink format files were converted to SGS format using software available through <http://genepi.med.utah.edu/~alun/software/index.html>. Genetic distance was interpolated using the Rutgers map (v3) [26]. Variants from 168 northern European 1KGP controls were used to construct a graphical model of linkage disequilibrium, used for simulating founder genotypes.

A pedigree SGS analysis considers sampled, distantly related cases (without genotyping from connecting relatives) and poses the question as to whether the length of consecutively shared loci (identified as identical-by-state, or IBS) among affected cases is longer than expected by chance [18]. IBS is established by determining if alleles at sequential loci are the same (phase is ignored). IBS is not the same as identity-by-descent (IBD; the same inherited segment from a common ancestor) which is our true interest; however, if the length of SGS shared IBS is significantly longer than expected by chance (given the known pedigree relationships) then IBD is implied. Observed sharing statistics (length of sharing) for a given pedigree are compared to simulated (“null”) shared statistics for an identical pedigree structure to determine empirical probabilities. Simulations involve assigning founder haplotypes (based on the graphical model of linkage disequilibrium) followed by simulating Mendelian

inheritance (with recombination likelihoods derived from Rutgers map) to all pedigree members (“gene-dropping”). For the length of a simulated segment to surpass an observed segment, the simulated shared segment must span the entire observed segment. For each pedigree, all possible combinations of affected cases were assessed for sharing. The total number of simulations to determine empirical significance for each subset of cases and for each chromosome ranged from 300,000 to 200,000,000. After a minimum of 300,000 simulations, no further simulations were performed on any particular subset where the most significant segment was observed at least three times in simulated data.

Thresholds for significance were determined using a modification of the technique previously described by Lander and Kruglyak (1995) for assessing significance in linkage analysis [27]. The modification incorporates extreme value theory and the law of large deviations in order to account for multiple testing (See supplemental methods). Briefly, empirical p-values for each marker were optimized by selecting the lowest p-value at each marker across all subsets, and reduced such that any shared segment in the optimized set was represented only once. The resulting distribution of p-values was used for threshold determination as described in supplemental methods. Additionally, to identify common segments shared across multiple families, the optimized p-values from each family were combined across families using Fisher’s method in R (metap, v0.7). There were two comparisons. First, we compared all families by combining optimized p-values across all three families. Second, we compared pairs of families by combining optimized p-values across each pair, then re-optimized at each marker using the combined p-values from each of the three pair combinations (8002-10001, 8002-25002, 10001-25002). Since these two comparisons are not independent, the thresholds were adjusted accordingly.

Although SGS infers IBD in shared segments, it only directly measures IBS. Therefore, for any significant SGS regions, the actual IBD segments are likely flanked by segments of random IBS, i.e. the actual interesting segment may be smaller than the SGS region identified. To address this, we utilized the presence of multiple affected siblings in each of our families to identify the most likely IBD segment, as follows. Consider a hypothetical family consisting of three affected cases. Case1 and Case2 are

siblings, and Case3 is distantly related. Case2 and Case3 share a segment spanning markers A, B, C, D, and E. However, when considering Case1, Case2, and Case3, the shared segment only spans A, B, and C, so therefore we infer that markers D and E were likely random IBS between the two distant cousins, and the actual IBD region is A, B, and C.

To assess preferential segregation of identified shared segments to ASD cases, significant segments were re-analyzed including all possible combinations also of unaffected siblings of cases, who serve as within-family controls. We assume variants associated with ASD risk may have reduced penetrance and/or may affect sub-clinical aspects of ASD that can occur in unaffected siblings; therefore, some degree of segregation to these siblings is not unexpected. However, excessive segregation to unaffected siblings may indicate a false positive result. Although we prioritized findings where less than one half of within-family controls shared the same segment, all segments with evidence of IBD are reported.

#### *Online bioinformatics for segment follow-up*

Base pair boundaries for each shared segment were defined as: previous unshared marker base position + 1bp before the shared segment to the first unshared marker base position - 1bp after shared segment. Genes within shared segments were identified using the Table Browser tool [28] from the UCSC Genome Browser [29,30]. Known ASD risk genes were identified using ASD@Princeton [31], a new, validated, web-based tool developed using machine-learning techniques based on known ASD genes and a human brain-specific gene network.

#### *Plotting*

Plots and visualizations were generated using the UCSC Genome Browser, R v3.2.2 [32] and ggplot2 v1.0.1 [33], Progeny v9.5.0.2 pedigree drawing software, and the Genome Decoration Page tool (<http://www.ncbi.nlm.nih.gov/genome/tools/gdp/>).



## Results

### *Shared Genomic Segments*

Significance thresholds for families and shared segments meeting those thresholds in each family are shown in Table 2 and Figure 2. One segment on chromosome 17 from family 10001 reached genome-wide significant (chr17q21.32,  $p=1.47 \times 10^{-8}$ ), with only a single unaffected sibling (76756) also sharing. Several segments in each family exceeded genome-wide suggestive thresholds. In family 8002, three shared segments reached the suggestive threshold, two on chromosome 4 and one on chromosome 13. Segregation of segments ranged from 3-9 internal controls out of 12 total. In family 10001, three segments reached the suggestive threshold, one each on chromosomes 3, 6, 14. Segregation ranged from 3-7 out of 8 total internal controls. In family 25002, two segments, one each on chromosome 11 and 15, reached the suggestive threshold with 3-7 out of 12 internal controls also sharing. Shared segments for each family are shown in Figure 3, along with genes in each segment. Genes were only included figure if they resided in the shared segment of interest (Figure 3). Only segments with genes are shown. Specific subset details are given in Supplemental Material.

### *Overlapping segments from multiple families*

Significance thresholds for segments identified in multiple families and shared segments meeting those thresholds in each family are shown in Table 3 and Figure 2. Several segments met the suggestive threshold. When considering multiple families, an additional 13 regions are identified. These segments are shown in Figure 3 along with the genes in each segment. Again, genes were included in the figure if they resided in the shared segment of interest. Only segments with genes are shown.

## Discussion

This analysis has revealed several familial shared genomic segments that may harbor variants contributing to Autism Spectrum Disorder risk in specific cases in our study. From within-family analyses, these segments are on chromosomes 3, 4, 6, 11, 13, 14, 15, and 17. Several of these

segments also showed evidence of sharing between families. Additional between-family shared segments were identified on chromosomes 6, 8, 11, and 18.

There were several genes in these segments that have been previously implicated in autism spectrum disorders and multiple others that have been associated with autism-like phenotypes. Specific genes of interest include *ATP8A1* in family 8002; *DOCK3*, *CACNA2D2*, and *ITGB3* in family 10001; *AMBRA1*, *FOLH1*, *DGKZ*, *MTHFS*, and *ARNT2* in family 25002; and several butyrophilin subunits (*BTN2A2*, *BTN3A1*, *BTN3A3*, *BTN2A1*, and *BTN1A1*) in segments shared across multiple families.

In family 8002, seven affected members share the segment on chromosome 4p13 containing ATPase phospholipid transporting 8 A1 (*ATP8A1*). *ATP8A1* was shown to be elevated in post-mortem human juvenile hippocampal tissue from subjects with ASD, and the same study showed deficits in mouse sociability behavior due to induced *ATP8A1* overexpression [34]. However, *ATP8A1* knockout mice also showed differences in hippocampal function [34,35], suggesting that precise regulation of *ATP8A1* is necessary for normal hippocampal development and function. Three of twelve unaffected siblings also share this segment, so it predominantly is segregating to affected cases.

In family 10001, all eight affected family 10001 members share the segment on chromosome 3p21.31-p21.2 containing calcium voltage-gated channel auxiliary subunit alpha 2 delta 2 (*CACNA2D2*) and dedicator of cytokinesis 3 (*DOCK3*). Three of eight unaffected siblings also share this segment. Variants in *CACNA2D2* have been associated with childhood epilepsy [36,37] and opioid sensitivity [38]. Furthermore, ASD subjects have been identified with de novo functional mutations [39] or duplication events [40] in *CACNA2D2* (see supplemental material from both studies). Interestingly, *DOCK3* has also been associated with epilepsy [41] as well as attention deficit hyperactivity disorder [42]. According to ASD@Princeton, *DOCK3* is a known ASD gene, regulating post-synaptic density through stimulation of axonal outgrowth [43] and is a target of Fragile-X mental retardation protein [44]. An analysis of runs of homozygosity (ROH) in ASD cases identified several affected individuals with ROH in *DOCK3* [45]. Five affected members of family 8002 also share this segment.

Also in family 10001, seven affected, and one unaffected sibling, share the segment on chromosome 17q21.32 containing integrin subunit beta 3 (*ITGB3*), which has been previously associated with ASD [46–53], although not in a cohort of Irish ASD cases [54]. Considered by ASD@Princeton to be a known ASD gene, *ITGB3* plays a role in cell adhesion, and regulates serotonin levels through an interaction with the serotonin transporter [55]. Interestingly, the lone affected case that does not share this segment is the MZ twin. Complications arising during twin pregnancy and birth have been shown to increase the risk for ASD [56]; it is possible these non-genetic risks contributed significantly to this case.

In family 25002, all seven affected family 25002 members share the segment on chromosome 11p11.2 containing autophagy and beclin 1 regulator 1 (*AMBRA1*) and diacylglycerol kinase zeta (*DGKZ*), shared with six unaffected siblings, and the segment just downstream (11p11.2-p11.12) containing folate hydrolase 1 (*FOLH1*, discussed below), shared with seven unaffected siblings. *AMBRA1* functions in regulating neurogenesis, is prominently expressed in the striatum, hippocampus, and cortex [57,58], and *AMBRA1* variations have been associated with schizophrenia [59,60]. Mouse models of *Ambra1* (+/-) heterozygous mutation show increased autism-like behavior exclusively among females coincided with lower protein expression among female vs. male heterozygous mice [58]. Likewise, variations in *DGKZ* were also linked to schizophrenia [59], and *DGKZ* is a target of FMRP [44].

Also in family 25002, the chromosome 15q25.1 segment, shared by six affected family 25002 members, and three of twelve unaffected siblings, contains 5,10-methenyltetrahydrofolate synthetase (*MTHFS*, discussed below) and aryl hydrocarbon receptor nuclear translocator 2 (*ARNT2*), which encodes a transcription factor involved in neurogenesis and plays a critical role in neurodevelopment of the hypothalamo-pituitary axis [61–63]. *ARNT2* regulates transcription of oxygen-responsive genes by dimerizing with hypoxia-inducible factor 1 alpha and binding to promoter and enhancer segments [64,65]. Variation in *ARNT2* has been associated with ASD and schizophrenia [66].

Several other segments also showed evidence of overlap between all three families. There is a gene cluster of butyrophilin genes (including *BTN2A2* and *BTN3A3*) on chromosome 6p22 in another segment of interest. A component of human milk fat, butyrophilin glycoproteins are members of the immunoglobulin superfamily of transmembrane proteins [67] and are involved in lipid and sterol metabolism, and interact with HLA genes. They may be of interest given current hypotheses of a possible immune component to ASD susceptibility [68], particularly in regards to molecular mimicry as exposure to butyrophilin proteins can induce, as well as suppress, T-cell immune response to myelin oligodendrocyte glycoprotein (MOG) implicated in the pathogenesis of multiple sclerosis [69]. Of related interest, *BTN2A2* has been associated with schizophrenia [70] and altered DNA methylation and expression of *BTN3A3* has been found in post-mortem brain tissue of individuals with bipolar disorder and schizophrenia [71]. Increased expression of *BTN2A2* was also found in postmortem brain tissue of individuals with schizophrenia [72].

ASD in these large, high-risk families is likely due to a combination of familial genetic variants, plus the interaction of each individual's unique environment, that have a range of effects on phenotypic risk. For some cases it is possible that these risk factors also combine with unidentified functional *de novo* variation. This study highlighted segments of the genome that are shared between affected cases in three high-risk ASD families. It is promising that the most significant finding of the three individual families, the segment on 17q21.32 in family 10001, contained known ASD-risk gene *ITGB3*. Furthermore, both shared segments in family 25002 contained genes involved in folate metabolism, *FOLH1* and *MTHFS*. In the intestines, *FOLH1* is required for folate absorption while in the brain, *FOLH1* is expressed by astrocytes, and facilitates the production of glutamate [73]. Variations in *FOLH1* have been associated with schizophrenia [74] and depression [75]. Activation of *MTHFS* intracellularly increases folate turnover and reduces folate concentrations [76]. A comparison of ASD cases and healthy controls found differences in the cerebrospinal fluid concentration of several folate metabolites [77], and folates have been explored or proposed as interventions in several neuropsychiatric disorders, including autism [77–79]. 5-methyltetrahydrofolate (5MTHF) crosses the

blood-brain barrier and is the predominant form of folate in the central nervous system [80]. Although cerebrospinal 5MTHF concentrations have been shown to differ between ASD cases and healthy controls [77], another study did not find a correlation between 5MTHF and ASD symptoms in a cohort of young ASD cases sampled twice over a period of several years [81]. Central folate deficiency was identified among females with Rett Syndrome in a European study [82], though Neul et al (2005) failed to replicate this finding among a US sample of females with Rett Syndrome [83]. The invasive nature of cerebrospinal fluid collection and unpredictable nature of CSF 5MTHF level fluctuation, precludes routine CSF collection for 5MTHF levels during autism assessment [81].

Shared genomic segments analysis has the ability to narrow the search for meaningful risk variants that may be more common and have reduced penetrance, plus the capacity to tie variants to specific phenotypes in a cohort of affected cases. Furthermore, SGS is not limited to exome variation with a much smaller universe of potential variants. This study has demonstrated each of these points, and follow-up analyses will focus on identifying risk variants in these families using whole genome sequencing.

### **Competing Interests Statement**

All authors declare that there are no competing interests to disclose.

### **Funding**

This research was supported in part by R01 MH0994400, R01 CA163353, R01 CA134674, Simons Foundation Project 388196, the Margolis Foundation, and GCRC M01-RR025764 from the National Center for Research Resources. Partial support for all datasets within the Utah Population Database (UPDB) was provided by the University of Utah Huntsman Cancer Institute.

### **Author Contributions**

LJ maintained and prepared the biological samples. JM and DB provided expert clinical insight. NC, VR, RS, HC and TD developed, tested, and executed the SGS analysis. TD and HC wrote the manuscript, with input from all authors. All authors reviewed the manuscript.

### **Acknowledgments**

Additional computing resources were utilized through the Center for High Power Computing at the University of Utah ([www.chpc.utah.edu](http://www.chpc.utah.edu)). The analyses required extensive use of gnu parallel [84].

Table 1. Description of individuals with ASD and unaffected siblings used for Shared Genomic Segments (SGS) analyses in three extended families

	Extended High-Risk Families		
Phenotype	10001	25002	8002
N Generations	8	8	8
<b>Cases</b>			
N	8	7	8
Proportion male	7   8	6   7	7   8
Age in years at assessment mean, (SD)	12.60 (10.60)	16.51 (12.51)	13.76 (3.72)
ADI Social domain, mean (SD) <sup>1</sup>	19.67 (7.87)	15.26 (9.27)	21.13 (6.45)
ADI Communication, mean (SD)	16.00 (5.02)	12.25 (7.52)	13.75 (4.80)
ADI Repetitive/ Restricted, mean (SD)	6.67 (2.16)	5.88 (3.80)	5.63 (2.88)
ADI Onset scale, mean (SD)	4.33 (1.21)	4.14 (1.22)	3.13 (1.37)
ADOS, mean (SD)	14.57 (4.86)	12.88 (6.92)	15.25 (4.59)
SRS total, mean (SD) <sup>2</sup>	84.13 (35.42)	78.29 (46.68)	84.71 (38.10)
Full Scale IQ, mean (SD) <sup>3</sup>	86.00 (35.83)	89.88 (29.67)	102.88 (26.70)
<b>Unaffected siblings</b>			
N	8	12	12
Proportion male	6   8	7   12	5   12
Age in years at assessment (SD)	20.34 (9.41)	24.34 (7.47)	13.39 (8.43)
SRS total, mean (SD)	13.63 (10.73)	10.17 (10.19)	31.67 (14.92)
Full Scale IQ, mean (SD)	109.57 (11.21)	117.55 (8.30)	112.36 (17.97)

<sup>1</sup> One individual with ASD in family 8002 did not receive the ADI assessment; one individual in family 10001 was given a lifetime DSM-IV diagnosis of autism but did not receive either the ADOS or ADI assessments.

<sup>2</sup> One individual with ASD in family 8002 and one in family 25002 did not have complete SRS assessments. Three unaffected siblings in family 8002 did not have a complete SRS assessments; however, parents reported no behavioral concerns. All three of these cases were adult siblings of case 73930.

<sup>3</sup> IQ was measured with an assessment appropriate for developmental level. In family 10001, IQ assessment was attempted but not completed for persons 64352 and 64356 due to significant cognitive delays. One unaffected sibling in family 8002, one in family 10001, and one in family 25002 did not complete the IQ assessment, but were reported by parents as performing at an average or above average level in school.

Table 2. Shared genomic segments with at least genome-wide suggestive evidence in each family.

Family	Locus	Cytoband	p-value <sup>a</sup>	Cases <sup>b</sup>	Controls <sup>c</sup>
<b>8002</b>	chr4:42,646,678-43,197,961	4p13	6.64x10 <sup>-7</sup>	7	3/12
<b>8002</b>	chr4:157,340,211-157,601,192	4q32.1	1.85x10 <sup>-7</sup>	8	6/12
<b>8002</b>	chr13:23,586,367-23,641,247	13q12.12	6.53x10 <sup>-7</sup>	7	9/12
<b>10001</b>	chr3:50,421,082-51,927,597	3p21.31-p21.2	5.51x10 <sup>-8</sup>	8	3/8
<b>10001</b>	chr3:52,348,365-53,160,359	3p21.1	5.51x10 <sup>-8</sup>	8	4/8
<b>10001</b>	chr6:130,537,006-131,438,062	6q23.1-q23.2	4.91x10 <sup>-8</sup>	7	3/8
<b>10001</b>	chr14:59,505,554-59,585,931	14q23.1	1.94x10 <sup>-7</sup>	8	4/8
<b>10001</b>	chr17:45,300,219-46,574,490	17q21.32	1.47x10 <sup>-8*</sup>	7	1/8
<b>25002</b>	chr11:46,309,969-47,394,304	11p11.2	1.20x10 <sup>-6</sup>	7	6/12
<b>25002</b>	chr11:47,449,073-50,519,186	11p11.2-p11.12	1.20x10 <sup>-6</sup>	7	7/12
<b>25002</b>	chr15:79,604,783-80,724,488	15q25.1	1.47x10 <sup>-6</sup>	6	3/12

<sup>a</sup>Significant and suggestive thresholds for family 8002 (9.90x10<sup>-8</sup> and 1.05x10<sup>-6</sup>); 10001 (1.92x10<sup>-8</sup> and 2.94x10<sup>-7</sup>); 25002 (2.20x10<sup>-7</sup> and 2.44x10<sup>-6</sup>)

<sup>a</sup>Number of affected cases sharing this segment

<sup>b</sup>Number of internal controls that share this segment with affected cases, shown with total number of possible internal controls

\*Genome-wide significant



Table 3. Overlapping shared segments with at least genome-wide suggestive evidence across families.

Families	Locus	Cytoband	p-value <sup>a</sup>	#Cases;#Controls <sup>b</sup>		
				8002 8;12	10001 8;8	25002 7;12
<b>8002;10001</b>	chr3:50,421,082-51,927,597	3p21.31-p21.2	1.20x10 <sup>-10</sup>	5;6	8;3	
<b>8002;10001</b>	chr3:52,348,365-52,467,323	3p21.1	1.20x10 <sup>-10</sup>	5;6	8;4	
<b>8002;10001</b>	chr11:51,566,910-55,584,108	11p11.12-q11	2.19x10 <sup>-10</sup>	6;7	6;4	
<b>8002;10001</b>	chr14:59,505,554-59,521,231	14q23.1	3.70x10 <sup>-10</sup>	7;9	8;4	
<b>8002;25002</b>	chr11:47,969,153-50,519,186	11p11.2-p11.12	4.53x10 <sup>-10</sup>	6;8		7;7
<b>10001;25002</b>	chr6:130,894,603-131,438,062	6q23.1-q23.2	2.54x10 <sup>-10</sup>		7;3	3;3
<b>All</b>	chr6:26,378,682-26,584,667	6p22.2	2.59x10 <sup>-12</sup>	3;3	7;6	6;10
<b>All</b>	chr6:36,342,182-36,700,832	6p21.31-p21.2	4.57x10 <sup>-12</sup>	8;8	6;4	2;0
<b>All</b>	chr8:64,001,937-64,194,109	8q12.3	4.82x10 <sup>-12</sup>	7;0	5;1	5;5
<b>All</b>	chr17:45,377,578-45,794,705	17q21.32	7.05x10 <sup>-13</sup>	8;6	7;1	6;0
<b>All</b>	chr17:46,319,649-46,509,446	17q21.32	7.05x10 <sup>-13</sup>	8;7	7;1	6;0
<b>All</b>	chr17:46,509,448-46,574,490	17q21.32	7.05x10 <sup>-13</sup>	8;7	7;1	6;0
<b>All</b>	chr18:60,050,642-60,090,078	18q21.33	6.48x10 <sup>-12</sup>	6;4	6;2	4;2

<sup>a</sup>Significant and suggestive thresholds for pairs of families (2.30x10<sup>-11</sup> and 4.53x10<sup>-10</sup>); all three families (2.00x10<sup>-13</sup> and 7.34x10<sup>-12</sup>)

<sup>b</sup>Number of affected cases and unaffected siblings sharing the respective segment in each family. The total number of affected cases and unaffected siblings per family is in the header line.

Figure 1. Family pedigree drawings.

Gender has been disguised, and sibling order has not been maintained in order to protect family privacy. Shaded symbols represent individuals with ASD. All family members with a numeric ID have available DNA. Boxed cases and unaffected siblings were used for SGS analyses. OmniExpress genotype data from family 8002 members 89891 and 100163 were not available at time of analysis (a). Family 10001 members 64049 and 64052 are monozygotic twins; 64049 was selected randomly for analysis (b). Family 25002 member 83681 was found during quality control to have a chromosomal anomaly and therefore the containing branch was not used for analysis (c).

Figure 2. Shared genomic segments meeting significance thresholds.

Shared genomic segments for each family and overlapping segments across multiple families are shown for each autosomal chromosome. Chromosome number is shown below each chromosome.

### Figure 3. Shared genomic segments with genes.

Figure 3 shows segments of the genome shared between affected cases in each family that meet significance thresholds. Chromosome segments are denoted by red boxes on each chromosome ideogram. Shared segments from family 8002 (red), 10001 (blue), 25002 (green) are shown below each ideogram. Thick portions of each family bar represent estimated actual shared segment. Thin bars for each family, if present, represent entire segment originally identified by SGS. Actual shared segments were determined by including any missing siblings from the original subset. Genes are displayed below shared segments.

## References

1. Buescher AVS, Cidav Z, Knapp M, Mandell DS. Costs of autism spectrum disorders in the United Kingdom and the United States. *JAMA Pediatr.* 2014;168:721–8.
2. Nordenbæk C, Jørgensen M, Kyvik KO, Bilenberg N. A Danish population-based twin study on autism spectrum disorders. *Eur. Child Adolesc. Psychiatry.* 2014;23:35–43.
3. Colvert E, Tick B, McEwen F, Stewart C, Curran SR, Woodhouse E, et al. Heritability of Autism Spectrum Disorder in a UK Population-Based Twin Sample. *JAMA Psychiatry.* 2015;72:415–23.
4. Tick B, Bolton P, Happé F, Rutter M, Rijsdijk F. Heritability of autism spectrum disorders: a meta-analysis of twin studies. *J. Child Psychol. Psychiatry.* 2016;57:585–95.
5. Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature.* 2012;485:237–41.
6. Buxbaum JD, Daly MJ, Devlin B, Lehner T, Roeder K, State MW. The autism sequencing consortium: large-scale, high-throughput sequencing in autism spectrum disorders. *Neuron.* 2012;76:1052–6.
7. De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Cicek AE, et al. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature.* 2014;515:209–15.
8. Terwilliger JD, Haghghi F, Hiekkalinna TS, Göring HHH. A bias-ed assessment of the use of SNPs in human complex traits. *Curr. Opin. Genet. Dev.* 2002;12:726–34.
9. Blangero J, Williams JT, Almasy L. Novel family-based approaches to genetic risk in thrombosis. *J. Thromb. Haemost. JTH.* 2003;1:1391–7.
10. Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R, et al. Functional impact of global rare copy number variation in autism spectrum disorders. *Nature.* 2010;466:368–72.
11. Hamdan FF, Daoud H, Rochefort D, Piton A, Gauthier J, Langlois M, et al. De novo mutations in FOXP1 in cases with intellectual disability, autism, and language impairment. *Am. J. Hum. Genet.* 2010;87:671–8.
12. Hamdan FF, Daoud H, Piton A, Gauthier J, Dobrzeniecka S, Krebs M-O, et al. De novo SYNGAP1 mutations in nonsyndromic intellectual disability and autism. *Biol. Psychiatry.* 2011;69:898–901.
13. Utah Population Database (UPDB) [Internet]. Available from: <http://healthcare.utah.edu/huntsmancancerinstitute/research/updb/>
14. Jorde LB. Inbreeding in the Utah Mormons: an evaluation of estimates based on pedigrees, isonymy, and migration matrices. *Ann. Hum. Genet.* 1989;53:339–55.
15. Jorde LB. Consanguinity and prereproductive mortality in the Utah Mormon population. *Hum. Hered.* 2001;52:61–5.
16. Lao O, Lu TT, Nothnagel M, Junge O, Freitag-Wolf S, Caliebe A, et al. Correlation between genetic and geographic structure in Europe. *Curr. Biol. CB.* 2008;18:1241–8.
17. Coon H, Villalobos ME, Robison RJ, Camp NJ, Cannon DS, Allen-Brady K, et al. Genome-wide linkage using the Social Responsiveness Scale in Utah autism pedigrees. *Mol. Autism.* 2010;1:1.
18. Thomas A, Camp NJ, Farnham JM, Allen-Brady K, Cannon-Albright LA. Shared genomic segment analysis. Mapping disease predisposition genes in extended pedigrees using SNP genotype assays. *Ann. Hum. Genet.* 2008;72:279–87.

19. Cannon DS, Miller JS, Robison RJ, Villalobos ME, Wahmhoff NK, Allen-Brady K, et al. Genome-wide linkage analyses of two repetitive behavior phenotypes in Utah pedigrees with autism spectrum disorders. *Mol. Autism*. 2010;1:3.
20. Lord C, Rutter M, Le Couteur A. Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J. Autism Dev. Disord*. 1994;24:659–85.
21. Lord C, Risi S, Lambrecht L, Cook EH, Leventhal BL, DiLavore PC, et al. The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J. Autism Dev. Disord*. 2000;30:205–23.
22. Constantino JN. *The Social Responsiveness Scale*. Los Angeles, CA: Western Psychological Services; 2002.
23. Constantino JN, Davis SA, Todd RD, Schindler MK, Gross MM, Brophy SL, et al. Validation of a brief quantitative measure of autistic traits: comparison of the social responsiveness scale with the autism diagnostic interview-revised. *J. Autism Dev. Disord*. 2003;33:427–33.
24. McVean GA, Altshuler (Co-Chair) DM, Durbin (Co-Chair) RM, Abecasis GR, Bentley DR, Chakravarti A, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012;491:56–65.
25. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M a R, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet*. 2007;81:559–75.
26. Matisse TC, Chen F, Chen W, De La Vega FM, Hansen M, He C, et al. A second-generation combined linkage physical map of the human genome. *Genome Res*. 2007;17:1783–6.
27. Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat. Genet*. 1995;11:241–7.
28. Karolchik D. The UCSC Table Browser data retrieval tool. *Nucleic Acids Res*. 2004;32:493D–496.
29. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. *Genome Res*. 2002;12:996–1006.
30. Rosenbloom KR, Armstrong J, Barber GP, Casper J, Clawson H, Diekhans M, et al. The UCSC Genome Browser database: 2015 update. *Nucleic Acids Res*. 2015;43:D670–81.
31. Krishnan A, Zhang R, Yao V, Theesfeld CL, Wong AK, Tadych A, et al. Genome-wide characterization of genetic and functional dysregulation in autism spectrum disorder [Internet]. 2016 Jun. Report No.: biorxiv/057828v1. Available from: <http://biorxiv.org/lookup/doi/10.1101/057828>
32. R Core Team. *R: A Language and Environment for Statistical Computing* [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2015. Available from: <https://www.R-project.org>
33. Wickham H. *ggplot2: elegant graphics for data analysis* [Internet]. Springer-Verlag New York; 2009. Available from: <http://ggplot2.org>
34. Kerr DJ, Marsillo A, Guariglia SR, Budylin T, Sadek R, Menkes S, et al. Aberrant hippocampal Atp8a1 levels are associated with altered synaptic strength, electrical activity, and autistic-like behavior. *Biochim. Biophys. Acta BBA - Mol. Basis Dis*. 2016;1862:1755–65.
35. Levano K, Punia V, Raghunath M, Debata PR, Curcio GM, Mogha A, et al. Atp8a1 deficiency is associated with phosphatidylserine externalization in hippocampus and delayed hippocampus-dependent learning: Neuronal plasma membrane flippase and Atp8a1. *J. Neurochem*. 2012;120:302–13.

36. Chioza BA, Aicardi J, Aschauer H, Brouwer O, Callenbach P, Covanis A, et al. Genome wide high density SNP-based linkage analysis of childhood absence epilepsy identifies a susceptibility locus on chromosome 3p23-p14. *Epilepsy Res.* 2009;87:247–55.
37. Pippucci T, Parmeggiani A, Palombo F, Maresca A, Angius A, Crisponi L, et al. A Novel Null Homozygous Mutation Confirms CACNA2D2 as a Gene Mutated in Epileptic Encephalopathy. Brusgaard K, editor. *PLoS ONE.* 2013;8:e82154.
38. Rhodin A, Grönbladh A, Ginya H, Nilsson KW, Rosenblad A, Zhou Q, et al. Combined analysis of circulating  $\beta$ -endorphin with gene polymorphisms in OPRM1, CACNA2D2 and ABCB1 reveals correlation with pain, opioid sensitivity and opioid-related side effects. *Mol. Brain.* 2013;6:1.
39. O’Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature.* 2012;485:246–50.
40. Gregory SG, Connelly JJ, Towers AJ, Johnson J, Biscocho D, Markunas CA, et al. Genomic and epigenetic evidence for oxytocin receptor deficiency in autism. *BMC Med.* [Internet]. 2009 [cited 2016 Nov 28];7. Available from: <http://bmcmedicine.biomedcentral.com/articles/10.1186/1741-7015-7-62>
41. Li J, Mi X, Chen L, Jiang G, Wang N, Zhang Y, et al. Dock3 Participate in Epileptogenesis Through rac1 Pathway in Animal Models. *Mol. Neurobiol.* 2016;53:2715–25.
42. De Silva MG, Elliott K, Dahl HH, Fitzpatrick E, Wilcox S, Delatycki M, et al. Disruption of a novel member of a sodium/hydrogen exchanger family and DOCK3 is associated with an attention deficit hyperactivity disorder-like phenotype. *J. Med. Genet.* 2003;40:733–740.
43. Namekata K, Harada C, Guo X, Kimura A, Kittaka D, Watanabe H, et al. Dock3 Stimulates Axonal Outgrowth via GSK-3-Mediated Microtubule Assembly. *J. Neurosci.* 2012;32:264–74.
44. Darnell JC, Van Driesche SJ, Zhang C, Hung KYS, Mele A, Fraser CE, et al. FMRP Stalls Ribosomal Translocation on mRNAs Linked to Synaptic Function and Autism. *Cell.* 2011;146:247–61.
45. Casey JP, Magalhaes TR, Conroy JM, Regan R, Shah N, Anney R, et al. A novel approach of homozygous haplotype sharing identifies candidate genes in autism spectrum disorder. *Hum. Genet.* 2012;131:565–79.
46. Weiss LA, Ober C, Cook EH. ITGB3 shows genetic and expression interaction with SLC6A4. *Hum. Genet.* 2006;120:93–100.
47. Weiss LA, Kosova G, Delahanty RJ, Jiang L, Cook EH, Ober C, et al. Variation in ITGB3 is associated with whole-blood serotonin level and autism susceptibility. *Eur. J. Hum. Genet. EJHG.* 2006;14:923–31.
48. Coutinho AM, Sousa I, Martins M, Correia C, Morgadinho T, Bento C, et al. Evidence for epistasis between SLC6A4 and ITGB3 in autism etiology and in the determination of platelet serotonin levels. *Hum. Genet.* 2007;121:243–56.
49. Ma DQ, Rabionet R, Konidari I, Jaworski J, Cukier HN, Wright HH, et al. Association and gene-gene interaction of SLC6A4 and ITGB3 in autism. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet. Off. Publ. Int. Soc. Psychiatr. Genet.* 2010;153B:477–83.
50. Napolioni V, Lombardi F, Sacco R, Curatolo P, Manzi B, Alessandrelli R, et al. Family-based association study of ITGB3 in autism spectrum disorder and its endophenotypes. *Eur. J. Hum. Genet. EJHG.* 2011;19:353–9.
51. Singh AS, Chandra R, Guhathakurta S, Sinha S, Chatterjee A, Ahmed S, et al. Genetic association and gene-gene interaction analyses suggest likely involvement of ITGB3 and TPH2 with autism spectrum disorder (ASD) in the Indian population. *Prog. Neuropsychopharmacol. Biol. Psychiatry.* 2013;45:131–43.

52. Schuch JB, Muller D, Endres RG, Bosa CA, Longo D, Schuler-Faccini L, et al. The role of  $\beta 3$  integrin gene variants in Autism Spectrum Disorders — Diagnosis and symptomatology. *Gene*. 2014;553:24–30.
53. Baig DN, Yanagawa T, Tabuchi K. Distortion of the normal function of synaptic cell adhesion molecules by genetic variants as a risk for autism spectrum disorders. *Brain Res. Bull.* [Internet]. 2016 [cited 2016 Dec 2]; Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0361923016303446>
54. Cochrane LE, Tansey KE, Gill M, Gallagher L, Anney RJL. Lack of association between markers in the ITGA3, ITGAV, ITGA6 and ITGB3 and autism in an Irish sample. *Autism Res*. 2010;3:342–4.
55. Mazaloukas M, Jessen T, Varney S, Sutcliffe JS, Veenstra-VanderWeele J, Cook EH, et al. Integrin  $\beta 3$  haploinsufficiency modulates serotonin transport and antidepressant-sensitive behavior in mice. *Neuropsychopharmacology* [Internet]. 2015 [cited 2017 Feb 2];40. Available from: <http://www.nature.com/npp/journal/v40/n8/abs/npp201551a.html>
56. Bilder D, Pinborough-Zimmerman J, Miller J, McMahon W. Prenatal, Perinatal, and Neonatal Factors Associated With Autism Spectrum Disorders. *PEDIATRICS*. 2009;123:1293–300.
57. Vázquez P, Arroba AI, Cecconi F, de la Rosa EJ, Boya P, De Pablo F. Atg5 and Ambra1 differentially modulate neurogenesis in neural stem cells. *Autophagy*. 2012;8:187–99.
58. Dere E, Dahm L, Lu D, Hammerschmidt K, Ju A, Tantra M, et al. Heterozygous Ambra1 Deficiency in Mice: A Genetic Trait with Autism-Like Behavior Restricted to the Female Gender. *Front. Behav. Neurosci.* [Internet]. 2014 [cited 2016 Nov 28];8. Available from: <http://journal.frontiersin.org/article/10.3389/fnbeh.2014.00181/abstract>
59. Rietschel M, Mattheisen M, Degenhardt F, Genetic Risk and Outcome in Psychosis (GROUP Investigators), Mühleisen TW, Kirsch P, et al. Association between genetic variation in a region on chromosome 11 and schizophrenia in large samples from Europe. *Mol. Psychiatry*. 2012;17:906–17.
60. Heinrich A, Nees F, Lourdasamy A, Tzschoppe J, Meier S, Vollstädt-Klein S, et al. From gene to brain to behavior: schizophrenia-associated variation in *AMBRA1* alters impulsivity-related traits. *Eur. J. Neurosci*. 2013;n/a-n/a.
61. Jain S, Maltepe E, Lu MM, Simon C, Bradfield CA. Expression of ARNT, ARNT2, HIF1 alpha, HIF2 alpha and Ah receptor mRNAs in the developing mouse. *Mech. Dev*. 1998;73:117–23.
62. Hosoya T, Oda Y, Takahashi S, Morita M, Kawachi S, Ema M, et al. Defective development of secretory neurones in the hypothalamus of Arnt2-knockout mice. *Genes Cells*. 2001;6:361–374.
63. Valen E, Pascarella G, Chalk A, Maeda N, Kojima M, Kawazu C, et al. Genome-wide detection and analysis of hippocampus core promoters using DeepCAGE. *Genome Res*. 2008;19:255–65.
64. Semenza GL. Hypoxia-Inducible Factors in Physiology and Medicine. *Cell*. 2012;148:399–408.
65. Bersten DC, Bruning JB, Peet DJ, Whitelaw ML. Human Variants in the Neuronal Basic Helix-Loop-Helix/Per-Arnt-Sim (bHLH/PAS) Transcription Factor Complex NPAS4/ARNT2 Disrupt Function. Kato M, editor. *PLoS ONE*. 2014;9:e85768.
66. Di Napoli A, Warriar V, Baron-Cohen S, Chakrabarti B. Genetic variant rs17225178 in the ARNT2 gene is associated with Asperger Syndrome. *Mol. Autism*. 2015;6:9.
67. Afrache H, Gouret P, Ainouche S, Pontarotti P, Olive D. The butyrophilin (BTN) gene family: from milk fat to the regulation of the immune response. *Immunogenetics*. 2012;64:781–94.
68. Gesundheit B, Rosenzweig JP, Naor D, Lerer B, Zachor DA, Procházka V, et al. Immunological and autoimmune considerations of autism spectrum disorders. *J. Autoimmun*. 2013;44:1–7.



69. Stefferl A, Schubart A, Storch M, Amini A, Mather I, Lassmann H, et al. Butyrophilin, a milk protein, modulates the encephalitogenic T cell response to myelin oligodendrocyte glycoprotein in experimental autoimmune encephalomyelitis. *J. Immunol. Baltim. Md 1950.* 2000;165:2859–65.
70. Glessner JT, Hakonarson H. Common variants in polygenic schizophrenia. *Genome Biol.* 2009;10:236.
71. Chen C, Zhang C, Cheng L, Reilly JL, Bishop JR, Sweeney JA, et al. Correlation between DNA methylation and gene expression in the brains of patients with bipolar disorder and schizophrenia. *Bipolar Disord.* 2014;16:790–9.
72. Sinkus ML, Adams CE, Logel J, Freedman R, Leonard S. Expression of immune genes on chromosome 6p21.3–22.1 in schizophrenia. *Brain. Behav. Immun.* 2013;32:51–62.
73. Šácha P, Zámečník J, Bařinka C, Hloučová K, Vícha A, Mlčochová P, et al. Expression of glutamate carboxypeptidase II in human brain. *Neuroscience.* 2007;144:1361–72.
74. Roffman JL, Brohawn DG, Nitenson AZ, Macklin EA, Smoller JW, Goff DC. Genetic Variation Throughout the Folate Metabolic Pathway Influences Negative Symptom Severity in Schizophrenia. *Schizophr. Bull.* 2013;39:330–8.
75. Ye X, Lai C-Q, Crott JW, Troen AM, Ordovas JM, Tucker KL. The folate hydrolase 1561C>T polymorphism is associated with depressive symptoms in Puerto Rican adults. *Psychosom. Med.* 2011;73:385–392.
76. Anguera MC, Suh JR, Ghandour H, Nasrallah IM, Selhub J, Stover PJ. Methenyltetrahydrofolate Synthetase Regulates Folate Turnover and Accumulation. *J. Biol. Chem.* 2003;278:29856–62.
77. Ramaekers VT, Sequeira JM, Quadros EV. The basis for folinic acid treatment in neuro-psychiatric disorders. *Biochimie.* 2016;126:79–90.
78. Roffman JL, Lamberti JS, Achtyes E, Macklin EA, Galendez GC, Raeke LH, et al. Randomized Multicenter Investigation of Folate Plus Vitamin B<sub>12</sub> Supplementation in Schizophrenia. *JAMA Psychiatry.* 2013;70:481.
79. Araújo JR, Martel F, Borges N, Araújo JM, Keating E. Foliates and aging: Role in mild cognitive impairment, dementia and depression. *Ageing Res. Rev.* 2015;22:9–19.
80. Hyland K, Shoffner J, Heales SJ. Cerebral folate deficiency. *J. Inherit. Metab. Dis.* 2010;33:563–70.
81. Shoffner J, Trommer B, Thurm A, Farmer C, Langley WA, Soskey L, et al. CSF concentrations of 5-methyltetrahydrofolate in a cohort of young children with autism. *Neurology.* 2016;86:2258–63.
82. Ramaekers VT, Hansen SI, Holm J, Opladen T, Senderek J, Häusler M, et al. Reduced folate transport to the CNS in female Rett patients. *Neurology.* 2003;61:506–515.
83. Neul JL, Maricich SM, Islam M, Barrish J, Smith E, Bottiglieri T, et al. Spinal fluid 5-methyltetrahydrofolate levels are normal in Rett syndrome. *Neurology.* 2005;64:2151–2152.
84. Tange O. GNU Parallel: The Command-Line Power Tool. *Login USENIX Mag.* [Internet]. 2011 [cited 2017 Jan 9];36. Available from: <https://pdfs.semanticscholar.org/7fcc/116f8760161ca5620858873901b27c9686a4.pdf>

Figure 1. Family pedigree drawings.

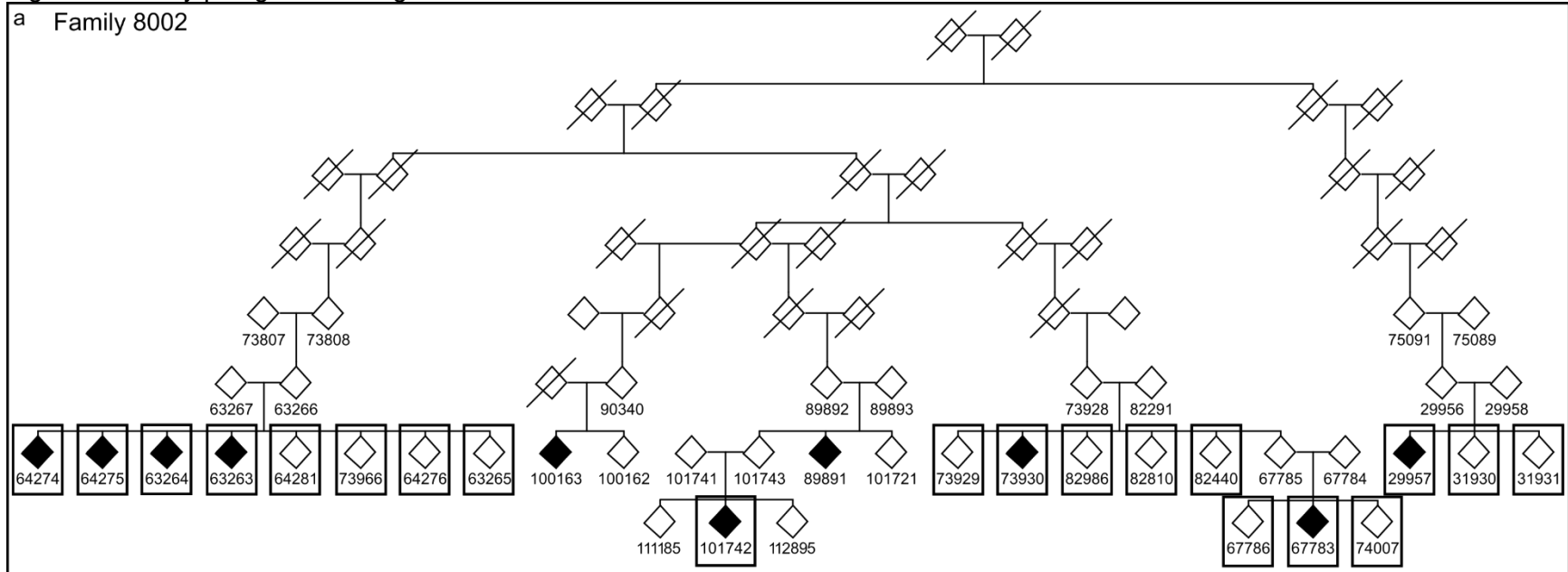


Figure 1 continued.

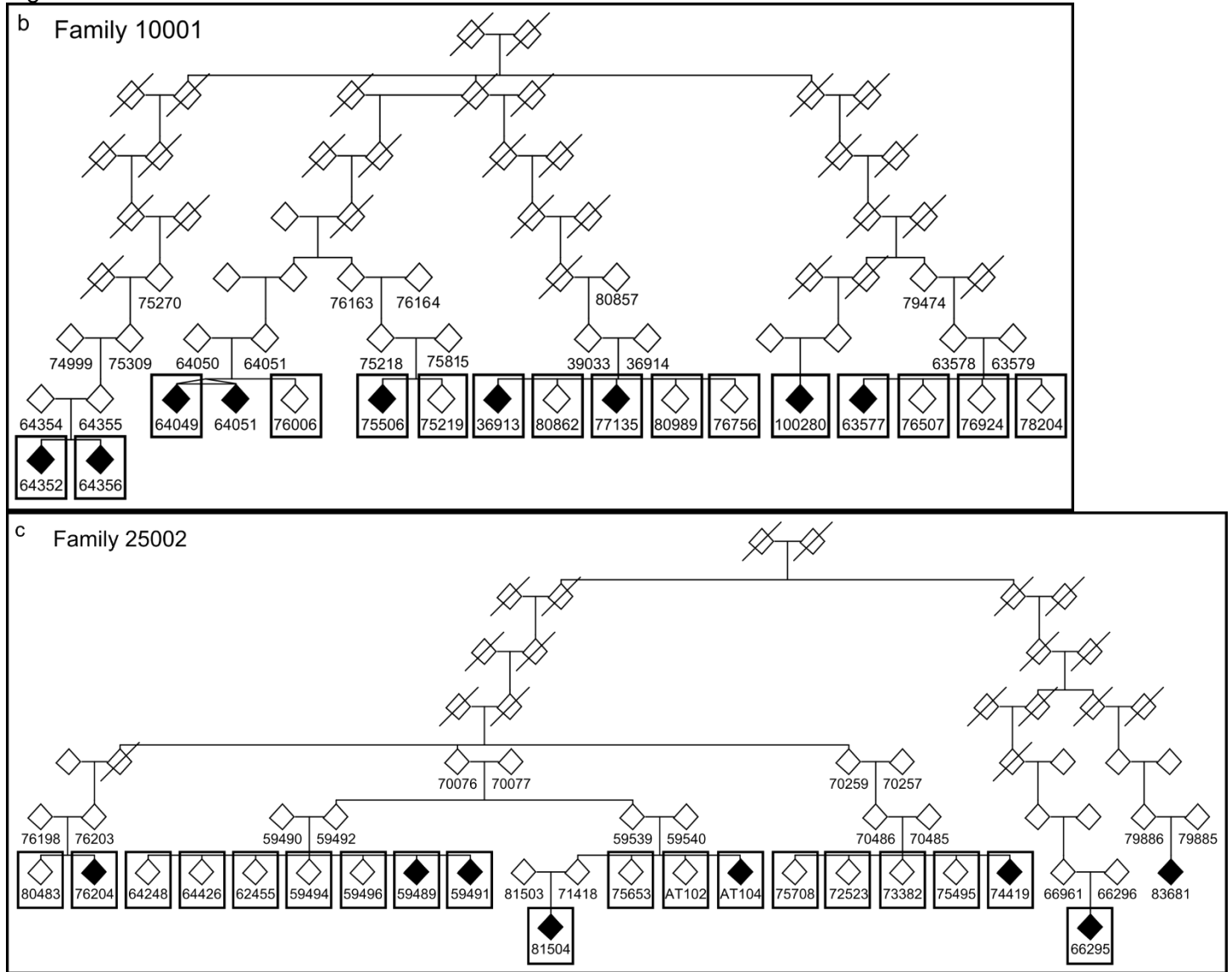


Figure 2. Shared genomic segments meeting significance thresholds.

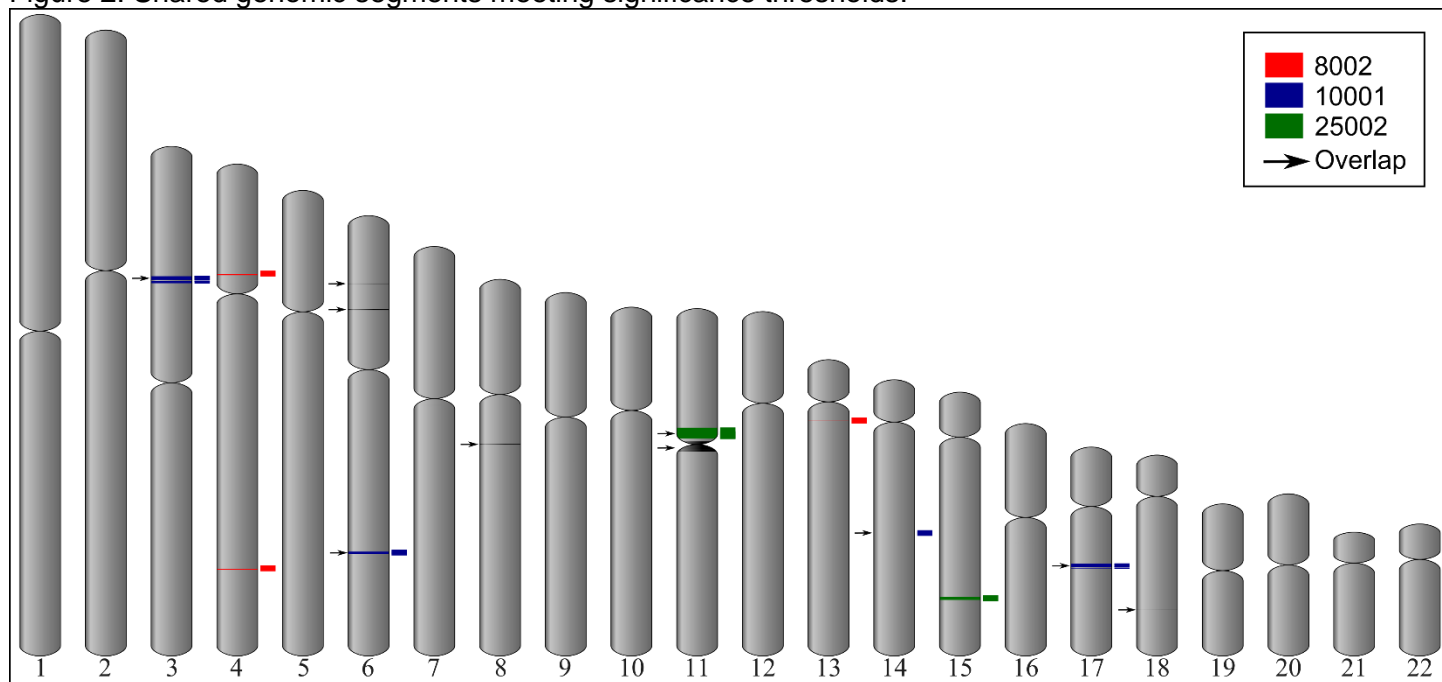


Figure 3. Shared genomic segments with genes.

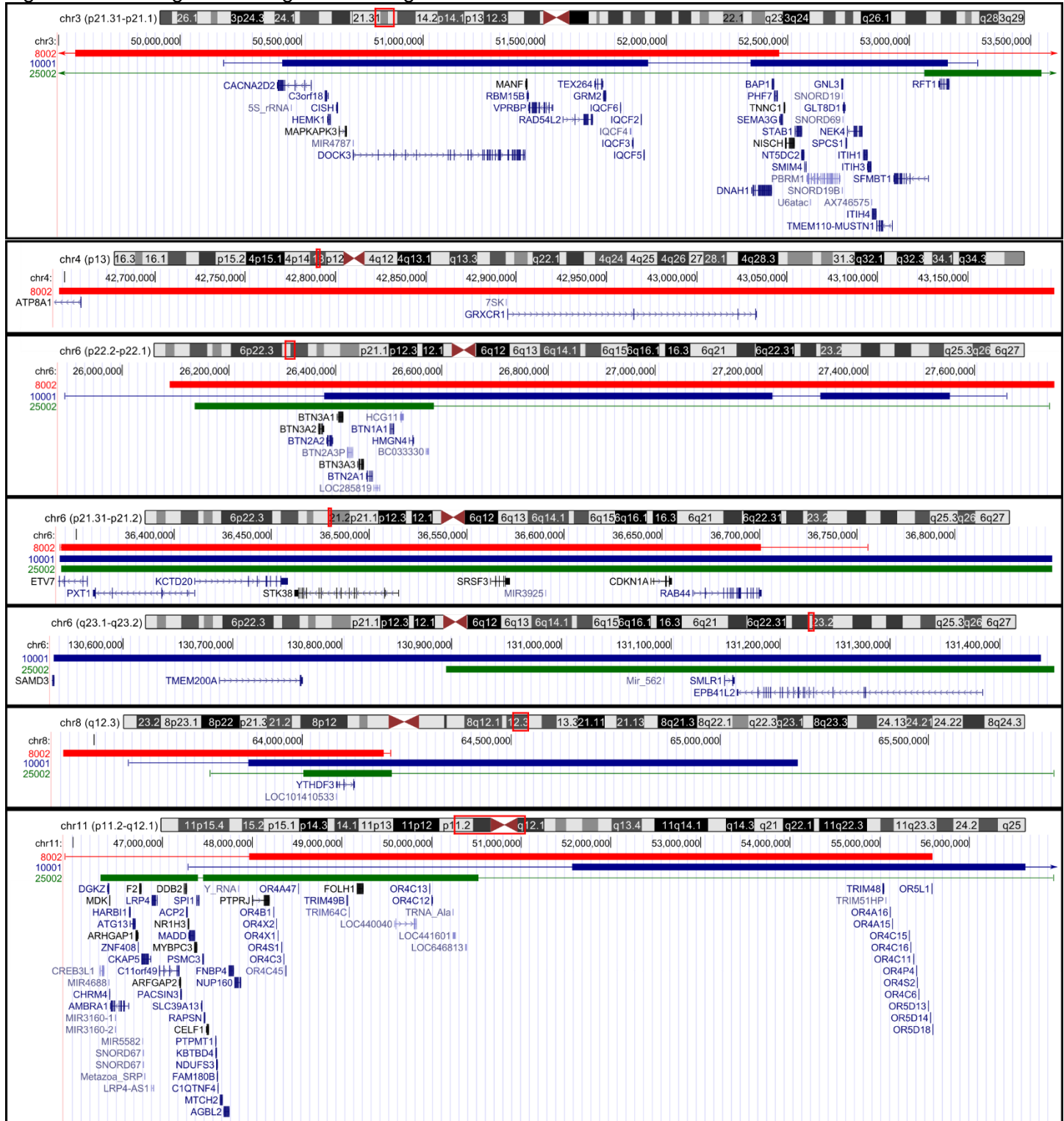


Figure 3 continued.

