

**Genetic etiologies
of Autism Spectrum Disorder**

Niklas Krumm

A dissertation submitted in partial
fulfillment of the requirements for the degree of
Doctor of Philosophy

University of Washington, 2014

Reading Committee:

Evan E. Eichler, Chair

Deborah A. Nickerson

Jay Shendure

Program Authorized to Offer Degree
Genome Sciences, University of Washington

ABSTRACT

Autism spectrum disorder (ASD) is a common, heritable neurodevelopmental disorder. In this thesis, I examine how different genetic etiologies, mutation types and specific genes contribute to the risk of ASD and how these factors can be used to expand our understanding of the neurobiological underpinnings of ASD. I develop a new bioinformatics method (CoNIFER: copy number inference from exome reads) to identify copy number variants (CNVs) using exome sequencing data, enabling much more sensitive identification of a previously under-ascertained class of small CNVs (<100 kbp in size). I estimate the precision of the algorithm using 366 exomes and show that this method can be used to reliably predict both *de novo* and inherited rare CNVs and can predict absolute copy number for loci with fewer than eight copies.

Next, I searched for disruptive, genic rare CNVs among 411 families with sporadic ASD from the Simons Simplex Collection (SSC) and identified additional small genic rare CNVs compared to high-density single nucleotide polymorphism (SNP) microarrays (~2X higher yield). I found that affected probands inherit more CNVs than their siblings ($p=0.004$; $OR=1.19$), and these CNVs affect more genes, are enriched for brain-expressed genes, and are transmitted preferentially from the mother. Finally, I found that the excess burden of inherited CNVs among probands is driven primarily by proband-sibling pairs with discordant social behavior phenotypes. I then created a combined set of both inherited and *de novo* single nucleotide variants (SNVs) and CNVs across 2,377 SSC ASD families, including 1,786 families with both an affected and unaffected child. I compared the burden of inherited and *de novo* mutations between affected and unaffected siblings and found that private inherited truncating SNV mutations in conserved genes are significantly enriched in probands ($OR=1.14$, $p<0.0002$)— an effect that became more pronounced with increasing gene conservation. I quantified ASD risk for *de novo* and inherited CNVs and SNVs by using a conditional logistic regression model. Independent from *de novo* mutations, private truncating SNVs and rare inherited CNVs contribute an increase in risk of 1.11 ($p=0.0002$) and 1.23 ($p=0.01$), respectively. These results confirm a statistically independent role for inherited mutations in ASD risk and identify additional candidate genes (e.g., *RIMS1*, *CUL7* and *CSMD1*) where inherited and *de novo* burden converge.

TABLE of CONTENTS

I. INTRODUCTION	6
1.1 SUMMARY	6
1.2 AUTISM SPECTRUM DISORDER	7
1.3 IDENTIFYING GENES AND PATHWAYS INVOLVED IN ASD AND ID	8
1.4 AN INCREASE IN <i>DE NOVO</i> LOSS-OF-FUNCTION MUTATIONS	11
1.5 CANDIDATE GENES YET FEW RECURRENT HITS	12
1.6 LARGE-SCALE RESEQUENCING OF CANDIDATE GENES	13
1.7 NOVEL CANDIDATES AND THEIR NEUROBIOLOGY	16
1.8 PROTEIN INTERACTION NETWORKS CONVERGE ON COMMON PATHWAYS	18
1.9 ASSAYING COPY NUMBER VARIATION	24
1.10 THESIS GOALS	25
II. COPY NUMBER VARIATION DETECTION AND GENOTYPING FROM EXOME SEQUENCE DATA	28
2.1 SUMMARY	28
2.2 INTRODUCTION	29
2.3 METHODS	29
2.4 RESULTS	32
2.5 DISCUSSION	41
III. TRANSMISSION DISEQUILIBRIUM OF SMALL CNVS IN SIMPLEX AUTISM	44
3.1 SUMMARY	44
3.2 INTRODUCTION	45
3.3 METHODS	46
3.4 RESULTS	48
3.5 DISCUSSION	57
IV. INHERITED SNV MUTATIONS IN AUTISM SPECTRUM DISORDER	65
4.1 SUMMARY	65
4.2 INTRODUCTION	66
4.3 METHODS	67
4.4 RESULTS	68
4.5 DISCUSSION	79
V. SUMMARY AND FUTURE DIRECTIONS	82
5.1 SUMMARY OF RESULTS	82
5.2 TOWARDS ASSAYING THE COMPLETE SET OF GENETIC VARIATION	82
5.3 UNDERSTANDING NORMAL VARIATION AND PATHOGENIC VARIANTS:	86
5.4 DEFINING SUBTYPES OF THE AUTISM SPECTRUM	88
5.5 DEFINING A GRADIENT OF SIMPLEX AND MULTIPLEX AUTISM	89
5.6 UNDERSTANDING COMPLEX GENETIC ETIOLOGIES AT A FAMILY LEVEL	90
5.7 FUTURE DIRECTIONS	92
REFERENCES	94
WEB AND SOFTWARE RESOURCES	106
APPENDICES	108

LIST of FIGURES

Chapter 1

Figure 1.1:	Estimating the number of ASD/ID risk genes	14
Figure 1.2:	Location of <i>de novo</i> truncating mutations in ASD and ID genes	16
Figure 1.3:	Genes disrupted by <i>de novo</i> mutations form a connected network	20
Figure 1.4:	<i>CHD8</i> and <i>CTNNB1</i> putative regulatory network for head size.....	22

Chapter 2

Figure 2.1:	Method overview and CNV discovery	34
Figure 2.2:	CNP locus genotyping of <i>RHD</i> and <i>C4A</i>	38
Figure 2.3:	Genotyping accuracy across 62 CNP loci.....	39

Chapter 3

Figure 3.1:	Discovery and validation of previously undiscovered CNVs using exomes.....	50
Figure 3.2:	Increased inherited CNV burden in ASD probands for large and small CNVs	52
Figure 3.3:	Inherited CNV burden correlates with SRS phenotype	54
Figure 3.4:	Genes in proband-only CNVs from SRS-discordant quads are more likely brain-expressed.....	56
Figure 3.5:	A combined model of inherited and <i>de novo</i> mutations	62

Chapter 4

Figure 4.1:	Network of genes with recurrent <i>de novo</i> hits	70
Figure 4.2:	Transmission disequilibrium of SNVs in ASD.....	72
Figure 4.3:	Transmitted mutations and their effect on phenotype.....	74
Figure 4.4:	Convergence of <i>de novo</i> and inherited mutations on <i>CSMD1</i>	77
Figure 4.5:	Combined risk model for SNVs and CNVs, inherited and <i>de novo</i>	78

Chapter 5

Figure 5.1:	High genomic GC nucleotide content (green histogram) hinders whole- exome sequencing	84
Figure 5.2:	Number of trios sequenced and expected rate for ASD-implicated genes	91

LIST of TABLES

Chapter 1

Table 1.1:	Six recent family-based exome studies of ASD and ID	10
Table 1.2:	Recurrent disruptive mutations in ID and ASD	15

Chapter 2

Table 2.1:	Cohorts analyzed	30
Table 2.2:	Precision of exome-based CNV calls in HapMap samples	36
Table 2.3:	Precision of exome-based CNV calls in autism trios	37

Chapter 3

Table 3.1:	Summary of transmitted CNVs in 411 ASD quads.....	49
Table 3.2:	Summary of IQ and SRS burden	55
Table 3.3:	Selected inherited CNVs	61

Chapter 4

Table 4.1:	Genes with new recurrent <i>de novo</i> mutations	69
Table 4.2:	Summary of mutations in IGF-related ASD network	71
Table 4.3:	Converging evidence for <i>RIMS1</i> , <i>CUL7</i> and <i>CSMD1</i> from <i>de novo</i> and inherited mutations	75
Table 4.4:	Summary of logistic regression model results	79

I. Introduction

1.1 Summary

Autism spectrum disorder (ASD) and intellectual disability (ID) are neurodevelopmental disorders with large genetic components, but identification of pathogenic genes has proceeded slowly because hundreds of loci are involved. This introduction describes how new exome sequencing technology has identified novel rare variants and has found that sporadic cases of ASD/ID are enriched for disruptive *de novo* mutations. Targeted large-scale resequencing studies have confirmed the significance of specific loci, including chromodomain helicase DNA binding protein 8 (*CHD8*), sodium channel, voltage-gated, type II, alpha subunit (*SCN2A*), dual specificity tyrosine-phosphorylation-regulated kinase 1A (*DYRK1A*), and beta-catenin (*CTNNB1*). I review recent studies and suggest that they have led to a convergence on three functional pathways: (i) chromatin remodeling; (ii) wnt signaling during development; and (iii) synaptic function.

The heritability and genetic etiology cannot be fully explained by *de novo* events, and I suggest that additional underlying genetic etiologies, mechanisms and effects also play a role in ASD. I describe three major topics of this thesis: (i) the development of new tools to more completely assay genetic variants using next-generation sequencing data (Chapter II), (ii) describing the role of inherited copy number variants (CNVs) in the risk of ASD (Chapter III), and (iii) exploring how multiple types of mutations, including inherited and *de novo* SNVs and CNVs, contribute to ASD (Chapter IV).

This introduction has been adapted from: Krumm, N., O'Roak, B. J., Shendure, J., & Eichler, E. E. (2014). *A de novo convergence of autism genetics and molecular neuroscience. Trends in Neurosciences*, 37(2), 95–105. doi:10.1016/j.tins.2013.11.005.

1.2 Autism spectrum disorder

Autism spectrum disorder (ASD) is a common (frequency 1/88–1/250 births) disorder that results in three distinct phenotypes which manifest early in development: 1) deficits in language ability, 2) impairment in reciprocal social interaction and communication, and 3) repetitive or stereotyped movements and interests. ASD encompasses a range of disorders, including Asperger's syndrome and Autistic Disorder itself. ASD and related disorders can commonly present with comorbidities of intellectual disability or developmental delay and epilepsy, but these are neither requirements for diagnosis nor observed in all cases.

The etiology of ASD has a strong genetic component, and the heritability of these disorders is widely estimated to be 80-90% based on studies comparing concordance in monozygotic and dizygotic twins (Lichtenstein et al. 2010; Bailey et al. 1995), although larger and more recent studies based on a broader set of familial relationships have shown significantly lower estimates of ~50% (Hallmayer et al. 2011). Additional evidence for a genetic etiology stems from increased proportions of ASD diagnoses in monogenic developmental disorders such as Fragile X and tuberous sclerosis, the identification of genes in families with Mendelian forms of autism, an increased rate of sibling recurrence (~8% for boys), and observation of a "Broad Autism Phenotype" in (unaffected) parents of children with ASD.

Understanding the genetic etiology of ASD has several important motivating factors. First, identification of specific genes (and mutations) underlying the pathogenesis of ASD will enable a new era of genetic diagnosis and classification of ASD. The observed ASD phenotypes are extremely heterogeneous, making diagnostic criteria based on behavior and phenotype necessarily broad as well. An understanding of the specific genes that cause ASD will enable a much more precise diagnosis, or refinement of an initial diagnosis.

Second, it is likely the genetic etiology of ASD encompasses two different broad mechanisms of genetic mutations: The first are so-called *de novo* mutations, which are

present in offspring but not their parents and arise either in the parent's germline or early in development; the second type of mutation is an inherited (or *transmitted*) mutation from parents to offspring. Both types of mutations are likely to contribute to ASD, and an understanding of their importance and pathogenic potential will provide a crucial “big picture” understanding of how genetics affect ASD risk.

Finally, an understanding of both the genes and the broader genetic mechanisms in ASD is requisite for the identification and development of therapeutics. Specifically, therapeutic development will follow the ability to identify genetic subtypes of ASD (both defined by gene and genetic etiology or mechanism) and will build off of an understanding of the function of individual genes identified. The often severe and debilitating nature of the ASD phenotype, combined with its high population prevalence, provides strong impetus to study and understand this disorder.

1.3 Identifying genes and pathways involved in ASD and ID

The identification of genes underlying ID and ASD has been most successful for syndromic Mendelian or monogenic disorders—for example, *FMRI* (Fragile X syndrome, Fu et al. 1991), *MECP2* (Rett syndrome, Amir et al. 1999) or *UBE3A* (Angelman syndrome, Matsuura et al. 1997). Together, however, these syndromes are estimated to account for less than 10% of ASD/ID, suggesting the presence of additional genes and etiologies. Initial population-based studies failed to identify single genes of major effect and few major common risk variants have been replicated, despite the strong observed heritability of these diseases (Steffenburg et al. 1989; Bailey et al. 1995; Hallmayer et al. 2011; Constantino et al. 2013). By contrast, targeted and genome-wide microarray studies revealed that large *de novo* CNVs were significantly enriched among probands when compared to unaffected siblings and/or controls (Marshall et al. 2008; Sebat et al. 2007; de Vries et al. 2005; Levy et al. 2011; Sanders et al. 2011; Cooper et al. 2011; Sharp et al. 2006), a finding that echoed the earlier discovery of large chromosomal aberrations in ASD and ID. Both initial and subsequent higher-resolution studies estimate that 8% of sporadic ASD cases carried a *de novo* CNV, as compared with only 2% of unaffected siblings (Levy et al. 2011; Sanders et al. 2011). Furthermore, among children

with general developmental delay (DD) and ID, rare large *de novo* CNVs are thought to account for up to 15% of disease burden (Cooper et al. 2011). Although individually rare, some of these CNVs were in fact recurrent mutations, mediated by locus-specific genomic instability (Sharp et al. 2006), and many of these *same* recurrent CNVs observed initially in patients with ID (Sharp et al. 2008) or ASD (Miller et al. 2009) have been identified in adults with epilepsy (Helbig et al. 2009), bipolar disorder (Ben-Shachar et al. 2009), or schizophrenia (Stefansson et al. 2008; International Schizophrenia Consortium 2008), suggesting overlap in the genetic etiology of these disorders.

The discovery of an aggregate burden of large *de novo* CNVs and the identification of recurrent events signaled a new paradigm for ASD and ID genetics. While specific CNVs are individually rare, combined they account for a significant fraction of cases, indicating the presence of considerable locus heterogeneity of ASD and ID. The *de novo* nature of these CNVs, together with their absence in the general population, suggests they represent a class of highly deleterious and highly penetrant mutations. Their underlying genetic model does not explicitly fit a recessive model of disease, since CNVs are primarily present as hemizygous deletions or duplications. These mutations alter the dosage of genes but do not completely abolish their presence. Collectively, these observations support a complex disease/rare variant model for ASD, in which a proportion of etiologic risk is conferred by very rare variants and *de novo* mutations.

The commoditization of next-generation or “massively parallel” sequencing represents a turning point in human genetics and makes it possible to discover sequence-level variants across nearly all coding regions (“the exome”) or the whole genome. These methods were first applied to confirm point mutations underlying Mendelian disorders (Ng et al. 2009c), and subsequent pilot studies demonstrated that family-based (trio) exome sequencing could discover pathogenic mutations in simplex ID (Vissers et al. 2010) or ASD (O’Roak et al. 2011). In the past year, this paradigm of *de novo* mutation discovery using exome sequencing of parent-child trios has been expanded to about one thousand ASD or ID families, resulting in the first detailed picture of how *de novo* coding mutations contribute to these disorders.

In this introduction, I synthesize the results of recent large-scale exome sequencing studies of ASD and ID (O'Roak et al. 2012b; Neale et al. 2012; Iossifov et al. 2012; Sanders et al. 2012; de Ligt et al. 2012; Rauch et al. 2012) and summarize their implications for human neurodevelopmental genetics. There are three themes: 1) Exome sequencing of ASD/ID families has revealed a significant excess of *de novo* mutations in probands when compared to unaffected siblings and has identified novel candidate genes contributing to the neurological deficits. I note that the strongest effects are observed for *de novo* loss-of-function (or truncating) mutations, which prematurely truncate the protein due to frameshift and nonsense mutations. 2) Both CNV and exome sequencing data suggest that no single gene will account for more than 1% of autism cases; rather, rare mutations in hundreds of genes may contribute to ASD or ID. 3) Analyses of network connectivity further implicate potentially important neurodevelopmental and synaptic pathways in ASD and ID.

Study details			N	Synonymous ^b	Missense ^c	Nonsense, splice site, indels
Iossifov 2012 (Iossifov et al. 2012)	ASD ^a	Quads	343	79:69	207:207	59:28
Sanders 2012 (Sanders et al. 2012)	ASD ^a	Quads	200	29:39	110:82	15:5 ^c
O'Roak 2012 (O'Roak et al. 2012b)	ASD ^a	Quads	50	14:16	40:31	6:3
		Trios	159	54	115	29
Neale 2012 (Neale et al. 2012)	ASD	Trios	175	50	101	18
de Ligt 201 (de Ligt et al. 2012)	ID	Trios	100	16	48	14
Rauch 2012 (Rauch et al. 2012)	ID	Trios	51	11	56	20
Sum				122:124	357:320	80:36
Odds Ratio (ASD)			1,078	0.98	1.29	2.41
(95% CI)				(0.73–1.31:1)	(1.01–1.63)	(1.58–3.75)

^aTrios/quads from the Simons Simplex Collection (SSC).

^bCounts refer to the number of mutations in probands or, when separated by a colon, to probands and siblings (e.g., probands:siblings).

^cNot including indels.

Table 1.1: Six recent family-based exome studies of ASD and ID

1.4 An increase in *de novo* loss-of-function mutations

Both *de novo* CNVs and single nucleotide variants (SNVs) can have, in principle, similarly disruptive effects on genes. Crucially, however, the detection of *de novo* SNVs yields gene-level specificity, thus allowing individual pathogenic genes and neurobiological pathways to be identified. Moreover, a small subset of the *de novo* mutations (~4% for unaffected and ~9% of affected; Iossifov et al. 2012; Sanders et al. 2012) are disruptive (e.g., frameshift, premature stop codon, splice-donor defect) with respect to the protein's biological function. Recurrent mutations of this type for a specific gene can strengthen the probability that the *de novo* mutation relates to phenotype. Since *de novo* protein-altering SNVs are collectively more common mutation events (~1/generation) than large *de novo* CNVs (~0.02/generation), there is the exciting possibility that this type of mutation may explain a larger fraction of genetic etiology of ASD.

In all, six recent exome sequencing studies of trios (mother, father and affected child) and quads (also includes an unaffected sibling) of sporadic ASD (Iossifov et al. 2012; Sanders et al. 2012; O'Roak et al. 2012b; Neale et al. 2012) or sporadic ID (de Ligt et al. 2012; Rauch et al. 2012), together comprising 1,078 families (Table 1.1) have been performed. Three of the ASD studies included unaffected siblings in order to compare mutation rates between affected probands and siblings. While these studies found a slightly elevated rate of mutation in probands versus their unaffected siblings (1.02 vs. 0.79 mutations per offspring), the type of mutation was critical: probands had two- to threefold more disruptive *de novo* mutations in comparison to their siblings, or to a random model of mutation (Sanders et al. 2012; Iossifov et al. 2012). Overall, among the 593 ASD quads, there were 80 such mutations in probands with ASD, but only 36 in siblings (OR = 2.41, $p < 1 \times 10^{-4}$, Fisher's exact test; Table 1.1). The reported enrichment of missense mutations in probands has been less robust, with study estimates for enrichment between 1- and 1.34-fold, but analysis of all quads does show weak statistical enrichment (OR = 1.29, $p = 0.03$). It is likely that some missense mutations are pathogenic while others are benign, a distinction likely dependent on the context of the mutation and affected proteins themselves.

Overall, these studies suggest that protein-truncating *de novo* SNVs contribute to the risk of ASD for about 10-15% of probands (Sanders et al. 2012; Iossifov et al. 2012), though this fraction is almost certainly a conservative estimate, as an unknown fraction of *de novo* events are still missed using current sequencing methods and bioinformatics tools. It is important to note that the six current exome studies focused primarily on a *de novo* mutation genetic model for the development of disease. Recent results highlight the effect of transmitted CNVs (Poultney et al. 2013; Krumm et al. 2013), as well as a renewed emphasis on the effect of common variation in ASD (Klei et al. 2012) based on study of data generated from the same samples.

1.5 Candidate genes yet few recurrent hits

Many strong neurobiological candidates have emerged from the genes disrupted by *de novo* mutations in these studies, including mutations in previously identified ASD/ID genes. Several mutations were identified in *NRXN1* and *NLGN1*; both are central components of the neurexin-neuroligin synaptic cell adhesion complex (Kim et al. 2008). Numerous *de novo* mutations were identified in genes or loci linked to Mendelian disorders, many of which have features of ASD or ID. These loci and genes include *MBD5* (mental retardation, autosomal dominant #1, OMIM 156200), *CHD7* (CHARGE syndrome, OMIM 214800), *PTEN* (Cowden syndrome, OMIM 158350), *DYRK1A* (in Down syndrome critical region, OMIM 190685), *TSC2* (tuberous sclerosis, OMIM 613254), *SETBP1* (Schinzel-Giedion syndrome, OMIM 269150), and *RPS6KA3* (Coffin-Lowery syndrome, OMIM 303600). Finally, mutations in several genes mapped to critical deletion regions or association intervals initially discovered by large CNVs, including mutations in *SYNRG* (17q12 deletion syndrome), *POLRMT* (19p13.3 deletion), and *CTTNBP2*—a potential candidate for the *AUTS1* (7q31) deletion locus.

Recurrently mutated genes, however, were few. In sum, only three genes (*CHD8*, *SCN2A*, and *SYNGAP1*) had two independent truncating *de novo* mutations in any single study, and no gene had more than three mutations. Models designed to estimate the significance of recurrent *de novo* mutations based on gene size and context found

nominal significance for *CHD8*, *NTNG1* (O'Roak et al. 2012b), and *SCN2A* (Sanders et al. 2012), but most genes could not be conclusively implicated. Notably, however, a review of case-control data by Neale and colleagues of 935 cases found three additional truncating *CHD8* mutations and one splice-site mutation in *SCN2A*, further strengthening the initial disease associations of these genes (Neale et al. 2012). In addition, within a few weeks of these initial reports, a *de novo* translocation was discovered mapping to *CHD8* (Talkowski et al. 2012).

1.6 Large-scale resequencing of candidate genes

Given the low rate of recurrence among genes with *de novo* mutations, estimates of overall locus heterogeneity for ASD have yielded between 300 and 1,000 genes that could confer increased ASD risk when subjected to *de novo* mutation (Figure 1.1). Even if exome sequencing prices continue to fall, the cost to confirm the association for a significant fraction of these genes remains impractically high, especially if thousands or tens of thousands of samples are required as has been suggested by CNV studies. Instead, targeted next-generation resequencing of candidate genes has proven to be instrumental in associating specific genes. In particular, de Ligt and colleagues resequenced five candidate genes in a confirmation series of 765 ID patients, identifying additional mutations in *CTNNB1* and *GATAD2B* and markedly strengthening their association with ID. Similarly, we have successfully used a molecular inversion probe (MIP) assay to capture and sequence 44 candidate genes in 2,446 ASD probands (O'Roak et al. 2012a). MIP resequencing generates complete sequence across targeted regions, can be performed at high scale and low cost (under \$1 per gene per sample), and delivers higher sensitivity for targeted loci than exome sequencing due to increased sequence coverage. Altogether, this assay yielded 27 new *de novo* mutations across 16 genes; of these mutations, 17 were disruptive SNVs, a fraction higher than expected by chance. The discovery of these mutations confirmed the association with ASD for *CHD8* and *DYRK1A* and provided significant statistical support for four novel genes: *GRIN2B*, *TBR1*, *PTEN*, and *TBL1XR1*.

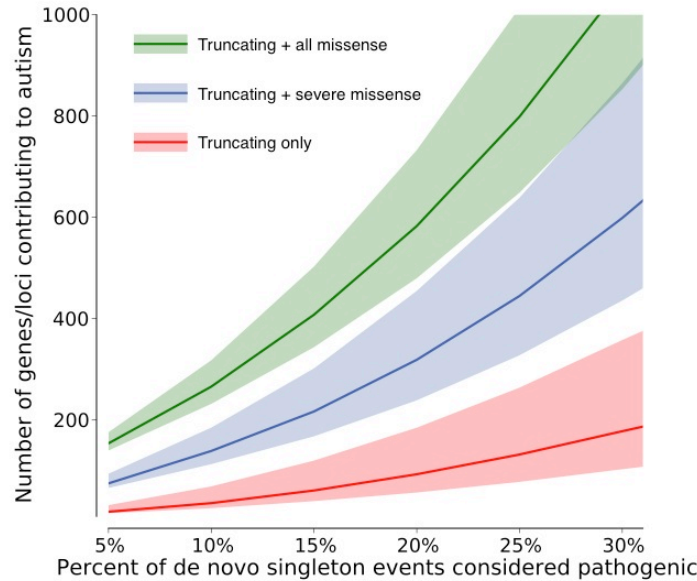


Figure 1.1: Estimating the number of ASD/ID risk genes. We estimate the number of ASD and ID genes, using an adaptation of the “hidden species problem” based on the ratio of genes with multiple de novo mutations to all genes with de novo mutations. For each estimate, all genes with recurrent de novo mutations are considered pathogenic, as well as a defined fraction of mutations in genes observed just once (since all de novo mutations are unlikely to be pathogenic). Including more of these singleton mutations as pathogenic, as well as including a broader range of mutation type, exponentially increases the number of ASD and ID risk loci. Thus, considering a disease model in which 15% of all truncating de novo mutations are sufficient and pathogenic, only ~50 genes are expected to be similarly sufficient in their pathogenicity; however, by including missense mutations, the number of loci rises dramatically (to over 400 loci when 15% of de novo SNVs are considered pathogenic). Taken together, this model highlights the locus heterogeneity underlying the genetic etiology of ASD and ID and suggests that the etiology of a large proportion of ASD/ID cases may not be due to a single de novo mutation (truncating or missense); rather, these cases may be the result of a complex set of interactions between multiple mutations, including SNVs, indels, and CNVs. Shaded area indicates 95% confidence intervals around estimate.

In sum, when considering only protein-disruptive mutations from six exome sequencing studies (four ASD and two ID) and including the resequencing of some of these candidate genes, a set of 11 genes (Table 1.2) show enrichment in cases with ASD/ID and account for approximately 2.2% of all cases. We have summarized the distribution of mutations, as well as the prevalence and coding location of mutations found in exome sequence from 6,503 samples from the NHLBI Exome Sequencing Project (ESP) in Table 1.2. For several of these genes with recurrent *de novo* hits in ASD probands (*CHD8*, *GRIN2B*,

DYRK1A), no truncating variants were observed in the ESP. Furthermore, while control mutations are sometimes found in genes in high frequency (e.g., frameshift in *SYNGAP1* at 3.2% frequency in controls), these mutations are found exclusively near the carboxy terminus of the protein and outside of functional protein domains and are unlikely to affect protein function (Figure 1.2).

Gene	ID Cases	ASD Cases	Summary	ESP Samples	
				Variants	Frequency
<i>CHD8</i>	-	9/2,446	2 (O), 7 (O*) [+ 3 (N*)]	0	0/6503
<i>SCN2A</i>	3/151	2/593	1 (L), 2 (R), 2 (S) [+ 1 (N*)]	1	7/6503
<i>SYNGAP1</i>	3/151	-	1 (L), 2 (R)	1	207/6503 ^a
<i>GRIN2B</i>	-	3/2446	1 (O), 2 (O*)	0	0/6503
<i>DYRK1A</i>	-	3/2446	1 (I), 1 (O), 1 (O*)	0	0/6503
<i>ZNF292</i>	1/151	1/593	1 (L), 1 (N)	1	2/6503
<i>POGZ</i>	-	2/593	1 (I), 1 (N)	1	1/6503
<i>KATNAL2</i>	-	2/593	1 (O), 1 (S)	1	1/6503
<i>TBR1</i>	-	2/2446	1 (O), 1 (O*)	0	0/6503
<i>CTNNB1</i>	1/151	1/2446	1 (L), 1 (O*), [+ 1 (L*)]	0	0/6503
<i>SETBP1</i>	1/151	1/593	1 (O), 1 (R)	3	58/6503 ^a
<i>ADNP</i>	-	2/2446	1 (O), 1 (O*)	1	1/6500
<i>LRP2</i>	1/151	1/593	1 (I), 1 (L)	6	53/6500
<i>ARID1B</i>	-	2/2446	1 (O), 1 (O*)	5	314/6500

Table 1.2: Recurrent disruptive mutations in ID and ASD Genes with two or more de novo truncating mutations observed in studies of ASD or ID. Summary indicates studies in which mutations were discovered. I, Iossifov et al. (Iossifov et al. 2012); L, de Ligt et al. (de Ligt et al. 2012); N, Neal et al. (Neale et al. 2012), O and O*: O’Roak et al. (O’Roak et al. 2012a; 2012b); R, Rauch et al. (Rauch et al. 2012); S, Sanders et al. (Sanders et al. 2012). Mutations found in secondary replication screens or case-control studies indicated in [brackets] with starred (*) reference. Truncating events found in the ESP database and their population frequencies are shown. (a) The truncating variants found in the EVS database in *SYNGAP1* and *SETBP1* genes fell at the extreme 3' end of the gene, suggesting that they do not adversely affect gene function.

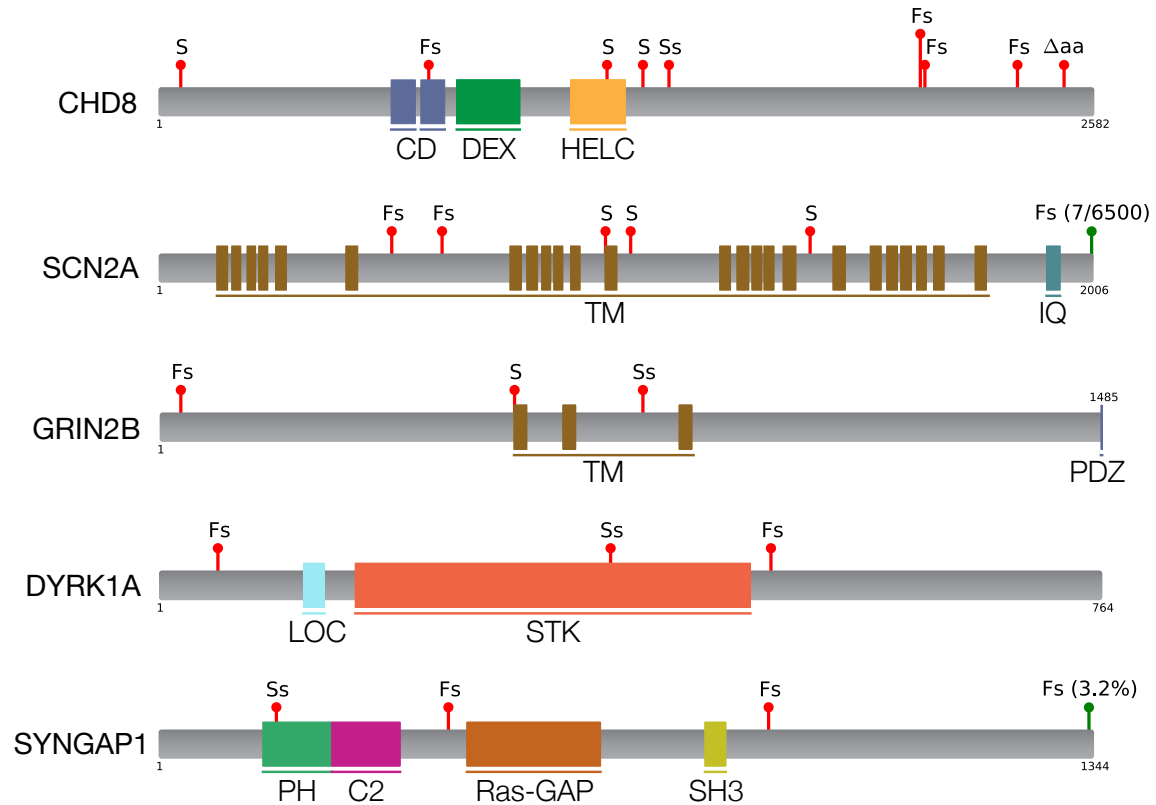


Figure 1.2: Location of de novo truncating mutations in top five ASD and ID genes. Red markers indicate locations of de novo mutations in ASD and ID cases; green markers indicate locations of truncating mutations in ESP database of over 6,500 samples (see Table 2.2 for details). Mutation codes: S, Stop-gain; Fs, Frameshift; Ss, Splice-site mutation; ΔAA amino-acid loss (non-frameshifting). Blocks indicate annotated protein domains from UniProt. Domain names, top to bottom: CD, chromodomain; DEX, Helicase ATP-binding; HELC, Helicase C-terminal; TM, transmembrane domain; IQ, IQ domain; PDZ, PDZ-binding motif; LOC, Bipartite nuclear localization signal; STK, Serine/Threonine protein kinase; PH, Pleckstrin homology domain; C2, C2 domain; SH3, SRC Homology 3 Domain.

1.7 Novel candidates and their neurobiology

Many of the top genes from recent exome studies are novel candidates for ASD and ID, including the strongest overall association: *CHD8*, an ATP-dependent chromodomain helicase that directly regulates *CTNNB1* (Beta-catenin; Nishiyama et al. 2012) as well as the p53 pathway (Nishiyama et al. 2009). The CHD8 protein has known binding activity with another chromodomain helicase, CHD7 (Batsukh et al. 2010), which is the key protein in CHARGE syndrome, a rare syndrome with high ASD comorbidity (Betancur

2011). In addition to directly interacting, both are homologues of the *Drosophila* trithorax group protein *kismet* and are components of large chromatin remodeling complexes thought to be important in neural crest cell differentiation (Bajpai et al. 2010). Overall, eight *de novo* truncating mutations were observed across 2,597 cases in this gene; in contrast, no such mutations were observed in control siblings, or in over 6,500 exomes in the ESP. The frequency of mutations in this gene is the highest of all genes screened thus far and nearly matches that of CNVs at 16p11.2, which is the most frequent recurrently deleted (0.5%) or duplicated (0.3%) locus in sporadic ASD (Kumar et al. 2009; Walsh and Bracken 2011).

The second strongest overall association, with two truncating mutations in ASD cases and three such mutations in ID cases, is *SCN2A*, a gene previously associated with epilepsy and seizure disorders (Kamiya et al. 2004; Ogiwara et al. 2009). *SCN2A* encodes a voltage-gated sodium channel (type II, alpha 1; Nav1.2) expressed throughout the brain, and is responsible for the generation and propagation of action potential in neurons. The phenotype associated with this gene appears to be highly variable. Given the smaller number of ID cases, the prevalence of mutations in *SCN2A* is higher in ID—however, one of the ID cases also shows signs of autism. Lastly, only one of the five cases had a history of seizures (de Ligt et al. 2012), suggesting that mutations in *SCN2A* have highly variable phenotypic outcomes.

Another striking candidate is *DYRK1A* (dual-specificity tyrosine phosphorylation-regulated kinase 1A), for which three truncating mutations have been discovered in autism probands (O’Roak et al. 2012b; 2012a) and *de novo* structural variants in ID probands (van Bon et al. 2011; Moller et al. 2008). *DYRK1A* is a highly conserved gene whose dosage imbalance has been implicated in the cognitive deficits associated with Down syndrome. The gene interacts with the SWI/SNF complex (Lepagnol-Bestel et al. 2009) and is considered a master regulator of brain growth, affecting diverse aspects of neurogenesis, including neuronal proliferation, morphogenesis, differentiation, and maturation (Mazur-Kolecka et al. 2012; Guedj et al. 2012; Yabut et al.). Mutations in the *Drosophila* ortholog (*mnf*) have been known for more than 20 years and result in a

“minibrain phenotype” where optic lobes and central brain hemispheres are reduced (Tejedor et al. 1995). Similarly, heterozygous mice knockouts for *Dyrk1A* (+/-) show a reduction of brain volume in a region-specific manner as well as mental impairment (Fotaki et al. 2002; Song et al. 1996). Consistent with these models, all three human loss-of-function autism patients are cognitively impaired and microcephalic (z-score < -2).

Three truncating mutations each of *GRIN2B* and *SYNGAP1*, and two mutations of *TBR1*, highlight the importance of excitatory/glutamatergic signaling in both ASD and ID—and are perhaps some of the most conclusive previously implicated genes to date. *GRIN2B* (found in an ASD case with ID) forms a subunit of an NMDA receptor associated with learning and memory, and targeted sequencing has linked it to neurodevelopmental disorders as well as its discovery in ASD (Endele et al. 2010). This receptor participates in a larger postsynaptic complex with *SYNGAP1*, in which three mutations in ID patients have been observed in the present cohorts, as well as in multiple previous screens of ID (Vissers et al. 2010; Hamdan et al. 2009). Interestingly, while no mutations of *SYNGAP1* were found in the Simons Simplex Collection (SSC) ASD cohorts, *SYNGAP1* mutations were recently implicated in several cases of ID with ASD (Berryer et al. 2012). Finally, *TBR1*, together with the *CASK* protein, regulates transcription of *GRIN2B* (as well as several other candidate ASD genes, such as *RELN* and *AUTS2*) (Bedogni et al. 2010).

1.8 Protein interaction networks converge on common pathways

Knowledge of molecular-level interaction between proteins has enabled the development of transcriptional networks (Voineagu et al. 2011) and protein-protein interaction (PPI) networks enriched for mutation in ASD and ID cases. These networks provide a powerful method to unify the landscape of mutations observed in genetically heterogeneous human disorders by leveraging regulatory interactions between genes and/or physical interactions between proteins. For example, Iossifov et al. found that 14/59 genes disrupted by *de novo* mutations were significantly enriched ($p < 0.006$) in a group of 842 genes previously defined (Darnell et al. 2011) as regulated by *FMR1*, the key protein disrupted in Fragile X syndrome, and noted that this was not true for mutations found in siblings (2/28 part of *FMR1*-regulated genes), the general population, nor for all missense

variants (Iossifov et al. 2012). Neale et al. performed a similar analysis to previously identified ASD and ID risk genes—including a core set of 31 synaptic genes identified from previous proteomic studies—and found that genes with nonsynonymous *de novo* mutations had a significantly reduced network distance (i.e., they were more closely associated in the network) than was a set of “comparator” genes derived from silent *de novo* mutations and sibling mutations (Neale et al. 2012). Lastly, we developed a network for interactions between proteins corresponding to genes with *de novo* mutations, revealing a single connected component for 39% (49/126) of genes with disruptive or likely disruptive missense *de novo* mutations (O’Roak et al. 2012b). Notably, in follow-up MIP resequencing, we targeted ~50% network and ~50% non-network genes and found that 94% (16/17) of the newly discovered truncating mutations fell within the network (or a similar, expanded 74-gene network)—an observation unlikely to have occurred by chance ($p = 0.0002$). In contrast, the non-network genes had only six total mutations, only one of which was a truncating mutation.

I integrated the results from the six exome studies by forming PPI networks using experimentally verified interaction data from StringDB (Jensen et al. 2009) (Supplemental Methods). I found the PPI network based on all truncating and missense mutations in probands was significantly more clustered, had more edges, and created larger connected components than randomly sampled or permuted networks ($p \leq 0.009$ for all tests; Supplemental Methods); in contrast, neither the genes with mutations in siblings, nor those with synonymous mutations (in either proband or siblings) showed any difference from the null distribution of networks (Table S1).

In order to summarize these PPI networks, I connected all truncating mutations as well as six genes with missense mutations with important roles in brain development (Figure 1.3; Supplemental Methods). The two largest connected components of this combined network encompass three broad functional pathways: the first connected component (13 proteins) forms a highly interconnected set of postsynaptic scaffolding proteins and receptors, including SYNGAP1, DLG4, GRIN2A/B, NLGN1 and NRXN1, while the second (9 proteins) contains both *WNT*-signaling functions of CTNNB1, DLL1, and

TBL1XR1 and chromatin remodeling functions, anchored by the CHD8 protein. It is important to emphasize that while the nodes in the displayed network are partially based on a manually selected set of genes, the connected components formed are a strict subset of the unbiased PPI simulations described above and are larger than any connected component that can be formed using disruptive mutations found in siblings, synonymous changes, or randomly chosen genes.

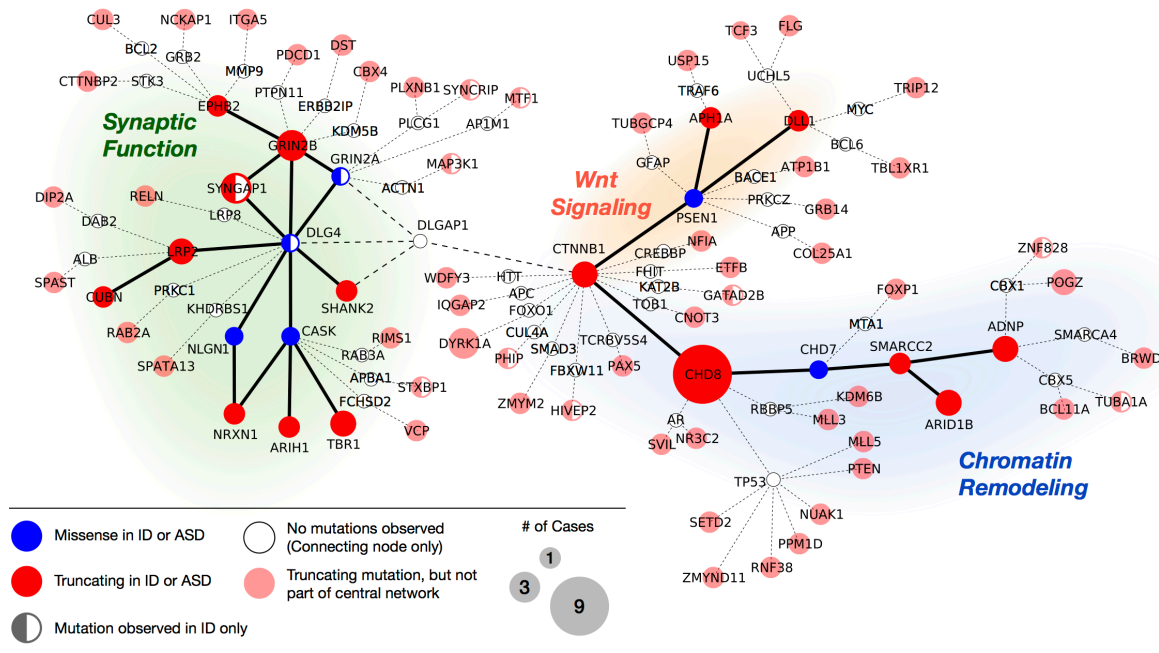
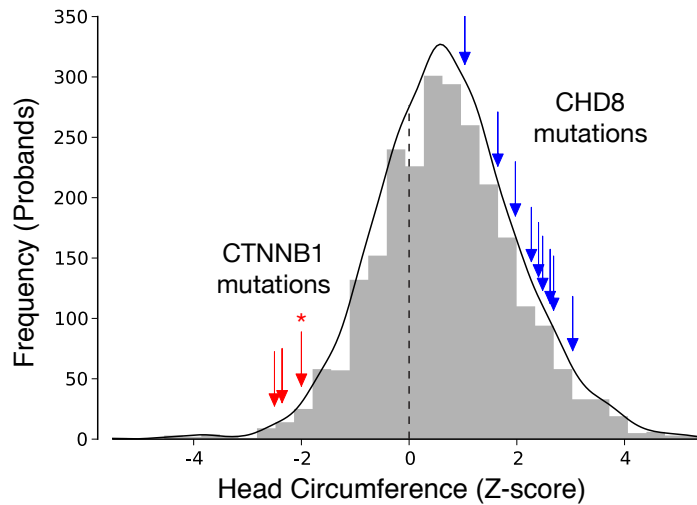


Figure 1.3: Genes disrupted by de novo mutations in ASD and ID form a central connected network. Genes with de novo truncating mutations (red nodes) or selected missense mutations (blue nodes) in four ASD exome studies and two ID exome studies are connected using experimentally derived PPI data from StringDB. Only medium- and high-confidence experimental interactions are shown, though we note that these may not always represent local interactions protein-protein interactions or interactions within the same subcellular compartment. Peripheral nodes (lighter shades) represent genes with additional truncating de novo mutations, which are separated from the central network by only a single node (white nodes; for this analysis we excluded SUMO1/SUMO2 and UBC, which are highly connected but nonspecific nodes).

Interestingly, Gilman et al., using a novel probabilistic framework (NETBAG) in conjunction with CNV data from SSC families, highlighted several genes and pathways with remarkable premonition and overlap to those found in the present exome-based studies (Gilman et al. 2011). In particular, their model showed enrichment of the

canonical *WNT* pathway, postsynaptic complexes, and dendritic spine development (e.g., *DLG4*, *SYNGAP1*) and several proteins involved in chromatin remodeling, including *BAZ1B* and *SMARCA2*, both of which interact with the central nodes of the chromatin remodeling network (Figure 1.3).

The *WNT* pathway and chromatin remodeling modules of the network are linked by interaction between *CHD8* and *CTNNB1*/Beta-catenin. Both of these proteins play important roles in neural development and growth: Beta-catenin, via downstream *WNT* pathways, influences neuronal migration, polarity and synaptogenesis (Salinas and Zou 2008), and constitutive overexpression of beta-catenin in mice results in macrocephaly (Chenn and Walsh 2003). *CHD8* negatively regulates beta-catenin via direct binding and, furthermore, downregulates beta-catenin responsive genes by recruitment to their promoter regions (Thompson et al. 2008). Strikingly, ASD cases with truncating mutations in *CHD8* have significant macrocephaly (O'Roak et al. 2012a), while all three cases with truncating mutations in beta-catenin have microcephaly (de Ligt et al. 2012; Bernier et al. 2014). These reciprocal phenotypes suggest that *CHD8* and beta-catenin form a regulatory network that controls head size by altering neuronal migration and growth during development (Figure 1.4). Other proteins with *de novo* mutations in this network include *TBL1XR1*, which binds beta-catenin (Cadigan 2008), and *DLL1*, which is expressed in neural progenitor cells and part of the Delta/Notch signaling pathway (Barton and Fendrik 2013).



b.



Figure 1.4: CHD8 and Beta-catenin/CTNNB1 putative regulatory network for head size. a) Truncating mutations in CTNNB1 (red arrows) and CHD8 (blue arrows) are found in patients with small and large head circumference, respectively. Gray histogram represents background distribution of age- and sex-corrected head circumference Z-scores for 2,446 probands from the SSC. (The exact head circumference for one case [marked with *] with clinically reported microcephaly could not be determined, so Z-score was estimated at -2.0 , or the clinical threshold for microcephaly). b) A putative regulatory model of head growth where CHD8 negatively regulates CTNNB1 (Thompson et al. 2008); CTNNB1 promotes head growth and constitutive over-expression of CTNNB1 in mice results in macrocephaly (Chenn and Walsh 2003).

Convergence on a second common pathway—chromatin remodeling—has primarily been driven by overlap between genetic syndromes and *de novo* mutations in sporadic ASD and ID. As discussed, CHD8 possesses ATP-dependent chromatin remodeling activity and directly interacts with CHD7 (Batsukh et al. 2010), which is responsible for CHARGE syndrome, a complex syndrome in which up to two-thirds of patients have been found to have ASD (Betancur 2011). Several *de novo* missense mutations in ASD cases have been noted in genes encoding for chromodomain helicase proteins, including CHD7 and CHD3, and a *de novo* frameshift in CHD2 was found by Rauch et al. in an ID case (Rauch et al. 2012). A second syndrome, Coffin-Siris syndrome (OMIM 135900), characterized by ID and severe speech delays, was recently attributed to truncating

mutations or disruptive CNVs in *ARID1B* (encoding a subunit of the SWI/SNF chromatin remodeling complex; Santen et al. 2012), and one *de novo* frameshift of *ARID1B* was found in a sporadic ASD case (O'Roak et al. 2012b). Additional disruptive *de novo* mutations recognized in *ASH1L*, *KDM6B*, and *MLL5* suggest that the chromatin remodeling activity of these proteins may be an underlying pathway implicated in ASD and ID (Iossifov et al. 2012). Finally, we note that mutations in *KANSL1* (né *KIAA1267*), a histone acetyltransferase with similar p53 regulatory activity to *CHD8*, were recently found to underlie 17q21.31 microdeletion syndrome, in which ID is a characteristic feature (Koolen et al. 2012). However, no mutations in *KANSL1* have been found in ASD cases, though this is likely due to exclusion of known clinical syndromes from these cohorts.

In addition to these newly proposed pathways, *de novo* mutations also highlight the importance of genes with roles in synaptic function and localization—a pathway previously suspected to be disrupted in ASD (Glessner et al. 2009). Many of these genes with *de novo* mutations form a closely related network of postsynaptic proteins, including the GTPase activating protein SYNGAP1, NMDA receptor subunits GRIN2B and GRIN2A, the scaffolding proteins DLG4 and CASK (the underlying mutation in CASK syndrome, OMIM 300749), and NRXN1, which has been previously associated with ASD (Kim et al. 2008). In conjunction with TBR1, CASK also transcriptionally activates several known neurodevelopmental genes, such as *RELN*, a gene with critical roles in neuronal development, synaptogenesis, and plasticity (Wang et al. 2004). Finally, this pathway is closely linked to *SHANK3*, a previously identified ASD protein with up to 1% mutation frequency in ASD cases (Durand et al. 2007; Moessner et al. 2007), although no mutations in this gene have been identified in the six studies presented here. While the reasons for this are not fully clear, it is likely that the high GC content of the gene impedes current short-read sequencing platforms.

Interaction networks (Figure 1.3) can also suggest novel targets for mutation screens or functional studies. For example, while DLGAP1 plays a central role in connecting the “Synaptic Function” component to beta-catenin, no mutations have been observed in

DLGAP1. Similarly, SMARCA4 connects BRWD1 to the in-network ADNP protein. These proteins, as well as other “nearby” proteins suggested by PPI networks, can provide novel targets for mutation screens and deeper functional/pathway study. It is likely that sequencing studies of patients will identify novel candidates for PPI networks, creating a reiterative process by which networks and genetics mutually inform.

Despite their widespread role in the current study of ASD and ID, PPI networks have several important limitations. First, protein interactions are difficult to assay experimentally and often are not at a proteomic scale, resulting in false negatives and false positives in databases. In addition, the extent to which the temporal and spatial nature of interactions is captured also limited, and in our network we do not distinguish between different interaction types (regulatory or physical) or cellular compartments. For example, while CASK binds NRXN1 presynaptically (Fairless et al. 2008), binding to the transcription factor TBR1 is in the nucleus (Hsueh et al. 2000). Second, although our PPI network only uses experimentally verified interactions, the impact and weight of interactions can vary considerably for different nodes, especially for “hub” nodes which can interact with hundreds of other proteins. Finally, current PPI networks do not take into account the functional impact of mutations on the proteins or the interactions themselves.

1.9 Assaying copy number variation

Recent advances in technology have made possible genome-wide discovery of rare CNVs and the estimation of copy number for copy number polymorphisms (CNPs). Most commonly, array comparative genome hybridization (array-CGH; Pinkel et al. 1998) or single nucleotide polymorphism (SNP) microarray platforms (Peiffer et al. 2006) have been used to interrogate the copy number of thousands to millions of positions within the genome. Briefly, these technologies work by hybridizing fluorescent-labeled DNA to targeted oligonucleotide probes on a glass slide and use a scanner to measure the hybridization intensity of each fragment. Copy number variation increases or decreases this intensity, either relative to a control sample (for array-CGH) or in absolute units (for

SNP microarray platforms); a CNV can be inferred (“called”) from the presence of multiple nearby probes with increases or decreases in signal.

However, microarray-based technology has several significant limitations. First, the minimum size of a CNV is limited by the number of probes printed on the microarray and further by the fact that generally at least 20 probes are required to distinguish a true CNV from random noise. Thus, even with high-resolution microarrays of 1 million or more probes, studies have typically examined CNVs greater than 50 or 100 kbp in size. A second limitation of microarray platforms is their poor performance within segmental duplications in the human genome (Bailey et al. 2002; Sudmant et al. 2010). Thus, the majority of studies to date have excluded the analysis of duplicated and CNP loci.

The advent of whole-genome sequencing using “next-generation” short-read technology has resulted in the development of several additional methods for CNV/CNP discovery and genotyping. Two broad approaches exist: the first are read-depth methods which leverage the relatively uniform distribution of sequenced reads throughout the genome (the “read-depth”) to estimate copy number; these methods can estimate the copy number of genomic segments as small as 1 kbp (Alkan et al. 2009; Yoon et al. 2009; Chiang et al. 2008). The second class of CNV discovery tools is based on the “paired-end” alignment of sequences from the two ends of fragment of DNA of known approximate size: when the ends of the fragment align to distant genomic loci, a deletion in the fragment can be inferred, and similar rules can be used to infer insertions and (unique to this method) inversions (Hormozdiari et al. 2009; Korb et al. 2009; 2007). Paired-end approaches are able to find very small CNVs less than 1 kbp in size, though they can suffer from a high false positive rate and require a large amount of sequence data as well as computational resources.

1.10 Thesis goals

This introduction and review of published literature has emphasized the role of *de novo* variation in ASD and ID and has cast light on how genes recurrently affected by *de novo* loss-of-function mutation in probands are part of functionally important and connected pathways. However, the rarity and effect size (versus unaffected siblings) of *de novo*

mutations in ASD cannot fully explain the overall heritability of the disorder, suggesting that additional underlying genetic etiologies, mechanisms and effects also play a role in ASD. Furthermore, given the extensive heterogeneity of both genotypes and phenotypes seen in ASD, it is likely that multiple genetic etiologies underlie ASD risk in families, and that any explanation of the genetics of ASD is incomplete if based on *de novo* mutations alone.

Therefore, the first aim of this dissertation is to fully explore the spectrum of mutations that contribute to ASD risk. To do this, I develop novel methods that leverage exome sequencing data to assay small CNVs, especially those under 100 kbp that have been missed using traditional assay methods such as array-CGH. Chapter one describes CoNIFER (copy number inference from exome reads), a novel computation method that uses exome read-depth data to infer copy number. First, I establish the sensitivity and precision of the method compared to standard high-resolution microarrays, and by forward validation of novel events. I demonstrate that exome-based detection of rare CNVs has up to 10-fold higher sensitivity for small events less than 10 kbp. In addition, I show that exome-based copy number correlates with true copy number for regions with 0-10 copies, suggesting that exome-based methods can be used for quantitative assessment of copy number. Knowledge of these smaller CNVs, many of which affect only a few exons of a single gene, fills out the spectrum of mutations, from the smallest single base-pair mutations through small and large CNVs and provides a more complete picture of the genes and pathways involved in ASD.

The second aim of this dissertation is to understand the role of rare inherited mutations in ASD. In chapter two, I apply CoNIFER to exome data from 411 ASD families and examine the pattern of inherited variation between the ASD-affected probands and their unaffected siblings. I show that rare inherited CNVs confer increased risk for ASD, even in the context of “sporadic” autism. These inherited variants are correlated with specific ASD-related phenotypes, including IQ and social ability. Finally, inherited variants found to be associated with ASD also are more likely to affect highly brain-expressed genes and are more likely to be part of existing disease and disorder pathways.

In chapter three, I expand the hypothesis that inherited events contribute to ASD risk to SNVs. Using an expanded resource of nearly 2,400 families, I demonstrate that ultra-rare SNVs carry pathogenic risk for ASD similar to that seen in CNVs and find that this risk is specifically related to loss-of-function SNVs in genes that do not tolerate functional (i.e., deleterious) variation in control populations. These SNVs also mirror CNVs in that they are correlated with specific phenotypes, in particular the IQ of ASD individuals.

This dissertation describes the integration of multiple genetic etiologies (*de novo* and inherited) and types of mutations in the context of simplex ASD. First in chapter two, and more fully in chapter three, I utilize a logistic regression model of ASD risk to directly compare the effects of *de novo* and inherited variation, for both CNVs and SNVs. These models suggest that all four combinations contribute an independent statistical component of risk for the development of ASD.

Finally, I also examine how multiple types and classes of mutation converge on specific genes and interactions and suggest new ASD candidate genes as well as an integrated etiology for risk of autism (Chapter III). In particular, convergence of *de novo* SNVs and inherited CNVs suggests that *CSMD1*, a complement control protein, may be an ASD risk factor. This gene is particularly interesting in the context of a neurodevelopmental disorder in that it displays strong and specific pan-brain expression, participates in dendritic spine restructuring, and has been implicated in other disorders with neurological basis, such as schizophrenia. I also examine how multiple mutations within one individual may lead to ASD. One case, described in chapter three, has mutations in two previously identified ASD genes that are part of the same complex (*NLGN2* and *NRXN3*). Underscoring the importance of the entire spectrum of genetic mutations, this interaction was discovered only by examination of inherited and *de novo* mutations, as the *NRXN3* mutation was an inherited CNV and the *NLGN2* mutation was a *de novo* SNV.

II. Copy number variation detection and genotyping from exome sequence data

2.1 Summary

While exome sequencing is readily amenable to single nucleotide variant discovery, the sparse and non-uniform nature of the exome capture reaction has hindered exome-based detection and characterization of genic copy number variation. We developed a novel method using singular value decomposition (SVD) normalization to discover rare genic copy number variants (CNVs) as well as genotype copy number polymorphic (CNP) loci with high sensitivity and specificity from exome sequencing data. We estimate the precision of our algorithm using 122 trios (366 exomes) and show that this method can be used to reliably predict both *de novo* and inherited rare CNVs involving three or more consecutive exons. We demonstrate that exome-based genotyping of CNPs strongly correlates with whole-genome data (median $r^2 = 0.91$), especially for loci with fewer than eight copies, and can estimate the absolute copy number of multi-allelic genes with high accuracy (78% call level). The resulting user-friendly computational pipeline, CoNIFER (copy number inference from exome reads), can reliably be used to discover disruptive genic CNVs missed by standard approaches and should have broad application in human genetic studies of disease.

This chapter has been published: Krumm, N., Sudmant, P. H., Ko, A., O’Roak, B. J., Malig, M., Coe, B. P., et al. (2012). Copy number variation detection and genotyping from exome sequence data. *Genome Research*, 22(8), 1525–1532. doi:10.1101/gr.138115.112

2.2 Introduction

Targeted capture and sequencing of coding exons (“exome sequencing”) has revealed common single-nucleotide polymorphisms (SNPs), rare sequence variants, short indels, and breakpoints of structural variation (Ng et al. 2009b; for review see Bamshad et al. 2011), but has been largely refractory to the discovery of copy number variants (CNVs). In contrast to whole-genome sequencing data, exome capture and sequencing results in non-uniform read-depth between captured regions and strong systematic biases between batches of samples. These biases, as well as the sparse nature of the capture, make exome sequencing unsuitable for “traditional” CNV detection algorithms, such as raw read-depth (Alkan et al. 2009), (Yoon et al. 2009), read-pair alignment (Hormozdiari et al. 2009) or split-read mapping (Karakoc et al. 2011). In this study, we combine read-depth data from exome sequencing with singular value decomposition (SVD) methods to discover rare CNVs and genotype known copy number polymorphic (CNP) regions from eight HapMap samples and 122 autism spectrum disorder (ASD) mother-father-proband trios sequenced as part of separate study to primarily discover *de novo* SNPs and indels (O’Roak et al. 2011). We validated the discovered events using orthogonal datasets, including whole-genome sequencing and tiling array comparative genomic hybridization (array-CGH) data for HapMap samples, and SNP array and quantitative PCR for events discovered in the autism trios. In light of the tens of thousands of exomes already sequenced, we believe this method will have widespread application for the discovery and association of both rare and common copy number variation in disease, and will complement existing methods to discover single-nucleotide variation from exome-sequencing data.

2.3 Methods

Samples and datasets

We used exome sequencing data from eight HapMap individuals (NA12878, NA15510, NA18507, NA18517, NA18555, NA18956, NA19129, and NA19240; available in the NCBI Sequence Read Archive under accession SRA039053 or SRP007298) and exomes from 122 mother-father-proband ASD trios (for 366 total individuals). In addition, we utilized exome data from 533 individuals from the NHLBI Exome Sequencing Project

(ESP) as a means to derive accurate estimates of the distribution of sequence coverage at each exon. Underlying exome sequence data are available from the Short Read Archive for the HapMap exomes (SRA039053) and from the dbGaP exchange area (for ASD exomes: phs000482.v1.p1; for ESP exomes: phs000 279.v1.p1, phs000290.v1.p1, phs000291.v1.p1, phs000281.v1.p1, phs000254.v1.p1), with additional cohorts pending; more information at <http://evs.gs.washington.edu/EVS/>). All exomes were captured using either the Roche NimbleGen EZ Exome SeqCap Version 2 (for ESP samples and ASD trios) or Version 1 (for HapMap samples) in-solution exome capture kits (44 Mbp captured, including 36 Mbp exon target). Short-read sequencing was performed using either an Illumina HiSeq2000 platform or an Illumina GAI, with a mix of 50 bp and 76 bp paired-end reads (Table 2.1; see Supplementary Note for additional details).

Cohort	# Samples	Capture Version	Passed QC	Average Number of Mapped Reads	Average Number of Mappings
HapMap	8	Roche NimbleGen EZ Exome SeqCap Version 1	8	138,593,483	158,568,475
Autism Trios	122 probands and 244 parents		366	119,461,629	143,574,053
NHLBI Exome Sequence Project	613	Roche NimbleGen EZ Exome SeqCap Version 2	533	127,125,719	152,787,950

Table 2.1: Cohorts analyzed

Mapping

Sequence reads were divided into non-overlapping 36 bp constituents and mapped to exons and the 300 bp flanking sequence of the repeat-masked hg19 reference sequence using mrsFAST (Hach et al., 2010), allowing for up to two mismatches per 36 bp. We calculated RPKM (reads per thousand bases per million reads sequenced; Mortazavi et al. 2008) values for 194,080 exome capture targets (see Supplementary Note) and excluded from further analysis 3,964 probes with a median RPKM of less than one, as these probes were likely failed or improperly targeted.

Singular value decomposition

RPKM values were transformed into standardized z-scores (termed Z-RPKM values)

based on the mean and standard deviation across all analyzed exomes and organized into an exon-by-sample matrix (X). Using SVD, we decomposed X into three matrices:

$$X = USV^T$$

In order to remove the strongest k components, we set $S_1 \dots S_k$ to zero to form S' , and then recalculate X as the dot product of U , S' and V^T (Fig. 2.1). We termed these final values SVD-ZRPKM values—each of which represents the normalized relative copy number of an exon in a sample.

Validation

For this study, we specifically selected samples that had been subjected to extensive prior experimental validation. Copy number variation of the eight HapMap samples was previously assessed by whole-genome shotgun sequencing and targeted clone sequencing (Kidd et al. 2008), and data from the 1000 Genomes Project (Sudmant et al. 2010).

Accurate estimates of copy number for duplicated loci were determined experimentally by single-channel array-CGH data and qPCR (Sudmant et al. 2010; Campbell et al. 2011). CNV data for the 366 autism exomes was obtained by SNP Microarray (Illumina 1M) and by targeted array-CGH as described previously (O'Roak et al. 2011; Sanders et al. 2011).

CoNIFER implementation

We implemented our algorithm as a collection of python programs under the name CoNIFER (copy number inference from exome reads), available at <http://conifer.sourceforge.net>. CoNIFER can accept files containing BAM alignment files or RPKM values from samples and outputs a number of charts (e.g., scree plots), a text file containing calls, and images corresponding to each call. Additionally, the raw SVD-ZRPKM values can be saved, facilitating genotyping of CNP loci and further analysis. The computational resources to run CoNIFER are lightweight. BAM-format files can be converted into read-depth files in approximately 20 to 30 minutes; then, given read-depth or read-count values for targeted exons or probes, the CoNIFER and the SVD-normalization can be run with minimal hardware requirements (e.g., 500 samples processed in less than one hour using 4 GB or less of memory).

2.4 Results

Our method exploits differences in sequence read-depth from exome datasets to predict copy number variation (Fig. 2.1). We focused on characterizing two distinct classes of genetic variation: rare CNVs and CNPs. The former are individually rare in populations (less than 1% frequency) and are predominantly found in unique regions of the genome. In contrast, CNPs are common, both between individuals and between populations, and are frequently associated with segmental duplications (Girirajan et al. 2010). The absolute copy number of multi-allelic CNPs embedded in segmental duplications ranges widely from zero to more than 40 copies, and this variation is typically referred to as multi-copy or multi-allelic (Sudmant et al. 2010). Crucially, as the total copy number of CNPs is estimated as the sum of both haplotypes (i.e., the copy number is not phased), independent re-assortment of parental haplotypes obfuscates the pattern of inheritance for CNPs between parents and offspring. Moreover, our approach utilizes relative read-depth values for each exon; for exons with highly diverse copy number across a population, the population standard deviation will be high as well, thus shrinking the range of relative values observed at that exon. In effect, this makes a threshold-based discovery algorithm less sensitive for CNPs and exons of high copy number diversity, but does not impact genotyping of these CNPs and exons when their location is known. Because of these fundamental differences, we chose to pursue the characterization of CNVs and CNPs differently: for CNVs, discovery within the exome data was unbiased by location, whereas for CNPs, we used *a priori* information regarding the location of copy number variable loci. Furthermore, when estimating the precision for rare inherited CNVs, we excluded calls within segmental duplications or CNP loci, as the inheritance pattern for these loci cannot be determined without phased copy number information.

For discovery of rare CNVs (Fig. 2.1, Fig. S1), we removed between 12 and 15 (k) singular values, a number we empirically adjusted based on the inflection point of the scree plot (Fig. S2). We set discovery thresholds at -1.5 or +1.5 SVD-ZRPKM for rare deletions and duplications, respectively, and required at least three exome probes to exceed the threshold (Supplementary Note). For genotyping CNP regions in the genome, we opted to remove only five components in order to prevent the SVD algorithm from

removing *bona fide* signal from highly CNP loci. The genotype value was calculated by determining the average of the SVD-transformed ZRPKM values for the exons/targets in the region of interest. As the output from our algorithm provides a relative value, we estimated absolute copy number from the SVD-ZRPKM values via two methods: 1) by using population frequency information of copy number states (Campbell et al., 2011) and 2) by creating a standard curve using copy number estimated from whole-genome sequencing data of matched HapMap samples (Sudmant et al. 2010); Supplementary Note).

An overview of our method is presented in Figure 2.1. Briefly, sequence reads from each exome were mapped to exons using mrsFAST (Hach et al. 2010), which allows for reads given a set edit-distance to map to multiple locations. Similar to RNAseq data analysis, we calculated RPKM values and transformed these into standardized z-scores (termed Z-RPKM values) based on the mean and standard deviation across all analyzed exomes. These were subsequently organized into an exon-by-sample matrix (X). Next, we implemented an SVD algorithm to overcome the systematic biases that pervade exome capture reactions. Since “singular values” can be used to examine the relative amount of contributed variance from each component, we used a plot of these singular values, known as a “scree plot” to identify this experimental noise. Our analysis reveals that the first 10–15 components disproportionally contribute to the variance of the data (Fig. S2). Given that we expect biological variation, in the form of rare CNVs as well as common CNPs, to be a minor contributor to the overall variance of the exon-by-sample matrix X , we formulated the basis of our algorithm by eliminating these strongest components. We selected the number of components for elimination based on the inflection point of the scree plot.

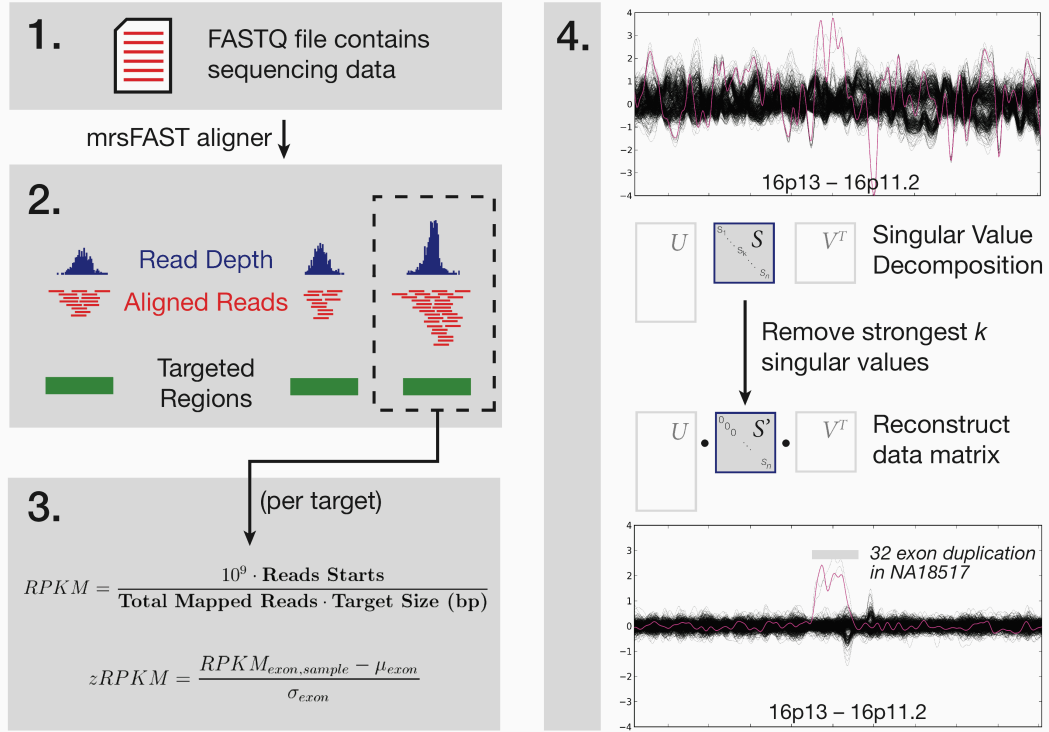


Figure 2.1: Method overview and CNV discovery (a) Exome sequencing reads from FASTQ files were divided into non-overlapping 36 bp constituents and (b) aligned to targeted regions, allowing for up to two mismatches per 36 bp alignment. (c) For each exon or targeted region, we calculated RPKM values and then transformed these into “ZRPKM” values based on the median and standard deviation of each exon across all samples. (d) ZRPKM values were inputted into the SVD transformation, where we removed the first 12-15 singular values. Finally, a centrally weighted 15-exon average was passed over the SVD-ZRPKM values in order to reduce false positives, and a ± 1.5 SVD-ZRPKM threshold is used to discover CNVs. Final image (e) shows ZRPKM values from 1,000 consecutive exons on chromosome 16, plotted for 533 ESP exome background samples (black traces) and NA18507 (pink trace). Blue bar corresponds to a rare duplication in NA18507 at the METTL9/OTOA locus at chr16p12.2 that was validated by SNP microarray CNV analysis.

The number of components selected for removal is an important parameter in our algorithm and warrants further consideration. Removing too few components leaves the algorithm at risk for residual systematic bias; conversely, removing too many components will begin to remove *bona fide* signal from exomes, especially at large, common segmental duplications within which a large proportion of analyzed exomes contribute strongly altered read-depth signal. However, individually rare CNVs do not contribute significantly to the overall variance of the sample-by-exon matrix, thus making

it unlikely that removing the first 12–15 components of the SVD decomposition results in loss of signal for these rare events. The SVD method depends on concurrently analyzing many samples, so that systematic noise becomes evident and can subsequently be removed. For the eight HapMap samples, we included an additional 533 ESP samples and removed 12 components. For analysis of the ASD trios, we combined the 122 trios (366 samples) with 366 randomly selected samples from the ESP dataset and removed 15 components. In our comparison of mrsFAST and BWA mappings, we used 492 ESP samples (for which BWA mappings were available) and the eight HapMap samples. Overall variance was lower in the BWA-based mappings, thus only six components needed to be removed during the SVD normalization.

Rare CNV discovery

To discover rare CNVs, we initially restricted our analysis to events where there was a change in copy number state for three or more consecutive exons. In order to assess the precision of our method, we intersected our exome-based deletion and duplication calls from five of the HapMap control individual genomes that had been previously analyzed using high-resolution array-CGH (2010). Of the 32 detected events (Table S1), seven were rare CNVs and 25 were CNPs; after intersecting with the reference set and requiring 10% reciprocal overlap, our method yielded 6/7 (86%) precision for rare CNVs and 16/25 (64%) for CNPs (Table 2.2). We also estimated sensitivity from this comparison: starting with 486 high-resolution array-CGH calls from the five HapMap samples that overlapped at least three exome probes, we restricted calls to those in unique/diploid regions of the genome (i.e., outside of segmental duplications, duplicated genes, and regions of somatic variation such as the *HLA* locus; Fig. S4). In this set of 41 calls (Table S2, lines annotated “Rare” and “CNP”), our algorithm identified 5/5 rare CNVs, but only 3/36 (8%) of CNPs (example calls in Fig. S7).

	Rare CNVs	Common CNPs	Total
$\geq 10\%$			
Reciprocal	6/7 (86%)	16/25 (64%)	22/32 (69%)
Overlap			
Any Overlap	7/7 (100%)	19/25 (76%)	26/32 (81%)
No Overlap	---	6/25 (24%)	6/32 (19%)

Table 2.2: Precision of exome-based CNV calls in HapMap samples

The relative paucity of rare CNVs in the HapMap cohort prompted us to estimate the precision of our method for rare CNVs using a larger set of 122 ASD trios. In our initial analysis, we applied the same filters for unique/diploid calls as above (Fig. S3), resulting in 191 calls among 97 probands. We identified eight putative *de novo* events (6.6% incidence; Table S3). For six of these, we were able to validate the event using available Illumina SNP microarray data as well as targeted array-CGH experimental data (Sanders et al. 2011). We could not confirm two *de novo* events, both of which were single-exon duplications of *FAF2* (data not shown). Next, we considered inherited events using our exome read-depth analysis and found that 128 events in the probands were inherited from either the mother or the father (Table S3a). For 117/128 (91.4%) of these events, the SVD-RPKM values of both the proband and the parent exceeded the detection threshold (± 1.5). However, for 11/128 (8.6%) of these calls, the SVD-RPKM values between proband and parents were just below the deletion or duplication threshold required for calling, and the inheritance status was determined by manual inspection.

Finally, inspection of the SVD-RPKM values for remaining 55 events (14 loci; Table S3b) revealed that these events strongly resemble CNP sites or mapped to genes for which processed pseudogenes exist. An example of such a locus can be seen at the *DAZZL* gene (Fig. S8: sample 13517.p1; chr3:15,636,820-17,640,105). The lack of phased copy number information precludes these loci from inheritance-based precision analysis (as the independent assortment of haplotypes can alter copy number in the offspring) and we, thus, excluded them from the precision analysis of rare CNV events. In total, we found

128 rare inherited events and validated 6/8 rare *de novo* events (excluding processed pseudogenes and duplicated exons), leading to a precision for the discovery of rare CNVs in the autism cohort of 134/136 (98.5%; Table 2.3). Example calls from this experiment are shown in Figure S8.

	Total	Validated
<i>de novo</i> CNVs	8	6 validated by SNP microarray
Inherited CNVs	127	116 have call made in parents (exceeded threshold) 11 manually inspected SVD-ZRPKM values revealed inheritance
Overall Precision		133/135 (98.5%)

Table 2.3: *Precision of exome-based CNV calls in autism trios*

We also gauged the accuracy of the exome-based CNV discovery against previously generated Illumina SNP microarray experiments (Sanders et al. 2011). SNP microarray data was available for 108 of 128 predicted inherited events and all eight predicted *de novo* events. We found 70% (76/108) of the inherited and 75% (6/8) of the *de novo* CNV events were confirmed by the SNP microarray (Table S3a). Given the high concordance rate of exome-based events within trios (>98%), the lower overlap vis-à-vis SNP microarray experiments likely reflects platform-specific differences in resolution and sensitivity and not an increased false-positive rate in the exome data.

Genotyping copy number polymorphic variants

We took two approaches in assessing our method's ability to determine the copy number of CNPs: 1) a relative correlation approach between the continuous SVD-ZRPKM values and whole-genome-sequence derived copy number estimates, and 2) an unsupervised clustering approach of exome-based genotype values in order to derive absolute copy number states for CNP loci.

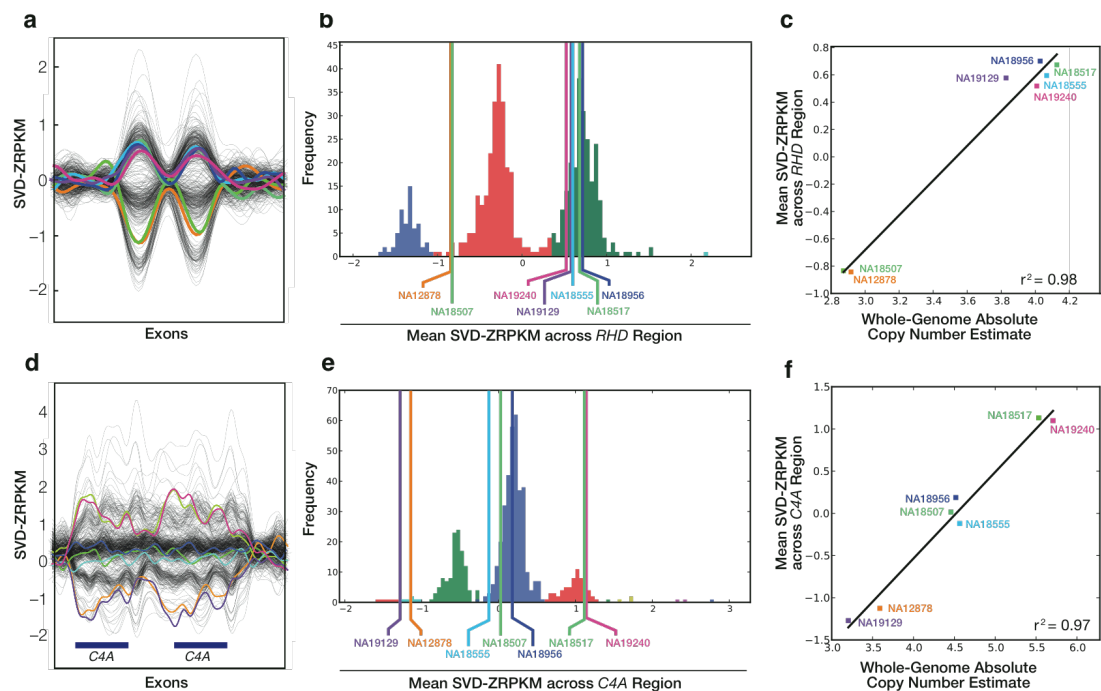


Figure 2.2: CNP locus genotyping of RHD and C4A (a) SVD-transformed values for exons for the Rhesus deletion factor locus (RHD/RHCE) show distinct copy number states across both paralogous genes. (b) Histogram of average SVD-ZRPKM values for ESP dataset (533 individuals) and seven HapMap samples. Clustering was performed using an unsupervised algorithm (Supplementary Note). (c) Correlation between SVD-ZRPKM genotype values (y-axis) and absolute copy number estimate (x-axis) based on whole-genome read-depth for seven HapMap samples and experimentally validated by array-CGH. (d-f) Similar to above, for C4A locus.

For the first approach, we selected 62 previously identified CNP loci and genes (from (Sudmant et al. 2010); Table S4) and calculated the copy number of each locus based on whole-genome read-depth data using previously described methodology, which has been experimentally validated using single-channel array-CGH intensity data⁹. For each locus, we correlated the estimated whole-genome copy number with the average of SVD-ZRPKM values for the exons in the locus (Fig. 2.2). The median r^2 value between exome-based and whole-genome-based genotyping at each locus was 0.91 (Fig. 2.3a; Table S4), indicating a high degree of reliability between exome and whole-genome copy number estimation for CNP loci. Furthermore, after stratifying the results by the median copy number of each locus, we found that for loci with median copy number of eight or less, 32 of 39 loci (82%) were highly correlated (r^2 value ≥ 0.9), but for loci with median copy

number greater than eight, the median locus r^2 was only 0.32.

Secondly, we assessed the accuracy of our approach in determining the absolute copy number of common CNPs. We leveraged available genotype information for seven of the HapMap samples in this study across 43 autosomal CNP loci previously studied by Campbell and colleagues (Table S5; Campbell et al. 2011). For each locus, we again used the locus-average of SVD-ZRPKM values and clustered these genotype values using an unsupervised clustering algorithm (Supplementary Note). Each cluster was then assigned the most likely copy number based on the most common copy number state previously identified. Using this unsupervised method, we correctly predicted absolute copy for 235/301 (78%) calls (Table S5) with an overall absolute genotype correlation across all 43 CNP loci of $r^2 = 0.74$.

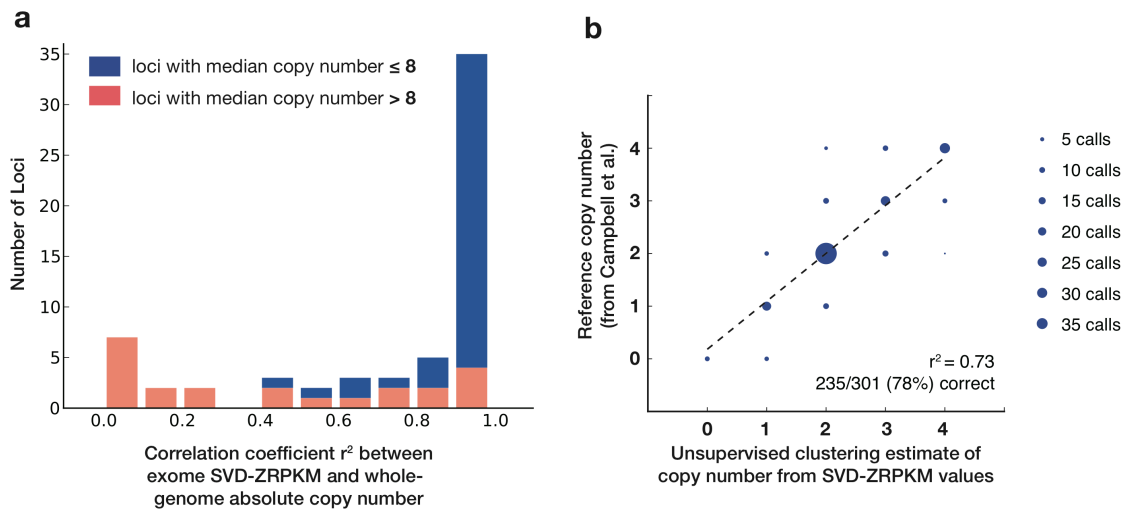


Figure 2.3: Genotyping accuracy across 62 CNP loci (a) Distribution of correlation coefficients of SVD-ZRPKM to whole-genome copy number estimate across 62 CNP loci for seven HapMap samples, split by the median copy number of each locus. (b) Correlation between SVD-ZRPKM score and relative (by median and standard deviation) whole-genome copy number estimate for 39 loci with ≤ 8 copies; and (c) for 23 loci with > 8 copies. Whole-genome read-depth copy number estimates for these specific sites and genomes were orthogonally validated using single-channel intensity data from previous array-CGH experiments.

Our algorithm uses relative read-depth values (introduced both by the Z-transformation and the SVD algorithm itself) in order to overcome significant batch biases in exomes, thus sacrificing the genome-wide linear model of read-depth and copy number exploited by whole-genome structural variation discovery algorithms. Nonetheless, the two approaches presented above can be used to “anchor” the relative SVD-ZRPKM values to absolute copy number. First, the strong r^2 correlation for many loci can be exploited as a “standard curve” for each locus, and the absolute copy number for exome samples can be estimated. Alternatively, SVD-ZRPKM values can be clustered (Supplemental Note) into copy number groups, thus facilitating absolute copy number estimates without the use of whole-genome data.

Comparison with other methods

As we generated our read-depth estimates from mrsFAST-based alignments, we were interested to see how our method would perform using BWA-based alignments. The BWA alignments were generated using commonly used parameters and filtering steps suitable for SNP-centric analyses, including removal of reads with multiple mappings (Supplementary Note). We calculated RPKM values from these BWA alignments for the HapMap samples and a subset of the ESP exomes. We observed that signal for rare deletions and duplications in the HapMap samples were attenuated (Fig. S5), and that the median signal-to-noise ratio for the seven rare deletions and duplications was 58% lower for the BWA-based mappings (Table S6; Supplementary Note). In addition, we genotyped 47/62 loci in Table S4 and found a striking difference in the correlation between BWA-based mappings (median $r^2 = 0.36$) and mrsFAST-based mappings (median $r^2 = 0.92$). The remaining 15/62 loci did not have any probes with adequate BWA read-depth, making them intractable and false negatives by this approach. The difference in correlation with mrsFAST mappings was mostly notable for loci with copy numbers ranging between 7 and 12 (Fig. S6b). These data highlight the importance of considering reads with multiple mappings, especially for loci with increased copy number (e.g., the *LRRC37A3* locus; Fig. S6c). These differences, however, do not solely reflect differences between the alignment algorithms, but rather of the entire alignment and post-processing pipeline.

Finally, we compared our algorithm to ExomeCNV (Sathirapongsasuti et al. 2011), which is designed to detect copy number aberration in the context of cancer using closely matched tumor-normal pairs of exomes. Nevertheless, we were interested to see if ExomeCNV could be used to detect germline variation. We analyzed (using default settings; see Supplementary Note) four HapMap exomes with NA19240 as the reference and compared the results to a validated call set from these genomes (2010). Overall, ExomeCNV predicted 450 CNVs, of which only 63 (14%) had more than 10% reciprocal overlap with the validated call set. In contrast, our algorithm identified 24 calls among these four samples, of which 21 (88%) overlapped the validated call set. We note that ExomeCNV uses uncalibrated read-depth to estimate copy number, and, depending upon batch effects, this can result in the algorithm reporting a significant fraction of the exome as non-diploid (Fig. S9). Furthermore, similar to the BWA-based alignments (see above), ExomeCNV has limited dynamic range in CNP loci and duplicated genes: the average r^2 correlation across tested CNPs was 0.57 (compared to our algorithm, $r^2 = 0.92$; Fig. S10).

2.5 Discussion

We have outlined a method for making read-depth data from exomes amenable to rare CNV discovery, as well as copy number genotyping of CNP loci. We used SVD normalization to overcome a host of coverage biases introduced by the capture and sequencing of exomes. Our method allows for differing sample preparations and capture reactions to be integrated into the same experiment, provided each “batch” is sufficiently large ($n \geq 8$). This includes correct normalization of the X chromosome, such that deletions and duplications can be assayed regardless of the sample’s sex. Additionally, our method can integrate exomes captured with different exome capture target designs: the eight HapMap exomes were captured using the Roche NimbleGen SeqCap EZ Version 1, while all other exomes in our experiments were captured using the SeqCap EZ Version 2 capture kit.

Remarkably, we find that sufficient dynamic range response remains to accurately predict the copy of duplicated genes up to eight copies. The upper limit of this response is likely an effect of the stoichiometry of the exome-capture reaction and we suggest that this may

be improved simply by adjusting the concentrations and targets of exome-capture platforms. Another important consideration in interpreting exome read-depth data is the presence of polymorphic processed pseudogenes. In our study of autism trios, we found that 14% (26/191) of events correspond to changes in the copy of processed pseudogenes residing elsewhere in the genome, often in segmental duplications. Such events have been difficult or impossible to discover using traditional SNP microarray approaches, as the probes for these assays often do not explicitly target the coding exons themselves. While such events may be easily inferred based on the absence of intronic sequence, a comprehensive catalog of polymorphic processed pseudogenes will improve detection of *bona fide* exonic deletions and duplications.

We envision a number of algorithmic improvements. Although using mrsFAST mappings both increases the signal-to-noise ratio for rare CNVs and improves genotyping accuracy for CNPs, these mappings often cannot distinguish between paralogous genes. By restricting the RPKM calculation to exons and regions that contain paralog-specific single nucleotide variants (Sudmant et al. 2010), we hope to be able to extend our method to genotype duplicated genes in a paralog-specific manner. We also expect to lower the minimum number of exons required to detect a CNV. We applied our method to genotyping single exons (such as the third exon of *GHR*; Santos et al. 2004) and found the SVD-ZRPKM values robustly distinguished different copy number classes. By developing a discovery set of copy number polymorphic exons, genes, and loci—as well as their copy number states in populations—future disease-association studies will be better informed. Finally, though array-based technologies have described many CNP-disease associations (Girirajan et al. 2010), discovery of loci has been limited to those with low median copy number, and our approach here will be able to examine CNP loci with higher copy number. Using our approach with large clinical cohorts currently undergoing exome sequencing, we expect to find new disease associations with rare CNVs, CNP loci, and paralog-specific copy number of known CNP loci.

ACKNOWLEDGEMENTS

We thank S. Ng, S. McGee, and T. Brown for helpful comments in the preparation of this manuscript, M. State and the Simons Simplex Collection Genetics Consortium for providing Illumina genotyping data, and A. Schachtel for suggesting the CoNIFER name. This work was supported by NIH grants HD065285 (E.E.E.), HHSN273200800010C (D.A.N.), and HL102926 (D.A.N.) and the Simons Foundation Autism Research Initiative (E.E.E.). E.E.E. is an Investigator of the Howard Hughes Medical Institute.

III. Transmission disequilibrium of small CNVs in simplex autism

3.1 Summary

We searched for disruptive, genic rare CNVs among 411 families with sporadic autism spectrum disorder (ASD) from the Simons Simplex Collection using available exome sequence and CoNIFER (copy number inference from exome reads). Our approach yielded increased sensitivity for smaller genic rare CNVs compared to high-density SNP microarrays (~2X higher yield), especially for CNVs smaller than 20 kbp. We find that affected probands inherit more CNVs than their siblings (453 vs. 394, $p=0.004$; $OR=1.19$), and these affect more genes (921 vs. 726, $p=0.02$; $OR=1.30$). These smaller CNVs (median size 18 kbp) are transmitted preferentially from the mother (136 maternal vs 100 paternal, $p = 0.02$) although this bias occurs irrespective of affected status. The excess burden of inherited CNVs among probands is driven primarily by sib-pairs with discordant social behavior phenotypes ($p < 0.0002$, measured by SRS score), in contrast to families where the phenotypes are more closely matched or less extreme ($p > 0.5$). Finally, we found strong enrichment for brain-expressed genes unique to probands, especially in the discordant SRS group ($p = 0.0035$). In a combined risk model, our set of inherited CNVs, *de novo* CNVs and *de novo* SNVs all independently contributed to the risk of autism ($p < 0.05$). Taken together, these results suggest that small transmitted rare CNVs play a role in the etiology of simplex autism. Importantly, the small size of these variants aids in the identification of specific genes as additional risk factors associated with ASD.

This chapter has been published: Krumm, N., O’Roak, B. J., Karakoc, E., Mohajeri, K., Nelson, B., Vives, L., et al. (2013). Transmission Disequilibrium of Small CNVs in Simplex Autism. *American Journal of Human Genetics*, 93(4), 595–606. doi:10.1016/j.ajhg.2013.07.024

3.2 Introduction

Discovering the mutations and the genes responsible for autism spectrum disorder (ASD) requires an assessment of the full-spectrum of genetic variation within families including both *de novo* and inherited events. There is compelling evidence that a diverse range of *de novo* mutations play an important role, including copy number variants (CNVs; Levy et al. 2011; Sanders et al. 2011; Sebat et al. 2007; Glessner et al. 2009; Pinto et al. 2010), single nucleotide variants (SNVs) and insertions and deletions (indels) (Iossifov et al. 2012; O'Roak et al. 2012b; Sanders et al. 2012; O'Roak et al. 2011). However, taken together, *de novo* variation does not fully explain the genetic etiology of ASD: only ~8% of probands carry a *de novo* CNV and ~10-20% carry a pathogenic *de novo* SNV or indel. Many of these mutations likely play a pathogenic role in the development of ASD, especially in the context of sporadic (or “simplex”) ASD. However, the heritability of ASD is estimated to be between 50% and 90% (Bailey et al. 1995; Hallmayer et al. 2011)—much higher than the to-date explained fraction of disease—suggesting that additional genetic factors contribute to the etiology of ASD.

The prevalence of rare CNVs smaller than 50 kbp has been underestimated in previous surveys using oligonucleotide microarrays (Levy et al. 2011; Sanders et al. 2011) and their role in ASD has yet to be explored. Such pathogenic events could in principle provide as much specificity as exonic *de novo* mutations with respect to genes and informative protein networks. Several recent methods based on exome sequencing read-depth data have enabled the discovery of small genic CNVs previously missed by microarray (Krumm et al. 2012; Fromer et al. 2012). In this study, we tested the hypothesis that small genic *inherited* CNVs also contribute to the genetic etiology of sporadic autism. Several lines of evidence are potentially supportive of this hypothesis, including increased prevalence of the broader autism phenotype (BAP) in parents of affected children (Losh et al. 2008; Davidson et al. 2012), trends for higher burden of extremely rare singly-transmitted CNVs in simplex families (Levy et al. 2011), and enrichment for large CNVs in cases versus unrelated controls (Pinto et al. 2010). In contrast other previous studies which have examined inherited CNVs in ASD found no significant excess of inherited burden in probands with ASD, although these studies were

mainly designed to detect *de novo* CNVs (Sanders et al. 2011).

Here, we present evidence for transmission distortion for smaller CNVs (median size ~18 kbp) by investigating families where both affected and unaffected siblings have been exome-sequenced. The availability of whole-exome sequence data for our samples has the advantage of increased sensitivity for small, genic CNVs affecting two or more exons, as well allowing us to integrate both rare SNV and CNV to develop a model to explain the genetic architecture of ASD.

3.3 Methods

CNV Detection from exome sequence data

We analyzed exome sequencing data from families ascertained as part of the Simons Simplex Collection (Fischbach and Lord 2010). Underlying FASTQ sequence data was obtained from 391 published ASD quads (O'Roak et al. 2012b; Sanders et al. 2012; Iossifov et al. 2012) and we generated sequence data for an additional 19 unaffected siblings from published trios (O'Roak et al. 2011). The data set include sequence data (median coverage >50x) from 411 families where a proband, unaffected sibling, mother and father (termed quad) all had been sequenced for a total of 1644 samples (see Table S1 and S2 for details). Sequence reads were split into 36mers, and mapped using the mrsFAST alignment program (Hach et al. 2010) to the Nimblegen EZ-SeqCap v2 targets (including 300 bp around each target and allowing two mismatches per 36mer). We used CoNIFER (Krumm et al. 2012) to calculate exon-level coverage and removed systematic bias between samples and targets. Using a custom pipeline (Figures S1 and Supplemental Methods), we 1) segmented our CoNIFER SVD-ZRPKM values using the DNACopy algorithm (Venkatraman and Olshen 2007), 2) minimized false-negatives by a quad-based genotyping method, 3) clustered CNVs into overlapping CNVRs, and 4) removed CNVs found in duplicated or repetitive genomic space. We limited our final call set to inherited CNVs (i.e., transmitted CNVs) that were present in 10 or fewer families (or approximately 1% population frequency), and we excluded CNVs which primarily fell within duplicated or highly polymorphic regions of the genome. We considered a CNV “rare” if it occurred in 10 or fewer families and a CNV private if it was observed only in

one family. Lastly, we did not include CNVs on the X chromosome in any analysis, and all *de novo* CNVs were excluded from burden analyses except where noted. Throughout this paper, we define “CNV burden” as the number of rare CNVs per individual.

Array comparative genomic hybridization

We designed a customized CGH microarray (Agilent SurePrint G3 4x180k CGH microarray; probe density ranging from 125 bp⁻¹ to 5 kbp⁻¹ depending on the size of the event to be validated) and selected 161 CNVs from a subset of 80 samples, stratified by proband/sibling (36 probands and 44 siblings), and by dataset (26 from Iossifov et al., 22 from O’Roak et al., 32 from Sanders et al.; Table S1 and S5). Minimum deletion and duplication thresholds for validation were determined by ROC curve analysis of known positive and negative control CNVs (Figure S3).

Phenotypic measures and models

Social Responsiveness Scale (SRS) was used as a quantitative measure of social deficits (Constantino and Gruber). We had complete phenotype information (SRS for both proband and sibling, and full-scale IQ for proband) for 389 families in this study (Table S7) based on data from the SSC. The probands in this study had a median SRS t-score of 82, significantly higher (i.e., more severely affected) than the median SRS score of our unaffected siblings (45; $p < 0.00001$, two-tailed paired t-test). We defined mild, moderate and severely affected individuals based on published thresholds (Constantino and Gruber).

Expression analysis

Gene expression data was from the Human U133A/GNF1H Gene Atlas (GEO: GSE1133), comprising 79 human tissues, including 18 nervous system tissues (Su 2004). Expression values were averaged across multiple probes when available. We defined a gene to be expressed in a given tissue if it ranked in the top 5% of all genes for that tissue. To measure enrichment, we compared the fraction of genes unique to either siblings or probands expressed in each tissue, and empirical p-values were calculated by shuffling proband/sibling labels 20,000 times and recomputing tissue-level expression enrichment. We FDR-corrected for 79 tests (i.e., for tissues) and statistical significance was assessed at $q < 0.05$.

Combined mutation model

We generated a list of truncating *de novo* SNV mutations (nonsense, frameshift or splice mutations) discovered in our 411 quads from published lists (Iossifov et al. 2012; O'Roak et al. 2012b; 2011; Sanders et al. 2012). Both *de novo* and inherited CNV burden was derived from this work (Table S3 and S4). We used a logistic regression model, which transforms the binary outcome (i.e., affected vs. unaffected) such that linear predictors can be used. The model shown in Figure 3.5 is summarized as: $\text{logit}[P(\text{Affected}=1)] \sim \text{intercept} + (\text{de novo CNV burden}) * (\text{inherited CNV burden}) * (\text{de novo SNV burden})$.

Data availability

The CoNIFER output files for 1,644 samples are deposited in the National Database of Autism Research (NDAR) under the NDAR Collection ID 1878 and the title of this manuscript.

3.4 Results

Samples and CNV discovery

We discovered a total of 847 transmitted, exonic, rare, autosomal CNVs (Table 3.1). This included 453 CNVs transmitted to probands and 394 transmitted to unaffected siblings. Overall, the median estimated CNV size was 18.1 kbp (range 150 bp – 5.18 Mbp, or 2 – 320 exons). The median size of inherited CNVs was slightly larger in probands (19.4 kbp) when compared to unaffected siblings (16.6 kbp) but this difference was not statistically significant. As expected, duplications outnumbered deletions (519 vs 328; $p < 1*10^{-10}$, binomial two-tailed test) and duplications were significantly larger than deletions (two-sided Mann-Whitney-U test, $p < 1*10^{-16}$). The excess of duplications depended upon the size of event. For example, rare CNVs involving 20 or more exons were overwhelmingly duplications (139 duplications vs. 25 deletions), while small events were not significantly different (73 duplications and 93 deletions for 2-exon CNVs). This difference is observed irrespective of disease status (Figure S4).

Category	CNVs	Dups	Dels	Median Size (est)	% of Samples	CNVs >500 kbp	CNVRs
All Proband CNVs	453	277	176	19.4 kbp	64%	21	390
All Sibling CNVs	394	242	152	16.6 kbp	60%	16	345
Father → Both	199	130	69	16.7 kbp	41%	7	94
Father → Proband Only	100	67	33	25.0 kbp	19%	7	93
Father → Sibling Only	82	52	30	15.4 kbp	18%	2	80
Mother → Both	233	127	106	15.0 kbp	48%	10	118
Mother → Proband Only	136	82	54	24.9 kbp	26%	5	128
Mother → Sibling Only	97	61	36	21.7 kbp	21%	6	94
Either Parent → Proband Only	236	149	87	25.0 kbp	39%	12	211
Either Parent → Sibling Only	179	113	66	19.3 kbp	36%	8	168
Mother → Either Offspring	466	270	196	17.8 kbp	86%	21	313
Father → Either Offspring	381	249	132	18.6 kbp	72%	16	252
Totals	847	519	328	18.1 kbp	62%	37	525

Table 3.1: Summary of transmitted CNVs in 411 ASD quads

Validation using single nucleotide polymorphism SNP microarray and targeted array-CGH

We assessed the specificity of our call set by comparing our larger calls to Illumina 1M/Duo SNP microarray data and then selecting a subset of 80 samples for validation of smaller CNVs by array comparative genomic hybridization validation. These 80 samples carried a total of 161 exome-based CNV calls of which 69 (43%) were confirmed by SNP microarray (Figure 3.1a). Using a customized microarray design (Methods), we were able to test 86/92 of the remaining calls and confirmed an additional 65 events (nearly a twofold increased yield of CNVs) (Table S5). Of the 27 events which were not validated by array-CGH, 14 (or 9% of all 161 calls) were found to be specifically part of processed pseudogenes (i.e., retro-transcribed mRNA), which masquerade as duplications in exome-based discovery of CNVs, indicating that these events— while not genomic CNVs— are in fact true duplications of these genes or exons. Thus, we estimate an overall false positive rate (FPR) of 4%–8% (7/155 tested, or 13/161 in total; Figure 3.1a), dependent on the number of probes (or exons) in each call: for calls with fewer than 10 exons, the false positive rate was ~7% (6/104), while only one calls with 10 or more exons did not validate (2% or 1/51). There was no difference in the FPR between probands and siblings (3/68 [4.2%] for probands, 4/80 [4.5%] for siblings; Table S6).

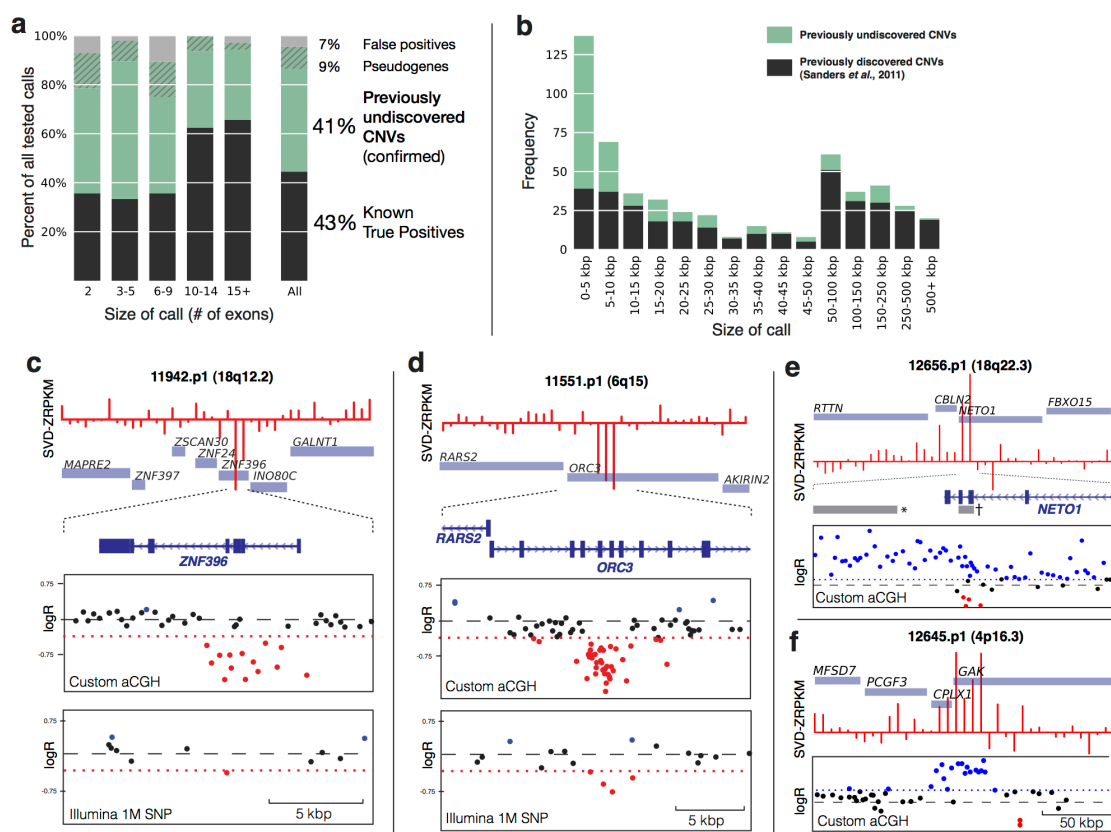


Figure 3.1: Discovery and validation of previously undiscovered CNVs using exomes. (a) Fraction of CNVs previously identified using Illumina 1M SNP microarray (gray, “known true positives”), the fraction of previously undiscovered CNVs identified and confirmed by targeted array-CGH in this study (green, “previously undiscovered CNVs”), confirmed processed pseudogenes (hatched green) and the overall false positive rate for unconfirmed CNVs (gray). (b) The majority (73%, 152/207) of all previously undiscovered calls (green) discovered using exomes were smaller than 20 kbp (c-d,f) Three examples of previously undiscovered CNVs in this study. Top: CoNIFER output and normalized coverage at each exon. Middle: targeted array-CGH at CNV locus, with threshold for deletion/duplication (dotted red line) as determined by ROC-curve analysis of known CNVs (Supplemental methods). Bottom: Illumina 1M SNP microarray data for locus, showing poor probe coverage (c and d only). (e) Exome-based CNV discovery affords high exon-level specificity, as indicated by duplication of NETO1 exons (†, CoNIFER call). Previous work (Sanders et al., 2011) had discovered this CNV (*), but the (incorrect) breakpoints did not extend into NETO1.

We also assessed the sensitivity (or false negative rate, FNR) of our calls versus the previously identified CNVs from SNP microarray data. We found that our pipeline identified 72% (FNR of 0.28) of all known CNVs intersecting at least two exons and

supported by 10 SNP microarray probes. False negative CNVs corresponded to samples with reduced mapped sequence coverage (Figure S2). For example, the Iossifov dataset (Iossifov et al. 2012) had an approximately twofold higher FNR, likely due to the lower overall sequence coverage in these exomes (a known factor in exome-based CNV discovery; Krumm et al. 2012; Fromer et al. 2012), and that the FNR for calls affecting only two exons was significantly higher than those with 3 or more exons (Table S6). We found no differences in the mapped coverage, estimated false positive rates or false negative rate among siblings and probands ($p > 0.3$, Fisher's two sided exact test and Table S6).

Increased inherited CNV burden among autism probands.

We compared the burden of inherited CNVs in the 411 probands and their siblings in terms of the total number of CNVs and the total number of genes "hit". We find that probands inherit more CNVs than siblings (453 vs. 394; Figure 3.2a) and these harbor more genes (921 vs. 726; Figure 3.2b). These comparisons are significant when using a paired t-test of probands-sibling pairs ($p = 0.02$ for genes and $p = 0.004$ for CNVs, two-tailed paired t-test) and when comparing the summed values for probands and siblings in aggregate ($p < 1 \times 10^{-6}$ for genes and $p = 0.046$ for CNVs, binomial two-tailed test). In order to ensure that these results were not driven by a few outlier families, we bootstrapped our data and calculated the confidence intervals for the proband-to-sibling burden (Figure S5). For CNVs, we found a median burden increase of 1.19 (95% CI: 1.09 – 1.29) and for genes a burden increase of 1.30 (95% CI: 1.10 – 1.52) across 10,000 bootstrap replicates, thereby rejecting the null hypothesis that probands have no increased inherited CNV burden in comparison to their siblings (Figure S5). Proband CNV burden was elevated over siblings across all size ranges, although individual quintile bins did not independently achieve statistical significance, due to their smaller size (Figure 3.2c). We find no significant enrichment of burden in either the smallest or the largest CNVs (by chi-square $\chi^2 = 1.18$, $p = 0.95$, $df = 5$, suggesting that the burden was not exclusively the result of either small or large CNVs).

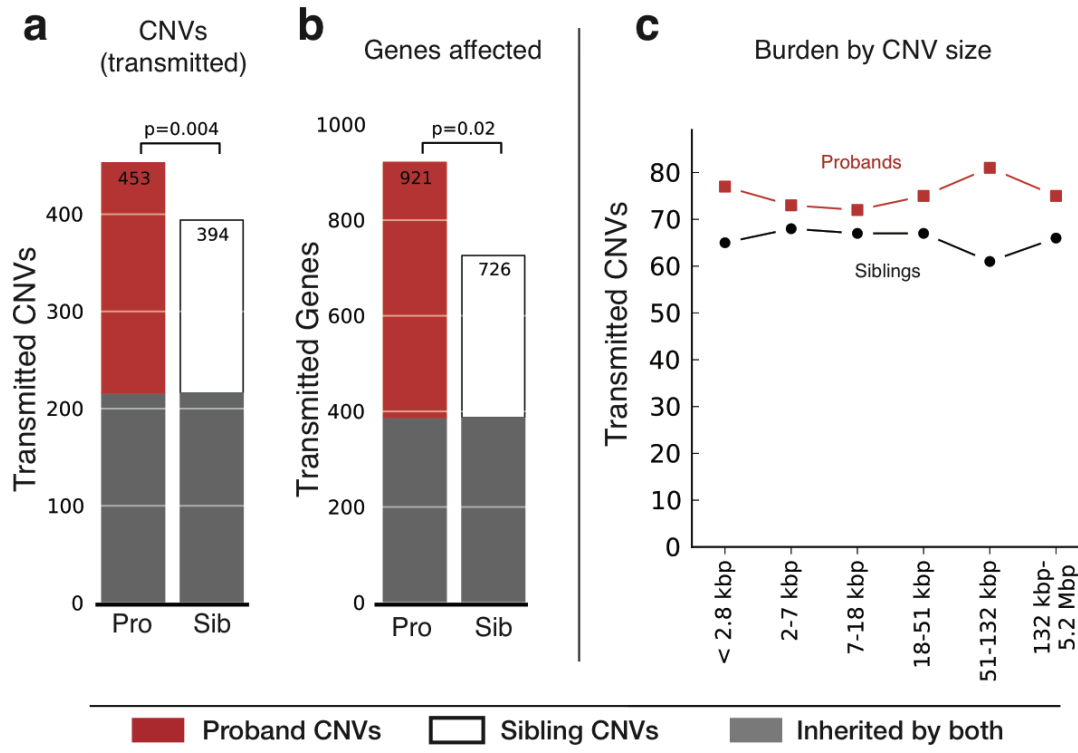


Figure 3.2: Increased inherited CNV burden in ASD probands for large and small CNVs. (a) Total number of rare (observed in fewer than 10 families) inherited CNVs (≥ 2 exons) for 411 ASD probands (Pro) and their unaffected siblings (Sib). (b) Total number of affected genes in rare inherited CNVs. P-values are two-tailed paired t-tests between proband and sibling counts. (c) Burden of inherited CNVs across six size categories.

Previous work has indicated that private or ultra-rare CNVs may be more likely to be pathogenic than simply “rare” (e.g., $< 1\%$ frequency) CNVs (Levy et al. 2011). We therefore examined if the inherited burden in probands was due to private CNVs in a small subset of the 411 families. We examined 271 private CNVs in probands and 245 private CNVs in siblings, but found no enrichment of private burden when compared to rare CNVs ($p = 0.74$, Fisher’s exact test; Figure S6a), nor did we find enrichment for the number of affected genes ($p=0.46$, fisher’s exact test; Figure S6b). (Note: the burden was in fact slightly increased when considering all rare events). We searched for additional factors which could underlie the proband-sibling burden differential. We found no significant differences in CNV burden dependent on the sex of the proband or the sibling, the concordance of their sexes, or the birth order of proband and sibling ($p > 0.5$, Fisher’s

exact test; Table S8). However, we note that the highest overall CNV burden was found in families with one affected proband and at least three unaffected siblings. In fact, there was a linear increase in burden between probands and siblings across increasing family size, culminating in a 1.38x higher burden of CNVs in probands with three or more unaffected siblings.

Finally, we analyzed our dataset for parent-of-origin effects, and found a greater number of maternally transmitted CNVs, (136 maternal vs 100 paternal, binomial two-tailed p-value = 0.02); but this effect was not significantly enriched in probands versus siblings (Fisher's exact test odds ratio = 1.14, two-tailed p = 0.49). Nonetheless, when we considered a null hypothesis in which a given transmitted CNV was equally likely to be transmitted to the proband only, the sibling only, or both (each with 1/3 probability), we found strong evidence that CNVs were not transmitted in equal fashion (Table 3.1; chi-square test with equal expected proportions, $p = 0.0058$, $\chi^2=16.4$, $df = 5$), and that CNVs transmitted from the mother to the proband only were significantly more common than other transmissions.

CNV burden-phenotype correlation.

We assessed whether the increased inherited CNV burden would segregate with markers of ASD phenotypic severity using phenotype data from the SSC. First, we utilized the Social Responsiveness Scale (SRS), a standardized parent- or teacher-completed questionnaire which measures the severity of autism symptoms in social settings (but is not a diagnostic indicator of ASD and was not used in ascertainment of the SSC). We partitioned our 411 families into two groups based on the SRS t-score: 1) We defined "Discordant SRS quads" as those where the proband was severely affected (SRS $t > 75$) and the sibling mildly affected (SRS $t < 60$), and 2) "Concordant SRS quads" as all others (Figure S7). The concordant group encompassed a range of moderately affected probands as well as some moderately affected siblings (Figure S7). There were a total of 276 discordant SRS proband-sib pairs and 115 concordant pairs based on this definition. We found a striking split between the discordant and concordant proband-sibling pairs: the increased CNV and gene burden was almost completely driven by the discordant pairs (Figure 3.3a; $p < 0.0002$ for CNVs, $p < 0.02$ for genes, two-tailed paired t-test), and there

was virtually no difference at a group or family level for concordant SRS pairs overall (1.04x, $p > 0.5$). Moreover, the burden ratio between probands and siblings was increased in the discordant group (for CNVs: 1.27x; for genes: 1.41x) over the ratio for the full set of 411 quads. Finally, we found that offspring (probands and siblings) with SRS scores ≥ 60 (“moderate” and “severe” range) had higher CNV burden than did all offspring with SRS score < 60 (361 CNVs in 390 mildly affected offspring [1.12] vs. 436 CNVs in 388 moderately/severely affected offspring [0.92]; two-tailed independent t-test $p < 0.0094$). There was no statistically significant difference in burden between probands and siblings within each group (i.e., SRS < 60 or ≥ 60), however the relatively low number of “affected” siblings and “unaffected” probands hampers these comparisons.

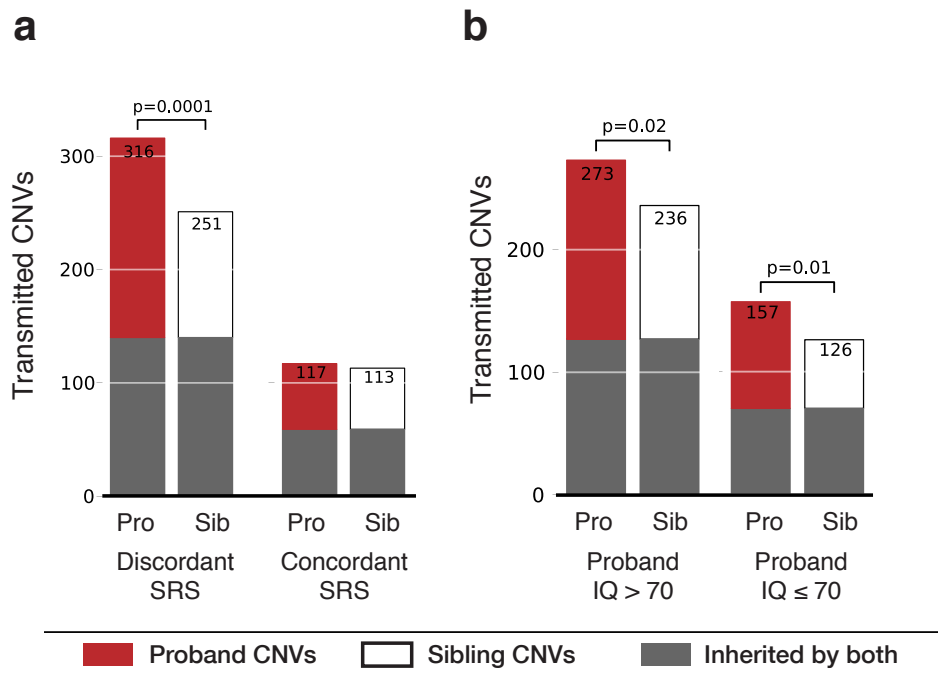


Figure 3.3: Inherited CNV burden correlates with SRS phenotype. The Social Responsiveness Scale measures autism features in social settings via parent report on 65 items. **(a)** We classified proband-sibling pairs with severely affected probands but mildly or unaffected siblings as “Discordant SRS” quads (276 quads), and all other quads as “Concordant SRS” quads (115 quads). Strikingly, the discordant SRS quads fully recapitulated the inherited CNV transmission bias, whereas the concordant SRS quads did not show a differential burden. **(b)** CNV burden was independent of full scale IQ (FSIQ), and probands with either low FSIQ (≤ 70) or high FSIQ had more CNVs than did their siblings. P values refer to two-tailed paired t-tests between probands and siblings.

	Proband FSIQ	Proband CNVs	Sibling CNVs	Ratio	Two-tailed t-test Probands vs. Sibs
All Quads	≤ 70	157	126	1.25	p = 0.014
	71 – 85	89	70	1.27	p = 0.029
	≥ 86	184	166	1.11	NS
Discordant SRS Quads Proband SRS < 60 Sibling SRS > 75	≤ 70	138	104	1.32	p = 0.004
	71 – 85	62	44	1.40	p = 0.012
	≥ 86	113	101	1.12	NS
Concordant SRS Quads	≤ 70	19	22	0.86	NS
	71 – 85	27	26	1.04	NS
	≥ 86	71	65	1.09	NS

Table 3.2: Summary of IQ and SRS burden. P-values represent two-tailed paired t-tests between probands and siblings in each group.

Next, we considered if the full-scale IQ (IQ) of the probands was affected by inherited CNVs. Since IQ scores were only available for probands (Table S7), we grouped quads into three groups: IQ ≤ 70 (“low”, consistent with a diagnosis of intellectual disability), between 71 and 85 (“intermediate”), and or greater than 85 (“high”). The CNV burden was significantly greater for probands in the “low” and “intermediate” proband IQ bins (1.25-1.27x burden, Table 3/2). Probands with “high” IQ did not show statistically significant enrichment over siblings, although a trend was still apparent (1.11x, Table 3.2). When we examined the effect of SRS and IQ together (Table 3.2, Table S8 and Figure S8), we found that the burden differential was strongest for the most severely affected probands (those with IQ ≤ 85 and part of discordant SRS quads), reaching 1.32–1.40x for CNVs (p = 0.004). However, there was *no* significant burden between probands and siblings in SRS concordant quads, even with “low” IQ probands (0.8x-1.09x, p>0.5; Table 3.2), indicating that the inherited burden may be most closely aligned with SRS score, and not IQ (however, we caution that there were only 22 quads total in this group).

Enrichment for brain-expressed genes in inherited CNVs:

We observed a trend for more of the proband-only genes to be highly expressed in brain-related tissues (19/317 or 6% proband vs. 6/224 or 2.7% for sibling only; Table S8; see Methods). The effect becomes most pronounced when considering discordant SRS quads

(15/256 genes (5.9%) in probands, and 2/170 (1.2%) in siblings ($p = 0.007$) (Figure 3.4). When we considered all genes highly expressed in at least one brain-related tissue, we found significantly more 57/411 (13.9%) of probands had a CNV than did their siblings (33/411, or 8.0%; OR = 1.85, $p = 0.009$, Fisher's exact test). These results suggest that a fraction of proband-specific genes are expressed in the nervous system tissues, and that this fraction is higher in proband-only genes than in sibling-only genes. While we caution that expression does not definitively imply pathogenicity, many of these genes and their biological pathways may be of interest for further study, both in these particular individuals and for ASD genetics in general.

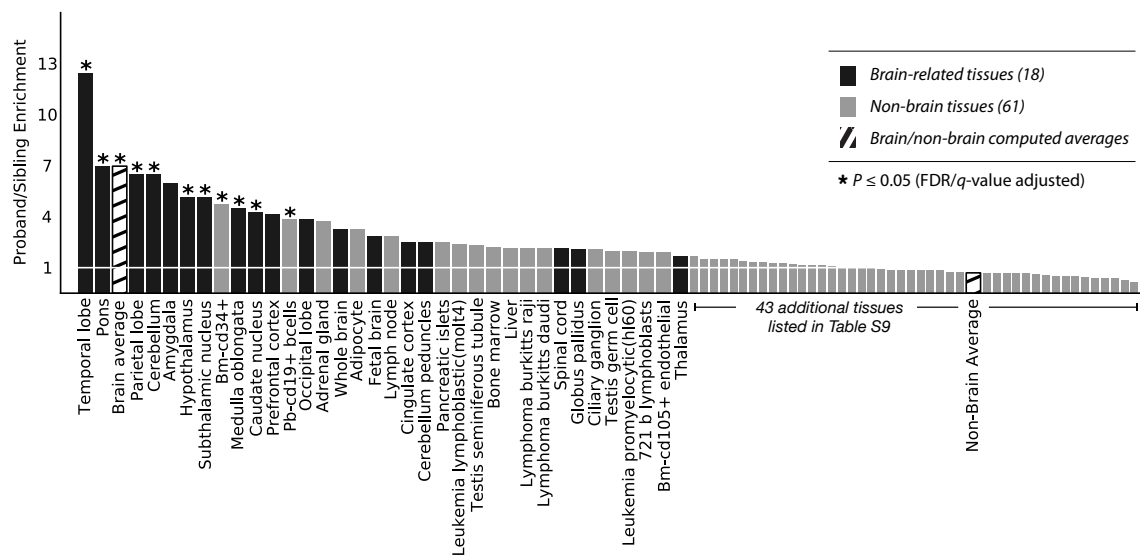


Figure 3.4: Genes in proband-only CNVs from SRS-discordant quads are more likely brain-expressed. We defined a gene to be expressed in a tissue if it ranked in the top 5% of all genes in that tissue, and calculated the fold enrichment of proband and sibling genes expressed in each tissue. Tissues part of brain structures had the strongest proband enrichment (black bars), as did a computed average of expression across 18 brain regions (“Brain average”) in comparison to the average expression across other regions. However, the particular brain tissues with the strongest apparent enrichment should not be considered as independently enriched, as expression values for individual genes between brain regions are highly correlated. Stars indicate a FDR-corrected p -value < 0.05 . See Figure S9 for results from all 411 quads.

We compared the genes detected in the CNVs in this study to a set of 1,560 genes that have been previously observed in autism/ASD, intellectual disability or schizophrenia (Table S10; Figure S10). Among SRS-discordant quads, we found significant enrichment

of autism genes among proband CNVs compared to unaffected siblings (66 vs 35 genes, $p=0.006$ two tailed paired t-test; Table 3.2a and Table S8) In contrast, there was no enrichment among “concordant SRS” proband-sibling pairs for previously observed genes (In fact, siblings had more genes: 17 vs. 24, $p = 0.069$). Overall, 16% of probands (44/276) in the SRS discordant group had a CNV in a previously observed gene, while only 10% of probands (12/115) in the concordant group had such an event. Intersecting the brain-expressed genes and previously observed disease genes we found that 13 genes matched both criteria, corresponding to 1.7% of all proband genes, and only two genes (0.3%) in siblings (Figure S10). The 13 convergent proband-only genes were found exclusively in discordant SRS families, indicating that these genes may be associated with more severe phenotypes (Table S11).

3.5 Discussion

In this study, we report a significant CNV transmission bias for autism, finding an enrichment of inherited CNVs in sporadic cases versus their unaffected siblings. The targeted nature of exome sequencing enabled us to explore a smaller CNV landscape largely inaccessible by high-density SNP microarray data (Pinto et al. 2010; Sanders et al. 2011; Levy et al. 2011). We estimate that the use of exome data increased our power to detect gene-disruptive CNVs smaller than 20 kbp by ~ 2.25 -fold. These CNVs provide potential insight in the pathophysiology of inherited CNVs in sporadic autism. We find that the CNV burden is more strongly correlated with measures of ASD phenotypes (such as the SRS score) as opposed to IQ; for proband-sibling pairs with concordant SRS scores, IQ was not dependent on the probands CNV burden. Genes already associated with autism and/or highly expressed in the brain are more likely to be disrupted. Private CNVs (seen once) were no more likely to be found in probands than simply rare variants (seen fewer than 10 times in our families). Burden was consistent across all sizes of CNVs, and we did not find any enrichment for either small or large events. Mothers are significantly more likely to be carriers of transmitted CNVs than fathers irrespective of disease status of the child. This finding is consistent with our recent analysis of “secondary” CNVs being transmitted from mothers to children with developmental delay and multiple CNVs (Girirajan et al. 2012). We also noted that the transmission bias

becomes more significant in probands from ASD families with many siblings as opposed to fewer individuals. Although this observation is inconsistent with the assumption that probands in larger families with many unaffected siblings are more likely to have an underlying sporadic genetic etiology, it likely reflects an ascertainment in selecting the “least affected” sibling in a large family as the “designated sibling” for the purposes of forming a quad (Sanders et al. 2012). This suggests that a significant fraction of the underlying genetic etiology in the SSC may be inherited, a notion that has been examined previously (Davidson et al. 2012).

Our study benefited from the quad-based design of the SSC (Fischbach and Lord 2010), which provided a robust genetic control for each ASD proband, as well as the detailed phenotypic information available, which sharpened the contrasts between severely affected and less affected probands and their siblings, some of which showed subtle signs of the Broader Autism Phenotype (BAP; Davidson et al. 2012). Most of our observations were strengthened or restricted to “SRS discordant” quads, where the proband was severely affected in terms of the SRS scale, but the sibling was unaffected. Approximately 67% (276/411) of the quads in this study were categorized as SRS discordant, and these quads explained virtually the entire overall CNV burden, encompassed the majority of brain-expressed genes and strengthened the association with previously implicated disease genes. This effect may be driven by inherent ambiguity in the simplex and multiplex classification scheme-- a scheme that is not truly binary but rather a continuous probability based on the number of unaffected siblings in the family (the more unaffected siblings, the greater the likelihood that the family is simplex). In essence, by focusing on the SRS discordant quads only, we have enriched not only for a more severe proband phenotype, but also a truly “*simplex*” genetic etiology (as opposed to an environmental and/or stochastic one), thus enhancing the observed transmission disequilibrium of CNVs.

Our results should be viewed carefully in the context of previous studies. Notably, two recent studies (Levy et al. 2011; Sanders et al. 2011) failed to find statistically significant enrichment of inherited CNVs in sporadic autism probands compared to their siblings.

These studies, which also analyzed families from the SSC, used high-density microarray platforms to discover CNVs in genome-wide fashion. It is possible that the increased sensitivity of our exome-based method for genic events— which are most strongly implicated by both *de novo* CNVs and *de novo* SNV studies— revealed the difference in burden between probands and siblings. Additionally, our study found that the differential burden was dependent on the SRS score and not IQ, a factor which has not been previously examined in the context of ASD and CNVs. In contrast, our results are in good agreement with those of the case-control study by Pinto and colleagues (Pinto et al. 2010), who found an overall case/control ratio of 1.19 for genic CNVs and no enrichment for “ultra-rare” CNVs seen only once in their cases (although this study was largely limited to CNVs larger than 50 kbp).

The smaller size of the CNVs provides increased specificity to define individual genes when compared to previous studies focused on large CNVs, which typically encompass dozens of genes. The deletion or duplication of a subset of exons can have, in principle, the same impact on gene function as disruptive point mutations. Accordingly, several genes in our brain-expressed/SRS-discordant set of CNVs have been previously identified as part of severe neurological disorders. Among these was a CNV affecting *DDHD2*, an intracellular phospholipase which plays an essential role in synaptic function, and which has recently been implicated in a recessive form of complex hereditary spastic paraplegia (HSP [MIM 615033])-- a syndrome characterized by early-onset intellectual disability and spastic paraplegia (Schuurs-Hoeijmakers et al. 2012). We did not observe any CNVs in this gene in 2,972 control exomes. Similarly, another proband (and 0/2,972 controls) carried an inherited CNV affecting only the *PACS2* gene (MIM 610423), part of six genes in critical region of 14q32 deletion syndrome, characterized by intellectual disability and mild facial dysmorphism (Holder et al. 2012). Lastly, in two families, we identified a previously unidentified small (~5 kbp), 2-exon deletion of the *ZNF396* gene (Figure 3.1c), which was identified as a candidate gene for Alopecia with Mental Retardation syndrome (MIM 613930) by microsatellite linkage analysis (in fact, *ZNF396* was the closest gene to the linkage peak; Wali et al. 2007). The frequency of this deletion in our control set was 3/2,972 (0.1%). Although these

identified genes and CNVs may play an important role in the pathogenesis of ASD on the basis of their previously identified roles in Mendelian disorders, we would like to emphasize that their individual rarity and overall small effect prevents them from being conclusively identified as having Mendelian effects.

Other genes disrupted by CNVs have functional roles in neural function, brain development or neurobehavioral phenotypes in model organisms (Table 3.3). For example, we identified two independent disruptions of *ORC3* gene (MIM 604972; one shown in Figure 3.1d) encoding a protein of the Origin Recognition Complex. The complex regulates dendritic spines and dendrite arborization in post-mitotic neurons, and has been implicated in olfactory learning and memory in *Drosophila* (Huang 2005). Notable also was a CNV affecting *CPLX1* (MIM 605032; Figure 3.1f), specific to the SNARE neuronal vesicle exocytosis pathway in neurons, as well as CNVs affecting neural receptors such as *HTR3E* (MIM 610123; a subunit of the ionotropic serotonin receptor) and *NETO1* (MIM 607973; Figure 3.1e), a key component of the NMDA-receptor complex and critical for synaptic plasticity and learning in mice (however, this CNV was transmitted to both proband and sibling; Ng et al. 2009a). Previous work has implicated the ubiquitin processing pathway (Glessner et al. 2009), and we found a rare CNV in *UCHL1* (MIM 191342), a ubiquitin-adduct processing enzyme which has strong and specific brain expression, knockout mice show specific neurodegenerative phenotypes (Wada et al. 1999), and recent work has shown it to regulate the *NCAM1* neural cell adhesion molecule (MIM 116930; Wobst et al. 2012) Finally, we found several interesting genes on the basis of brain expression pattern, including 1) an inherited deletion of the *IQSEC1/BRAG2* gene (MIM 610166), which is strongly expressed in the prefrontal cortex and involved in clathrin-mediated endocytosis of AMPA receptors critical to long-term potentiation in mice (Scholz et al. 2010), 2) a duplication of the *ZNF251/ZNF517* cluster on 8q24.3, which have tissue-specific expression highest in the fetal brain and cerebellum (Peter Lorenz 2010), and 3) duplication of *AQP4* (MIM 600308), the primary water transporter in brain glial cells, especially in the amygdala and prefrontal cortex and has been implicated in epilepsy (Binder et al. 2012).

Sample	Chr/Position (hg19)	Size (kbp)	State	# exons	Freq. in 411 quads	Genes in transmitted CNV
12647.p1	1 32,084,793 32,110,465	25.7	Dup	12	2	HCRTR1, PEF1
11872.p1	1 65,730,593 65,831,879	101.3	Dup	4	1	DNAJC6
12719.p1	1 146,715,494 146,767,190	51.7	Del	23	1	CHD1L
12997.p1	2 230,632,269 230,724,290	92	Dup	39	1	TRIP12
12394.p1	2 241,538,067 241,709,123	171.1	Dup	42	3	KIF1A, GPR35, AQP12B, AQP12A, CAPN10
12534.p1	3 12,940,888 12,978,197	37.3	Del	13	2	IQSEC1
13099.p1	3 97,486,951 97,634,880	147.9	Del	19	1	ARL6
12645.p1	4 818,279 845,762	27.5	Dup	5	1	CPLX1, GAK
11773.p1	4 2,641,461 2,835,561	194.1	Dup	34	1	TNIP2, FAM193A, SH3BP2
11066.p1	4 41,258,993 41,259,143	150 bp	Dup	2	3	UCHL1
13385.p1	4 169,083,678 169,086,477	2.8	Del	3	1	ANXA10
13293.p1	5 619,104 644,540	25.4	Dup	9	2	CEP72
12758.p1	6 24,454,242 24,523,153	68.9	Dup	20	1	ALDH5A1, GPLD1
11551.p1	6 88,315,634 88,318,947	3.3	Del	3	1	ORC3
11459.p1	6 88,317,390 88,366,700	49.3	Del	10	1	ORC3
13412.p1	7 33,102,179 33,185,976	83.8	Dup	7	3	RP9, BBS9, NT5C3
11722.p1	7 48,308,576 48,416,169	107.6	Del	20	1	ABCA13
11716.p1	8 38,090,512 38,117,639	27.1	Del	16	1	DDHD2
13412.p1	8 86,351,940 86,575,726	223.8	Dup	14	1	CA3, CA2, REXO1L1
12534.p1	8 145,947,028 146,033,780	86.8	Dup	18	1	ZNF251, ZNF34, ZNF517, RPL8
11356.p1	9 139,634,401 139,651,044	16.6	Dup	16	1	LCN6, LCN10, LCN8
13162.p1	10 5,203,384 5,260,723	57.3	Del	12	1	AKR1C4, AKR1CL1
13843.p1	11 43,772,460 43,775,671	3.2	Del	2	1	HSD17B12
11241.p1	12 120,875,929 120,884,632	8.7	Dup	7	2	GATC, COX6A1, TRIAP1
12396.p1	14 105,836,177 105,861,009	24.8	Dup	17	1	PACS2
11479.p1	15 43,696,610 43,701,294	4.7	Dup	5	1	TP53BP1, TUBGCP4
13843.p1	15 55,475,512 55,497,903	22.4	Dup	6	2	RAB27A, RSL24D1
12837.p1	15 57,730,197 57,754,090	23.9	Dup	7	2	CGNL1
13543.p1	15 91,488,121 91,520,001	31.9	Del	25	2	RCCD1, PRC1, UNC45A
13215.p1	16 15,596,178 15,609,285	13.1	Del	6	1	C16orf45
14201.p1	16 68,710,287 68,713,877	3.6	Dup	5	2	CDH3
12100.p1	16 70,714,696 70,714,928	232 bp	Dup	2	3	MTSS1L
12373.p1	16 81,314,461 81,396,216	81.8	Dup	10	1	GAN, BCMO1
12697.p1	18 24,436,174 24,628,467	192.3	Dup	10	1	CHST9, AQP4, CHST9-AS1
12869.p1	18 72,229,281 72,251,798	22.5	Dup	8	1	CNDP1
11356.p1	18 77,470,345 77,891,075	420.7	Dup	28	2	KCNG2, RBFA, CTDPI, ADNP2, TXNL4A, PQLC1
13296.p1	19 6,681,951 6,686,913	5	Dup	8	1	C3
11298.p1	19 18,704,375 18,704,917	542 bp	Dup	2	1	CRLF1
13815.p1	19 57,835,049 57,932,849	97.8	Del	15	1	ZNF547, ZNF304, ZNF17, ZNF548, ZNF543
13396.p1	21 19,628,825 19,632,603	3.8	Del	3	1	CHODL
13327.p1	21 35,742,777 35,899,047	156.3	Dup	8	2	KCNE2, RCAN1, KCNE1, FAM165B

Table 3.3: Selected inherited CNVs

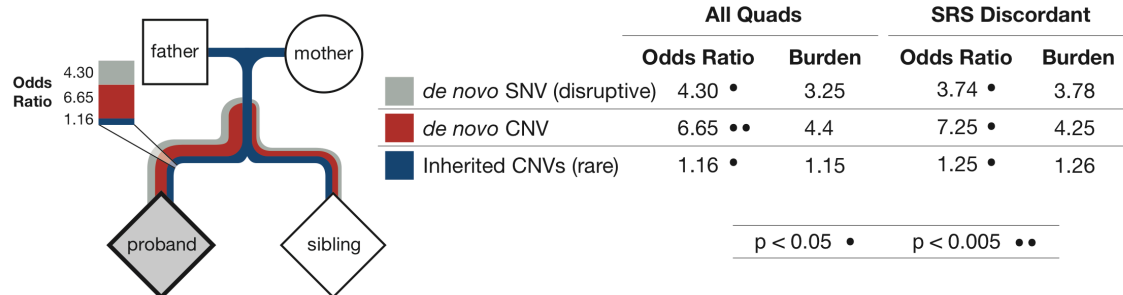


Figure 3.5: A combined model of inherited and *de novo* mutations reveals independent risk for both. A logistic regression model estimates the odds ratio for each inherited CNVs (blue), *de novo* CNV (red) or disruptive *de novo* SNV variant (gray; nonsense, splice and indels only) in probands and siblings. Odds Ratios and burden (proband vs. sibling ratio) given in accompanying table, revealing independent risk for each type of mutation. The line width for each type of mutation in the figure indicates if a bias has been observed for new mutations arising on the maternal or paternal haplotypes (see also: for SNVs, O’Roak et al. 2012 and for CNVs: Hehir-Kwa et al. 2011).

Since the patients and families have been analyzed for both *de novo* CNVs and SNVs, we can develop a model to assess the relative contribution of each class of genetic variant to autism. First, we confirmed that inherited CNVs were enriched in the set of probands without other known *de novo* CNVs or SNVs (368 inherited CNVs in probands vs. 327 CNVs in siblings of 336 quads; $p < 0.03$, two-tailed paired t-test). Second, we developed a logistic regression model, in which the binary outcome of either proband or sibling is predicted by the count of disruptive *de novo* SNVs, *de novo* CNVs and the count of our rare transmitted CNVs. We performed regressions on both the set of all 411 quads, as well as the set of 276 proband-sibling pairs with discordant SRS scores. The results (Figure 3.5 and Table S12) reveal a strong) effect for disruptive (nonsense, splice and frameshift) *de novo* SNVs (OR 4.30, $p < 0.001$) and CNVs (OR 6.65, $p < 0.02$), and also confirmed a statistically independent effect for transmitted CNVs (OR 1.16, $p < 0.04$); again, in this model, the effect was primarily driven by discordant SRS quads. Although the strength of *de novo* SNVs strongly outweighs the pathogenic effect of inherited CNVs, our model predicts that the inherited CNV contribute significantly to sporadic disease, especially in the case of discordant SRS pairs (where the OR increased to 1.26, $p < 0.015$). We did not find any significant interactions between our predictors, reflecting

the relative infrequency of co-occurring CNVs and *de novo* SNVs but also the limited sample size. It is also possible that careful consideration of rare and disruptive, inherited SNVs could statistically interact with other classes of mutation, but we did not take these into account in building our model. Taken together, our model suggests that disruptive *de novo* SNVs and both inherited and *de novo* CNVs contribute independently to the risk of autism. We believe that future studies of ASD and other complex neurological disorders will make significant strides in understanding the genetic underpinnings of disease, especially if an integrated approach considering *all* disruptive mutations— inherited and *de novo*, CNV and SNV, small and large— is applied.

ACKNOWLEDGEMENTS:

We thank the National Heart, Lung, and Blood Institute, NIH Grand Opportunity (GO) Exome Sequencing Project and its ongoing studies, which produced and provided exome variant calls for comparison: the Lung GO Sequencing Project (HL-102923), the Women's Health Initiative Sequencing Project (HL-102924), the Broad GO Sequencing Project (HL-102925), the Seattle GO Sequencing Project (HL-102926), and the Heart GO Sequencing Project (HL-103010). We are grateful to all of the families at the participating Simons Simplex Collection (SSC) sites, as well as the principal investigators (A. Beaudet, R. Bernier, J. Constantino, E. Cook, E. Fombonne, D. Geschwind, E. Hanson, D. Grice, A. Klin, R. Kochel, D. Ledbetter, C. Lord, C. Martin, D. Martin, R. Maxim, J. Miles, O. Ousley, K. Pelphrey, B. Peterson, J. Piggot, C. Saulnier, M. State, W. Stone, J. Sutcliffe, C. Walsh, Z. Warren and E. Wijsman). We also acknowledge M. State and the Simons Simplex Collection Genetics Consortium for providing Illumina genotyping data, T. Lehner and the Autism Sequencing Consortium for providing an opportunity for pre-publication data exchange among the participating groups. We are grateful for helpful discussion and manuscript preparation help from T. Brown and P. Sudmant, as well as all members of the Eichler lab. We appreciate obtaining access to phenotypic data on SFARI Base. This work was supported by the Simons Foundation Autism Research Initiative (SFARI 137578 and 191889; E.E.E., J.S. and R.B.) and NIH HD065285 (E.E.E. and J.S.). E.B. is an Alfred P. Sloan Research Fellow. E.E.E. is an Investigator of the Howard Hughes Medical Institute.

IV. Inherited SNV mutations in Autism Spectrum Disorder

Towards a holistic model of genetic variation in ASD

4.1 Summary

We describe the creation of a combined dataset and resource of inherited and *de novo* SNVs and CNVs across 2,377 Simons Simplex Collection (SSC) families. The dataset includes 1,786 parent-child-unaffected sibling “quads”, which enable comparison of the burden of inherited and *de novo* mutations between affected and unaffected siblings in simplex autism families. We find that private inherited truncating SNV mutations in conserved genes are significantly enriched in probands (OR=1.14, $p < 0.0002$, two-tailed paired t-test), and we observe that this effect becomes more pronounced with increasing gene-level conservation (assessed via the RVIS score). Likewise, we confirm previous reports of transmission disequilibrium for inherited CNVs. Transmission disequilibrium of SNVs was strongest in probands with diagnoses of Autism Disorder or Pervasive Developmental Disorder, and we did not observe a significant enrichment in probands with Asperger’s Disorder; similar results were observed when stratifying by IQ. We quantified ASD risk for *de novo* and inherited CNVs and SNVs by using a conditional logistic regression model, and found that inherited private truncating SNVs and rare inherited CNVs contribute an independent increase in risk of 1.11 ($p=0.0002$) and 1.23 ($p = 0.01$), respectively. Our results confirm a statistically independent role for inherited mutations in ASD risk and identify additional candidate genes (e.g., *RIMS1*, *CUL7* and *CSMD1*) where inherited and *de novo* burden converge.

This chapter is in preparation for publication.

4.2 Introduction

Autism spectrum disorder (ASD) is a common neurodevelopmental disorder diagnosed in approximately 1/88 children and manifests as deficits in social behavior and language development, as well as restricted or stereotyped interests. Several large studies have confirmed initial observations that ASD is a highly heritable, and consensus estimates suggest that ~50-60% of ASD etiologies are genetic (Hallmayer et al. 2011; Bailey et al. 1995; Constantino et al. 2013; Steffenburg et al. 1989). In particular, *de novo* mutations have been implicated as the underlying genetic cause in cases, and these mutations have provided a rich source for understanding the pathogenic genes and neurobiological mechanisms of ASD. However, *de novo* mutations are rare, and their overall contribution is estimated to be 25-35% (Krumm et al. 2014; Ronemus et al. 2014), shy of the overall heritability estimate, suggesting that other genetic etiologies contribute to ASD.

Previous reports have suggested additional genetic models for ASD, in which rare inherited copy number variants (CNVs) are disproportionally inherited by affected probands, rather than their unaffected siblings (Krumm et al. 2013; Poultney et al. 2013; Pinto et al. 2010). These studies describe a transmission disequilibrium for mutations of very low population frequency, suggesting that the pathogenic CNVs for ASD are of relatively young age and under strong purifying selection. In the present study, we hypothesize that inherited single nucleotide variants (SNVs) also contribute to a model of rare inherited variation underlying the genetic etiology of autism. We leverage a family-based study design of simplex autism, in which one offspring carries a diagnosis of autism and an unaffected offspring acts as a genetic control, to discover specific ASD risk genes and integrate inherited factors with *de novo* factors in a general ASD risk model.

This study, in conjunction with the National Database for Autism Research and the Simons Simplex Consortium, also makes available a resource of uniformly processed raw exome sequence (bam) files as well as raw variant (vcf) files from multiple variant calling tools. We have generated both inherited and *de novo* SNVs and CNVs across 2,377 families, include 1,786 quads. We envision that this data becomes a resource for

further study and investigation in the autism research community. The complete set of raw underlying data, variant calls and methods are available through the National Database for Autism Research (NDAR) under the Study ID 334 (see Web Resources for link)

4.3 Methods

Dataset

We analyzed exome data from 2,377 ASD families after quality control of 2,391 families from the Simons Simplex Collection (Fischbach and Lord 2010), including 1,786 quads and 591 trios (total $n=8,917$ exomes). A subset of these families ($n=752$ families) were sequenced as part of three previous publications (Sanders et al. 2012; Iossifov et al. 2012; O’Roak et al. 2012b). We note that Iossifov and colleagues report on aspects of sequence generation, and *de novo* rates (Iossifov et al., in preparation, *Nature*, 2014). This study was approved by the institutional review board of the University of Washington.

Alignment and SNV discovery

We aligned exome data to the GRCh37 reference genome using BWA-MEM (Li 2013; v0.7.5a) and post-processed alignments using the GATK “best practices” pipeline, including indel realignment and BQSR. Exome data was matched to existing SNP barcodes (generated from Illumina 1M/1MDuo SNP microarrays and/or 96-SNP fingerprints collected by the Rutgers sample distribution center) in order to eliminate sample identity/paternity mix-ups. We called SNVs and indels with both GATK HaplotypeCaller (McKenna et al. 2010; v 2.7-4) and FreeBayes (Garrison and Marth 2012; v0.99) to within 20 bp of the exon targets, and were calls were annotated using SnpEFF and merged into union and intersection sets. Allele frequency was estimated by counting non-reference alleles across all parents ($n=4,754$ parents). We define all stop-gained, frameshift and splice-site variants as “Likely Gene Disrupting” or LGD variants.

For *de novo* events, we applied minimum a read-depth of six alternate alleles in offspring and a depth of >10 reference reads in parents, including no more than two low-quality bases of the *de novo* allele. To exclude common artifacts, we only accepted unique *de*

*nov*o sites across all families. Inherited events were derived from the intersection set of both algorithms, with filters depth and quality filters set to DP > 20 and QUAL > 50.

CNV discovery

We used CoNIFER (Krumm et al. 2012) and XHMM (Fromer et al. 2012) algorithms to discover copy number variation from exome data at high exonic resolution. Calls from each algorithm were reconciled, merged, and genotyped within each family to determine inheritance patterns. In order to maximize both sensitivity and the precision of the callset, we used targeted *in silico* genotype information based on available SNP microarray data of each call (n=1,266 families; CRLMM algorithm (Scharpf et al. 2011), see supplementary methods). In order to focus our analysis to those CNVs most likely relevant to ASD pathogenesis, we restrict our analysis to rare CNVs found at less than 0.8% frequency (< 10 events/1,266 families) and outside of repetitive genomic elements.

4.4 Results

Discovery and validation of inherited and de novo SNVs

Starting from raw sequence data, we reprocessed 8,917 exomes from the Simons Simplex Collection in order to standardize their analysis and allow comparison among the entire data set of 2,377 families. Our pipeline entailed remapping using bwa-mem and variant calling using both GATK HaplotypeCaller and FreeBayes, which amounted to over 1,000,000 CPU-hours of computation using Amazon Web Services. Using FreeBayes and GATK, we found a median of 26,920 transmitted variants per family (95% CI 23,394–31,401). Overall, 81% of all transmitted variants were found by both FreeBayes and GATK, 12% by FreeBayes alone and 7% by GATK. Of all transmitted mutations in the intersection set, an average of 341 (95% CI: 133–632) sites per family were novel and not observed in dbSNP (v137); 98.6% of sites were in dbSNP with a mean concordance rate of 99.7% (for the union of transmitted variants, 93.4% of variants were found in dbSNP and 99.5% were concordant; for events discovered only by GATK, 76% were found in dbSNP and these were on average 96.7% concordant; for events discovered only by FreeBayes, 64% were found in dbSNP and these were 99.4% concordant). For intersecting variants, the Ti/Tv ratio was 2.94 (95% CI 2.79–3.03) for all sites, 2.95

(2.83–3.04) for dbSNP sites and 1.94 (1.05–2.75) for novel sites. The median Ti/Tv ratios for transmitted variants specific found only by GATK was 1.34 overall and 1.54 for dbSNP sites; for variants specific to FreeBayes discovery, the Ti/Tv was 2.15 overall and 2.35 for dbSNP sites.

New de novo mutations (SNVs and CNVs)

Our analysis benefited from the use of newer bioinformatics tools, allowing us to discover 1560 new *de novo* mutations previously not detected. We tested a small subset of these (n=75) and validated set of 23 new likely gene disrupting (LGD) mutations and 26 new missense *de novo* mutations in probands (Table 4.1 and Table S1). Notably, these validated mutations established *ASH1L* as a newly recurrently truncated gene, added a new LGD mutation to *GIGYF1* for a total of three LGD *de novo* mutations observed. In addition, the new mutations established recurrent hits in seven new genes, including *GIGYF2* (see below), *ATP1B1* (a gene with strong brain expression), *SSPO* (a brain-secreted protein involved in axon growth), and *JAKMIP1* (previously implicated in ASD and with strong and specific brain expression).

	Mutations (this work)	Previously identified mutations	Status	Siblings
<i>ASH1L</i>	1 LGD	1 LGD	Newly Recurrent LGD	
<i>GIGYF1</i>	1 LGD	2 LGD	Multiple recurrent LGD	
<i>GIGYF2</i>	1 Ms	1 LGD	Newly recurrent	
<i>ATP1B1</i>	1 Ms	1 LGD	Newly recurrent	
<i>SSPO</i>	1 LGD	1 Ms	Newly recurrent w/ new LGD	
<i>JAKMIP1</i>	1 Ms (NV: 12607.pl)	1 LGD	Newly recurrent	1 Ms, 1 LGD
<i>RNF213</i>	1 Ms	1 LGD	Newly recurrent	2 Ms
<i>UBR5</i>	1 Ms (NV: 11102.pl)	1 LGD	Newly recurrent	
<i>ZBTB45</i>	1 LGD	1 Ms	Newly recurrent w/ new LGD	

Table 4.1: genes with new recurrent *de novo* mutations. All mutations validated, except those marked NV (No validation attempted)

We used an intersection of both CoNIFER and XHMM to discover small, exonic CNVs and validated these using custom array-CGH and targeted in-silico genotyping using Illumina SNP microarray data. Of 52 tested CNVs, we validated 21 new (previously unvalidated or undetected) *de novo* CNVs; several of these affected genes recurrently hit by *de novo* SNVs, including *DSCAM*, *CHD2*, and *TNRC6B*.

Investigation of the newly recurrent genes found that three genes from this list (*GIGYF1*, *GIGYF2* and *TNRC6B*) and three additional genes with single *de novo* mutations (*GRB10*, *RBM12*, and *ZNF598*) are closely linked with one-another using protein-protein interaction data (Figure 4.1; Table 4.2). Gene ontology annotation of the genes in this network suggests involvement of the IGF (Insulin Growth Factor) signaling pathway (*GIGYF1*, *GIGYF2*, *GRB10*; accession GO:0048009), which has been previously implicated in the development of ASD (Bozdagi et al. 2013) Furthermore, *GIGYF2* and *ZNF598* form part of the m4EHP mRNA binding complex and have widespread translational repression roles, especially in the brain and lungs (Masahiro Morita 2012).

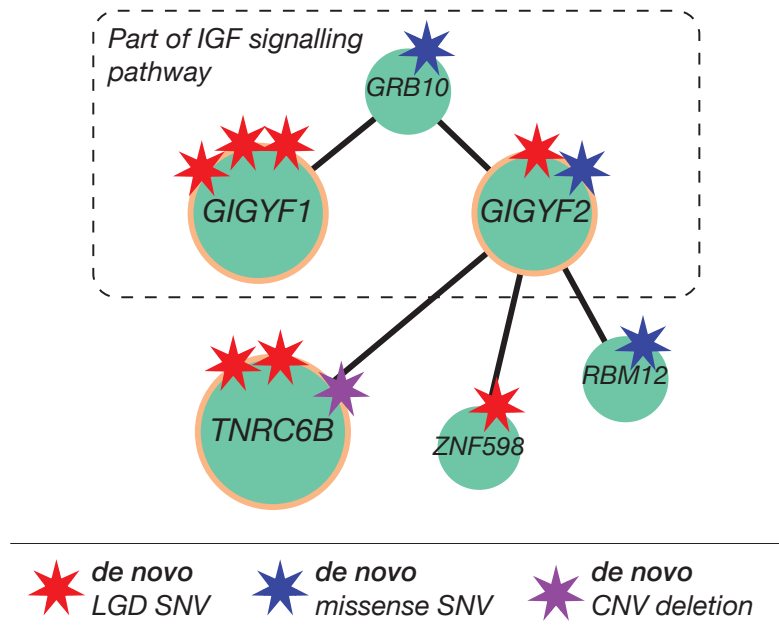


Figure 4.1: Network of genes with recurrent *de novo* hits, based on new *de novo* mutations identified in this study. Red Stars: *de novo* LGD mutations (Frameshift, Stop-gained, Splice-site); Blue stars: *de novo* missense mutations; Purple star: CNV deletion (see Figure S1)

	<i>De novo</i> mutations in probands	ESP6500 Rare LGD SNVs Allele Count [^]	Brain Expression	Part of IGF Pathway
<i>GIGYF1</i>	Fs, Fs*, SS	3	+++ (Cerebellar)	Yes
<i>GIGYF2</i>	Stop, Ms*	0		Yes
<i>GRB10</i>	Ms*	1	+	Yes
<i>TNRC6B</i>	Fs, Stop, CNV del*	101&	+++ (Cerebellar)	
<i>ZFN598</i>	Stop	4	++ (Cerebellar)	
<i>RBM12</i>	Ms	0		

Table 4.2: Summary of mutations in IGF-related ASD network

* Mutations newly identified in this study

[^] Must have minimum read depth of 10 + bi-allelic

& 99/101 are a single 5' frameshift variant (AA position 4949/5502)

Transmission disequilibrium of SNVs between probands and siblings

We tested for transmission disequilibrium between probands and siblings in three ways:

i) by a Fisher's exact test, ii) by paired Student's t-test or Mann-Whitney U test, and iii) by logistic regression (where the dependent variable was if the variant was found in a proband or sibling). We used only variants called by both FreeBayes and GATK variant callers in order to minimize false positives.

We found no statistically significant overall burden when considering all rare or private protein-altering mutations (LGD + missense) together, even when considering additional hypotheses based on highly brain expressed genes, mutations with high CADD scores, or mutations in genes with *de novo* mutations in other probands ($p > 0.05$ in all comparisons). In contrast, we found private LGD mutations in genes which are intolerant of deleterious mutations (as informed by the RVIS (Petrovski et al. 2013) in the lower 50% of all scores) were statistically enriched overall in probands (OR=1.14, $p < 0.0002$, Fisher's exact test) and at a family level ($p < 0.0001$, two-tailed paired t-test; Figure 4.2A). These effects persisted even for all LGD mutations in genes (regardless of frequency) with RVIS scores <50% (OR=1.06, $p=0.03$ Fisher's exact test; $p=0.02$ two-tailed paired t-test). Furthermore, the RVIS score was a significant predictor of proband

or sibling inheritance in a logistic regression model build on all LGD mutations ($p=0.028$, $OR=1.01$ per RVIS percentage point). As suggested by this model, the burden of private LGD mutations in genes with progressively lower RVIS scores continues to increase (Figure 4.2B). At the extreme, the burden between probands and siblings in genes with the lowest 1% of all RVIS scores reaches an odds ratio of 1.4, although this comparison at the extreme is not yet statistically significant (due to the small number of mutations present in this bin).

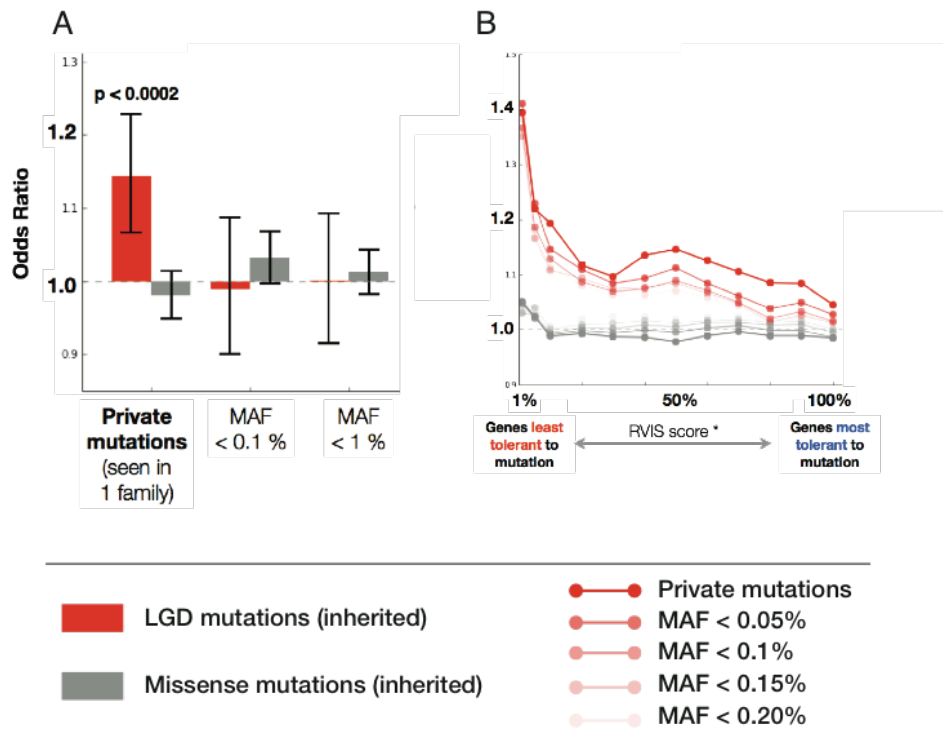


Figure 4.2 Transmission disequilibrium of SNVs in ASD **A)** Private LGD (red bars) inherited SNVs in genes which are not tolerant to functional variation were significantly enriched in probands. The analysis examines only SNVs in genes with an RVIS score in the lower 50%. Non-private rare variants, or missense (gray bars) inherited SNVs are not enriched in probands. **B)** The RVIS score is a critical determinant for enrichment in probands: Burden was highest (reaching $OR=1.4$) for private inherited LGD SNVs amongst genes with the lowest RVIS scores.

We looked for a relationship between the set of private LGD mutations in RVIS-restricted genes and the phenotype of probands in the SSC (Figure 4.3). First, we examined how the overall clinical diagnosis impacted burden: for the 1,575 probands

with a diagnosis of “autism” or “pervasive developmental disorder”, the odds ratio was 1.14 and 1.18 ($p=0.001$ and 0.05), respectively; in contrast, probands with a diagnosis of “Asperger’s” ($n=205$) had a lower odds ratio of 1.05 ($p> 0.7$; Figure 4.3A). Consistent with this, we found that probands with full-scale IQ lower than 70 had an odds ratio of 1.18 ($p = 0.014$, $n=530$), whereas those with IQ above 100 had a lower, non-significant odds ratio of 1.06 ($n=454$; Figure 4.3B). In contrast to IQ, there was no difference in transmission disequilibrium between probands and siblings with highly differential Social Responsiveness Scale scores (“discordant SRS” quads) and those with less extreme scores (data not shown).

Transmission disequilibrium of CNVs between probands and siblings:

There were 2,891 total autosomal CNVs detected in this study with child specific event counts of 854 in probands and 743 in siblings (ratio=1.25, $p=0.006$, binomial two-sided test) with 47.4% of probands and 44% of siblings having a CNV. Overall, proband CNVs (median=40.6 kbp) were slightly larger than sibling events (median=38.4 kbp) but not statistically significant ($p = 0.09$, Wilcoxon). The overall ratio of duplications to deletions was 1.6 consistent with previous results for a smaller SSC dataset (Krumm et al. 2013). Lastly, the number of proband CNVs >500 kbp ($n=85$, median size=1,211 kbp) identified in probands was 2.3-fold higher than in siblings ($n=37$, median size=889 kbp).

Phenotypic measures and autism CNVs

A previous publication has identified significance for CNVs in proband-sibling pairs that are discordant for their Social Responsiveness Score (SRS, where discordant is defined as a proband with $SRS > 75$ and an unaffected sibling with a score < 50) and not in those that are concordant (Krumm et al. 2013). Here, we confirm these results: we find that probands in discordant pairs have a significantly higher burden of CNVs (OR=1.16, $p=0.008$; Figure 4.3C). In contrast, probands were not enriched for transmitted CNVs when their SRS scores were concordant or unremarkable in comparison to their siblings (OR=1.02, $p > 0.1$; Figure 4.3C). When examining IQ, we find that probands with low IQ ($FSIQ < 70$) are enriched for inherited CNVs in comparison to their siblings (OR=1.16, $p=0.04$; Figure 4.3D), but that probands with higher IQs (>70) are not enriched. This is in agreement with our previous report, where we found that probands with low IQ (either

<70 or <85) were enriched for inherited CNVs *versus* their siblings, while those with higher IQ (>85) did not have a higher burden.

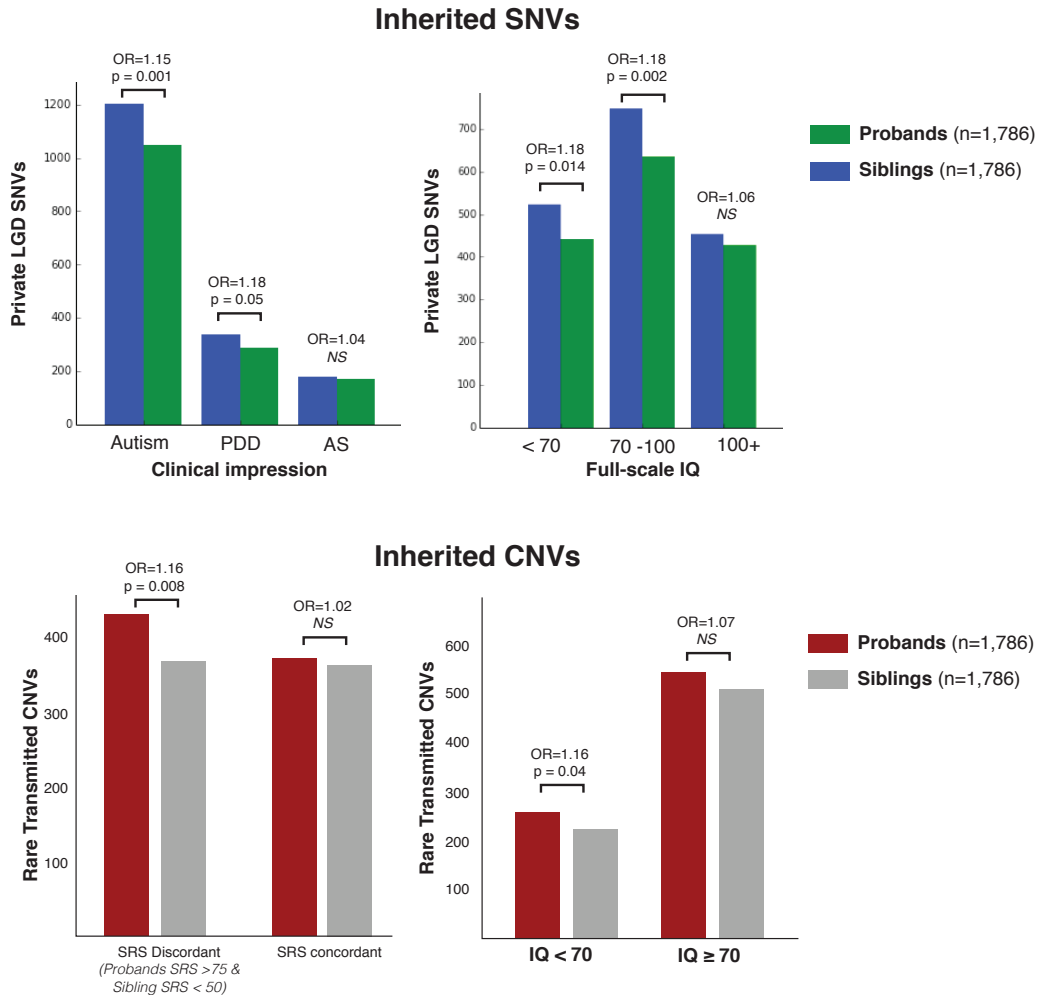


Figure 4.3. Transmitted mutations and their effect on phenotype. Clockwise from top left: (a) Private inherited LGD SNVs enriched in probands with Autism and Pervasive Developmental Disorder (PDD) diagnoses, but not Asperger's Syndrome (AS). (b) Private inherited LGD SNVs primarily enriched in cases with lower IQ than average (<100). (c) We observe transmission disequilibrium of rare inherited CNVs in SRS Discordant families (Proband SRS score > 75, Sibling < 50), but not in families where the SRS score is mild or more balanced between proband and sibling. (d) Rare inherited CNVs are enriched in probands (versus their siblings) with IQ lower than 70, but the effect is not significant in probands with IQ > 70. All tests and reported p-values are paired t-tests based on proband-sibling pairs.

Integration of mutational spectrum suggests new ASD candidate genes

We jointly examined SNVs and CNVs at a gene level in order to suggest new ASD candidate genes (Table 4.2). Events were tabulated based on mutation type (SNV/CNV) and inheritance class as presented throughout this manuscript. In particular, we counted all *de novo* CNVs and LGD or missense SNV events, private LGD-inherited SNVs in genes with an RVIS score < 50%, and rare inherited CNV, in which at least one gene had an RVIS score <50%. From these values, we calculated p-values for *de novo* SNVs (as in O'Roak et al. 2012a) and inherited SNVs and CNVs (using a binomial test). Genes were ranked based on the Fisher's combined p-value heuristic. Finally, in order to remove common 'false-positive' genes, we restricted our analysis to genes with low RVIS scores (<10%) or those with no events in sibling.

	Mutations	RVIS	Notes/Function
<i>RIMS1</i>	- 2 <i>de novo</i> LGD - 2 private inherited LGD - 6 rare inherited LGD - 3 rare inherited LGD in siblings (2 shared with probands)	3.3%	Strong and specific brain expression; Previous candidate in ASD studies
<i>CUL7</i>	- 2 <i>de novo</i> missense SNV - 2 private inherited LGD SNVs	3.4%	Neuronal dendrite patterning function
<i>CSMD1</i>	- 3 <i>de novo</i> missense SNV - 4 private inherited LGD SNVs - 5 inherited CNVs (4 focal to CSMD1)	<0.5%	Strong and specific brain expression; previous assoc. w/ SCZ

Table 4.3: *Converging evidence for RIMS1, CUL7 and CSMD1 from de novo and inherited mutations*

The combined gene-level table identifies several new candidate genes. In particular, the three highest ranked genes—*RIMS1*, *CUL7*, and *CSMD1*—each display brain-specific patterns or have identified neural functions. The highest ranked gene, *RIMS1*, has two *de novo* LGD mutations and two private LGD-inherited mutations in probands. Additionally, there were six additional LGD non-private inherited mutations in probands (two of which are shared with siblings). *RIMS1* has been previously suggested as an ASD candidate by Iossifov and colleagues (Iossifov et al. 2012), and has been also observed by

Mathew State and colleagues (personal communication).; it displays brain-specific brain expression, and disruption of the gene in mice leads to increased post-synaptic density and impaired learning.

CUL7 has two *de novo* and two LGD-inherited mutations in probands (none in siblings); functionally, it is a E3 ligase with high cerebellar brain expression and a selective role in neural dendrite patterning and growth (Litterman et al. 2011; *CUL7* is also the causative gene in 3M syndrome [OMIM:273750], which curiously is *not* associated with abnormal mental development).

Finally, *CSMD1* appears as a strong ASD risk factor candidate. Among the 1,386 quad families, we found multiple types of mutations and events in *CSMD1*: three *de novo* missense mutations, one shared inherited LGD SNVs, and one four rare inherited focal CNVs (one shared with siblings). Overall, there are eight events in probands and two in siblings. In addition, there were three additional LGD-inherited SNVs in probands within the trios. *CSMD1* has the fourth-lowest RVIS score (0.02 percentile) of all genes, suggesting it is highly intolerant to functional mutation; this is born out in examination of mutations in the ESP6500, where *CSMD1* has only six LGD mutations. Comparison of the *CSMD1*-focal inherited CNVs seen in probands with the events seen in the Database of Genomic Variants (DGV) suggests that they are private to the ASD families (i.e., not observed in DGV). Furthermore, the ASD-specific events occur at the exon-dense 5'-end of *CSMD1*, a region nearly devoid of exonic CNVs in the DGV (Figure S#).

Functionally, *CSMD1* exhibits strong and specific brain expression; it functions within the complement control pathway, which has been implicated in synaptic pruning. *CSMD1* in particular has been associated with schizophrenia (Håvik et al. 2011), and damaging variants of the gene segregated in two ASD families with distantly related probands (Cukier et al. 2014).

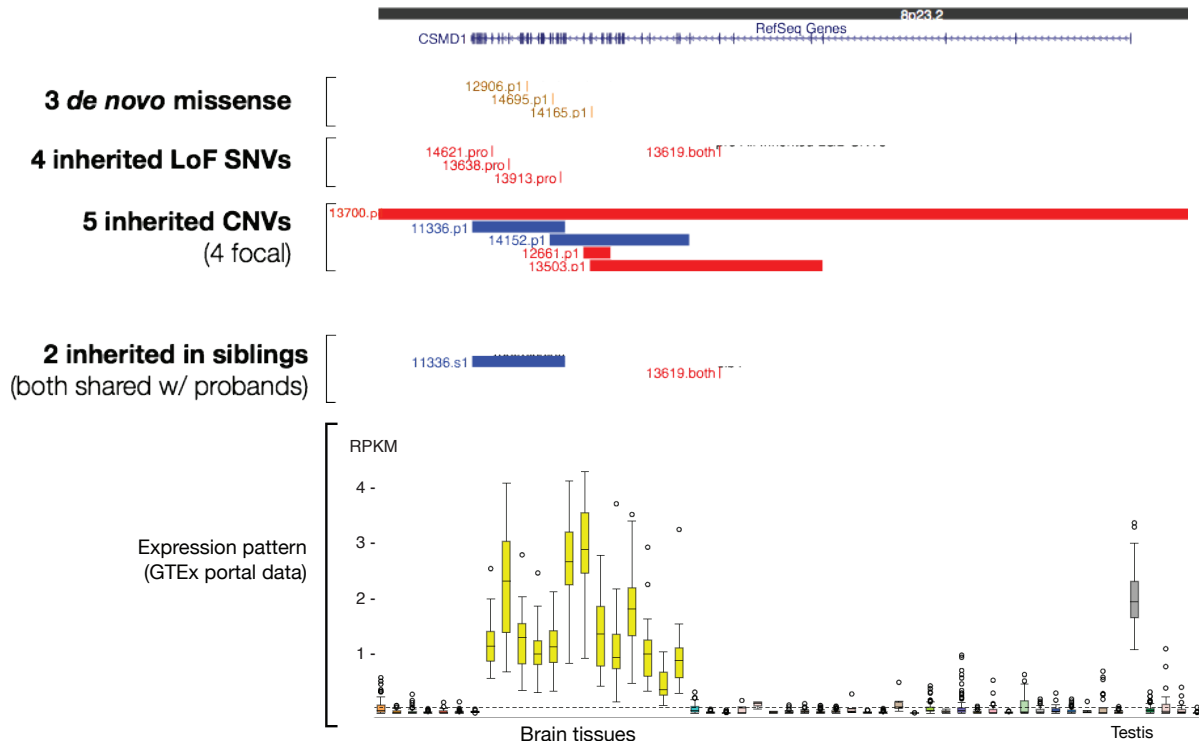


Figure 4.4: Convergence of de novo and inherited mutations on CSMD1. From top: (a) RefSeq gene model of CSMD1 (RVIS score < 1%), (b) three de novo missense mutations in probands, (c) four inherited LGD SNVs in probands, (d) five inherited CNVs, four are focal to CSMD1 alone, (e) of all mutations, only two are shared with siblings (and none are specific to siblings), (f) Expression profile of CSMD1 shows strong brain tissue expression (data from GTEx consortium).

Integration of ASD risk across SNVs and CNVs

We quantified the risk for ASD of *de novo* and inherited CNVs and SNVs by using a conditional logistic regression model (methods; Figure 4.5 and Table 4.4). In this model, the binary outcome of ASD proband or unaffected sibling is predicted by four independent counts: 1) the number of *de novo* CNVs, 2) the number of LGD *de novo* SNVs, 3) the set of rare inherited CNVs and 4) the set of private LGD-inherited SNVs in genes in the lower 50% percentile of RVIS scores. Additionally, we accounted for familial stratification effects by adding a family-level stratum to the model. Using data from the 1,786 quads, we found robust effects for *de novo* events— each *de novo* CNV increased the risk for ASD by 2.05-fold, while each *de novo* SNV increased risk by 1.72-fold ($p = 0.0004$ and $p < 1 \times 10^{-7}$, respectively; Table 4.4). In addition, the results from

this analysis confirm a statistically independent role for inherited mutations in ASD risk: Rare inherited CNVs contribute an increase in risk of 1.23 ($p = 0.01$), and private LGD SNVs have an odds ratio of 1.11 ($p=0.0002$). These results suggest that each of the four domains of mutations modeled additively contribute to the risk of ASD, and that they may do so in statistically independent manner.

We examined the ‘differential’ for each category of mutation in the model, calculated by examining the difference between the percent of probands and percent of siblings with at least one of the mutation type, in order to estimate the proportion of ASD which might be attributable to each type of mutation individually. The strongest differential was seen for *de novo* SNVs (6.6% differential), suggesting this class of mutations is a large portion of ASD heritability. Although inherited LGD SNVs had a low differential in the model (0.1%), we note that the RVIS score of the mutations plays a crucial role: when examining only inherited LGD SNVs with an RVIS score of 10 or lower, the differential jumps to 2.7% (probands=50.5%, siblings=47.9%). Furthermore, even stronger differentials were observed when examining SRS discordant quads only (3.7%), while SRS concordant quads had only a 1.6% differential in inherited LGD SNVs (with RVIS < 10).

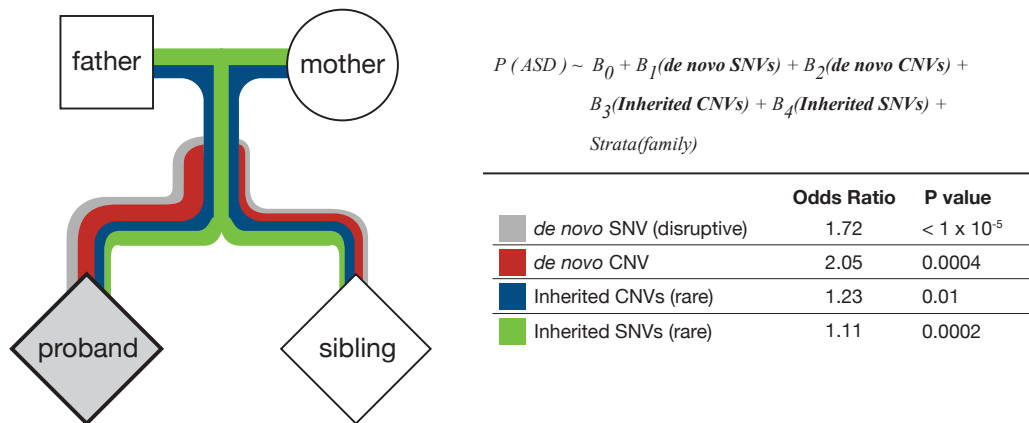


Figure 4.5 Combined risk model for SNVs and CNVs, inherited and *de novo*
Integrative risk model for ASD, based on *de novo* and inherited events, and covering both SNVs and CNVs. The model used is a stratified logistic regression model, which uses proband-sibling pairs to estimate the odds ratio (i.e., risk of ASD) for each type of event.

	% Probands	% Siblings	% Differential	P-value	Odds Ratio
Inherited CNVs	26.1	23.7	2.4%	0.01045	1.23
Inherited SNVs	92.1	91.2	0.1%	0.00024	1.11
<i>De novo</i> CNVs	3.8	1.8	2.0%	0.00039	2.05
<i>De novo</i> SNVs	15.3	8.7	6.6%	0.00000	1.72

Table 4.4: *Summary of logistic regression model results*

Our statistical risk model did not uncover any statistically significant interactions between the main effects, reflecting the relative rarity of each effect type in each individual. In addition, we found no evidence in the data for the presence of non-linear, exponential risk based on the summed number of mutations (methods).

4.5 Discussion

We present the complete ascertainment of inherited and *de novo* mutations in 2,388 families from the Simons Simplex Collection (SSC) of autism. Using a complete “ground-up” reanalysis of the data, and multiple variant discovery tools, we developed a resource of raw data and genetic variants for use throughout the community. Together with the extensive phenotype information present in the SSC, we believe that this resource will enable new and innovative research on the genetic basis and impact of autism.

In the present analysis, we have explored the effect of rare inherited variation on the risk of autism. Our results extend previous work in understanding the role of rare inherited CNVs in ASD (Krumm et al. 2014; Poultney et al. 2013; Pinto et al. 2010), by providing crucial evidence that rare inherited SNVs are also a risk factor for simplex ASD. In particular, we find that private inherited SNVs which likely disrupt the protein product are enriched in probands—but crucially, that these SNVs which disrupt genes intolerant to functional variation in control populations (measured by the RVIS score in the ESP6500 data set) are most enriched in ASD. Disruptive mutations in these genes face

strong selective pressure, suggesting that they may have significant phenotypic consequences.

We have also used the set of 1,786 quads in this analysis to confirm the results of previous work (Krumm et al., 2013) which examined the effect of inherited CNVs in simplex ASD cases on their phenotypes. In particular, we confirm that the SRS score—and especially the differential in scores between probands and their siblings (defined as discordant or concordant SRS scores)—is an important discriminant in CNV burden. Furthermore, we also confirm that probands with lower IQ scores are enriched for inherited CNVs, but those with higher IQ are not enriched in comparison to their siblings. These results suggest that more severely impacted ASD cases (as measured by SRS or IQ) are also enriched for

Exome sequence data provides the basis for detailed, gene-level examination of variants. In this study, we leverage exome sequence data to discover SNVs and CNVs, and use the convergence of inherited and *de novo* events to identify new ASD risk factors. We hypothesize that rare inherited mutations can highlight genes in one of two ways. First, they can narrow the focus onto those genes with identified *de novo* mutations but no or few recurrent *de novo* mutations. In this study, we identify several genes, such as RIMS1 and CSMD1, for which a combination of inherited and *de novo* mutations of both SNVs and CNVs paints a strong picture of ASD risk for these genes.

In a second approach, we find that examination of multiple mutations within each probands reveals a ‘multiple hit’ model for ASD. Critically, this study is the first to examine the complete genetic picture at an individual level in the context of autism. In particular, we found an inherited two-exon intra-genic deletion of *NRXN3* and a *de novo* missense mutation of *NLGN2* in 13367.p1. Both of these genes have been identified as ASD risk factors, but crucially, they are also protein-protein interacting partners. The neuroligin-neurexin interaction has long been hypothesized to be a key underlying pathway in ASD pathology (for review see Abrahams and Geschwind 2008), but to our knowledge this is the first identification of a case with mutations in both binding partners.

Finally, our ground-up reanalysis resulted in the identification of several new *de novo* mutations in genes, and added to the extensive existing work on *de novo* mutations in the SSC. The additional 49 validated *de novo* mutations discovered in the present analysis added several new genes to the “recurrently hit” list of genes (Table S2). Several of these “newly recurrent” genes form a network of protein-protein interactions (Figure 4.1), suggesting an underlying neurobiological pathway for ASD. Interestingly, three of these genes (*GIGYF1*, *GIGYF2*, and *GRB10*) all participate in IGF-pathway signaling, dysregulation of which has been previously suggested as an underlying neurobiological cause of ASD (Chen et al. 2014).

V. Summary and Future Directions

5.1 Summary of results

This thesis describes the development of a method (CoNIFER) to find a new class of genetic variation (small CNVs) and its use in more comprehensively assaying genetic variation in simplex autism families. Using exome sequence data from over 1,800 quads and nearly 600 trios in conjunction with CoNIFER and newer bioinformatics methods, I discovered new *de novo* variants and assayed uniformly for the first time in both inherited CNVs and SNVs. Integration of the spectrum of genetic variants yielded new insight into their relative contribution to simplex ASD risk and highlighted how convergence of multiple mutation types can identify new ASD risk genes. Although these results are in strong agreement with the highly heritable nature of ASD established through twin and sibling studies, and exome sequencing has identified over a dozen ASD-related genes, clear and specific genotype-phenotype correlations have not yet been established. Here, I summarize and highlight some of the conclusions from this thesis and contextualize them in light of remaining challenges and questions.

5.2 Towards assaying the complete set of genetic variation

The importance of *de novo* and ultra-rare transmitted variants in the etiology of ASD makes the sensitivity and specificity of variant detection and discovery algorithms a critical factor in our understanding of ASD genetics. In this thesis, I took two approaches to increasing sensitivity and/or specificity of the variants discovered.

In chapter one, I describe CoNIFER, a new algorithm that extends the use of exome sequence data and leverages it to find genic CNVs. In particular, the targeted nature of the exome sequence data to the exons provides a powerful way to assay very small CNVs that disrupt genes. These CNVs are often too small to be detected using standard genome-wide microarrays, suggesting they have not been previously accounted for in our understanding of ASD. In chapter two, I found that CoNIFER was able to identify over 40% of novel, validated gene-containing CNVs, which previously *had* been missed by

high-density Illumina 1M/1M Duo (with 1.1 and 1.3 million probes, respectively). Many of these CNVs disrupted individual genes, which were enriched for brain expression or previous associations with neurodevelopmental disorders. In addition, the high sensitivity afforded by CoNIFER sharpened the contrast between probands and unaffected siblings, and I was able to demonstrate a significant increased burden of CNVs in the affected probands. This differential burden was especially strong for proband-sibling pairs who were strongly discordant for social phenotype (via the SRS score). Finally, the results suggested that inherited CNVs in probands were maternally inherited.

In chapter three, I use multiple tools (i.e., FreeBayes and GATK, CoNIFER and XHMM) to establish both a highly specific set of variants (based on the intersection of each pair) and a highly sensitive set of variants (based on the union). The use of multiple algorithms resulted in an increased yield of loss-of-function *de novo* variants, added two new genes (*ASH1L* and *IRF2BPL*), to the list of recurrently hit genes with multiple observed *de novo* likely gene disrupting (LGD) mutations and implicated a new functional network of IGF-related proteins. Furthermore, the highly specific intersection set of these tools provided the basis for comparing rates and burden between probands and unaffected siblings.

However, exome sequencing is biased by amplification, sequencing and enrichment biases, which are especially dependent on the %GC content of targets. These biases reduce coverage or prevent sequencing altogether of regions with high %GC nucleotide content. As can be seen in Figure 5.1, GC-bias can prevent adequate coverage of entire domains and exons in genes, in turn preventing variant discovery. Importantly, these regions harbor previously identified autism and neurodevelopmental genes, such as *SHANK3*. *De novo* mutations of *SHANK3* were expected to account for up to 1% of autism cases, yet no *de novo* mutations have been observed so far using exome sequencing (in contrast, several mutations in *SHANK2* have been observed). These results suggest that a large fraction of ASD-related mutations and genes are not assayed using current data. Several methods are available to assay the GC-rich portion of the exome, including molecular inversion probe (MIP)-based resequencing and whole-genome sequencing, although both are still prone to some bias. Future studies will need to not

only critically evaluate which portion of the genome/exome is missed but will need to systematically overcome this bias.

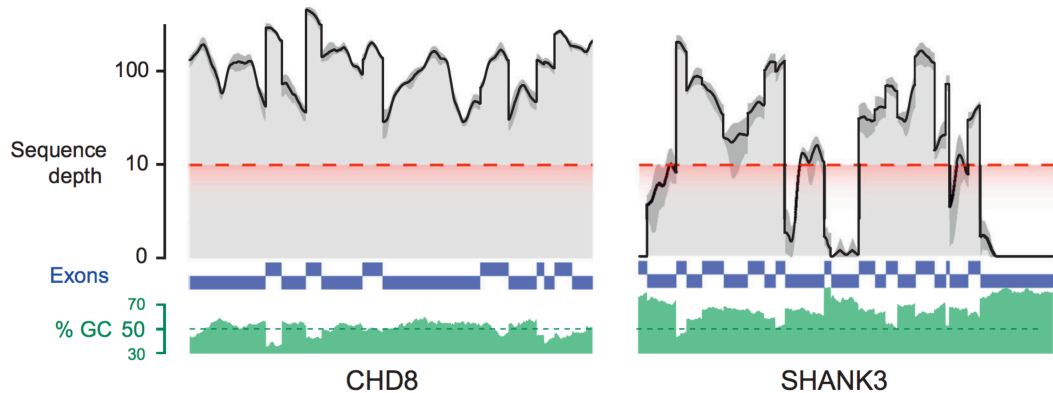


Figure 5.1: High genomic GC nucleotide content (green histogram) hinders whole-exome sequencing for some genes, such as *SHANK3* (right). Individual coding exons are shown in blue with non-genic sequences removed. Black lines indicate mean sequence depth for a single ASD trio and dark gray intervals indicate maximum and minimum depth across the family. The red dashed line indicates the minimum threshold required for accurate variant detection.

Whole-genome sequencing (WGS), in particular, is likely to provide the underlying basis for a “complete” view of genetic variation, even if results are initially “masked” to the genic portion of the genome. In addition to being much less prone to GC bias, the uniform coverage of WGS across all portions of the genome enables algorithms that capitalize on paired-end sequence information. Such algorithms are well suited for identifying fine or complex structural variation, such as very small CNVs (affecting only one or two exons), which still elude exome-based methods (as a single exon or “probe” is nearly always insufficient to identifying a CNV).

Moreover, WGS addresses two important remaining spaces of the human genome that have not been fully assayed to date: First, non-genic space and regulatory elements, and second, genes or regions that are part of segmentally duplicated portions. Both are more fully described below.

Mutations in non-genic regulatory elements of the genome are potentially a pervasive component of ASD genetic etiology, especially in the context of a susceptible genetic

background. Such regulatory backgrounds may in fact be relatively more common than the mutations seen in exons and genes. There is little present understanding of the expected penetrance of such regulatory mutations, and it is possible that the penetrance of ASD-implicated regulatory mutations mirrors that of the genes they regulate—that is, regulatory mutations of highly penetrant genes such as *CHD8* may be equally penetrant (e.g., as in mutations affecting *CFTR* expression, see Rowntree and Harris 2003), while other mutations may be of significantly lower penetrance (Walsh et al. 2008). Crucially, however, these mutations and backgrounds are still expected to be very rare and fall below the frequency threshold of genome-wide association studies.

Another possibility for regulatory mutations is that they contain a subset of highly penetrant *de novo* mutations with similar properties to loss-of-function mutations seen in genes. Yet, given the difficulty in identifying important regulatory elements in the genome, our limited understanding of their effects, and the enormous genomic space within which these mutations can fall, such mutations may be extraordinarily difficult to identify in a *de novo* paradigm. Instead, such mutations may be more readily identifiable using large families with recurrent autism and no strongly identifiable genic mutations.

Finally, studies of mutations in ASD have systematically ignored or filtered genes in segmental duplications, due in part to the ambiguity and difficulty involved in accurately mapping sequence data to incomplete or truncated haplotypes. However, segmental duplications may be very important to a complete understanding of ASD, as they contain over 1,000 genes and are enriched for neurodevelopmental and brain-related functions (Sudmant et al. 2010). These duplications represent the youngest structural changes to the human genome—and many are specific to the *homo* lineage. Furthermore, while segmental duplications are known to be a critical driving factor in non-allelic homologous recombination (NAHR), it is not yet understood how the breakpoints of NAHR events may affect the genic content of segmental duplications. Study of these regions has not been comprehensively possible using either array-based or exome sequence data; however, WGS data and analysis may reveal that these regions can contribute additional genetic risk for ASD.

5.3 Understanding normal variation and pathogenic variants

The extreme locus and variant heterogeneity of ASD has made statistical association and identification of genes and variants difficult. To combat this heterogeneity, many strategies rely on either reducing the genetic space searched or adding additional abstraction to the genetic data examined. In this thesis, a critical insight was the use of the RVIS score in chapter four. The RVIS score was able to stratify private, LGD-inherited SNVs into two groups: a group of variants that affected genes where functional variation was *commonly observed*, and a group of variants that truncated genes which were *intolerant* to functional variation. This is broad distinction at a gene level and reduction of the genetic/genomic *space* (i.e., it reduced the number of genes studied) was critical in establishing private, LGD-inherited SNVs as a risk factor for ASD as well as suggesting new candidate ASD genes.

However, gene-based scores such as the RVIS score may only be a stepping stone to understanding genetic variants at a variant level. Currently, the agreement between variant pathogenicity scores (such as PolyPhen2, SIFT, MutationTaster, etc.) is low and poorly stratifies risk versus non-risk variants in ASD. Although some methods, such as the CADD score (Kircher et al., 2014), are designed to aggregate multiple variant-level scores and their genomic context, I found that the CADD score did not substantially add to the identification of SNVs with ASD risk (notably, the CADD model does *not* include the RVIS score).

In order to improve our understanding of ultra-rare variants and their functional impact, additional work in three domains is needed. First, while the NHLBI's ESP 6500 project has provided an excellent base for understanding rare variation, cohorts of 50,000+ are likely needed to understand the spectrum of ultra-rare variation in human populations. In addition, trio- or family-based studies of control individuals are needed in order to understand the specific pattern of *de novo* variation.

Second, a critical next step in understanding functional variation in humans is improving the definition and identification of gene and transcript models. Over 90% of transcripts undergo alternative splicing, and over 60% are tissue-regulated (Wang et al. 2008). Crucially, these gene and transcript models are the first step in the identification of mutations as non-genic, synonymous, nonsynonymous, or likely gene disrupting (LGD). In many cases, multiple transcripts exist for a single locus/gene, resulting in multiple annotations for just one SNV. In this case, many publications (including this one) take the “most severe” effect as the effect of that SNV, a practice that biases towards more severe effects to be reported, or the “longest transcript”, which likely over-estimates the number mutations that fall in coding regions.

I analyzed the frequency of variant annotation disagreement for variants on currently known transcripts using a set of 2,529 *de novo* missense, stop-gained and frameshift variants from published sources, by taking the “most severe” functional annotation for each variant (O’Roak et al. 2012b; Iossifov et al. 2012; Sanders et al. 2012). Of these, 151 (6%) were also annotated as intronic (i.e., not protein-altering), and this fraction was slightly higher for frameshift and nonsense mutations (32/418, 7.7%) than for missense mutations (119/2,111, 5.6%), suggesting a measurable bias due to the selection of the “most severe” effect at each variant.

Long-read technology and RNAseq has the potential to dramatically improve our understanding of transcript models and increase our confidence in functional variant annotation (for reviews of this topic, see Garber et al. 2011 and Mutz et al. 2013). These improvements will not only create a canonical reference set of transcripts for each gene but have the potential to create tissue- or cell-type specific transcripts (Lonsdale et al. 2013). This information will make it possible to annotate variants to a much smaller subset of transcripts, improving the specificity of the variant effects. Finally, larger transcript-level datasets will also improve our understanding of splice-site variation. Current annotation tools crudely annotate a variant as a “splice-site disrupting variant” based on proximity to exons alone; future models of variant effects and transcripts will undoubtedly improve on this.

Finally, the wealth of information available at variant, gene, and gene-network levels continues to be a challenge to the interpretation of genomic variants. Future interpretation schemes will need to integrate information across levels, including, for example, the impact of an SNV on a gene at an amino acid level, the transcripts and expression levels of that gene within specific tissues of interest, and how that gene is connected within functional pathways and networks. Encouragingly, such improvements can provide both a wealth of value for new experiments as well as dramatically improve the value of existing data, such as the exome data from the SSC..

5.4 Defining subtypes of ASD

Exome sequencing provides a unique opportunity not only to understand ASD from a genetic viewpoint but also from a phenotypic point of view. This “genotype-first” approach, in which “genetic subtypes” of ASD are defined by mutations in genes or pathways promises to create meaningful distinctions within the ASD spectrum for both researchers and patients (Stessman et al. 2014). By assaying and stratifying patients based on the genetic data, the “search space” for possible associated genotypes is greatly reduced, and individual patterns (not significant at a study level) become more pronounced. One example of such ASD “genetic subtypes” is the *CHD8* mutation, associated with larger head circumference, gastro-intestinal problems and distinctive facial features. Notably, however, mutations in *CHD8* are not specifically associated with low IQ, and nearly half of patients with disruptive *de novo* mutations in *CHD8* have normal or above normal IQ. Follow-up study of *CHD8* mutations in zebrafish has confirmed these phenotypic effects (including slower gastrointestinal motility and wider spaced eye cusps, a proxy for brain or head size; Bernier et al. 2014).

The ability to find genetically and phenotypically similar subgroups within the larger ASD spectrum may also lead to the identification genes which are “specific” to the social and language deficits of ASD, rather than those that broadly impair IQ or cortical functioning. In chapters three and four, I explored how rare inherited CNVs were enriched in probands, and specifically in those probands with the poorest (i.e., highest)

scores on the Social Responsiveness Scale (SRS), which tracks a child's ability to communicate and interact socially. The families with the greatest spread in SRS score between the affected probands and his/her unaffected sibling also showed the greatest probands enrichment for inherited CNVs. In contrast, I saw no additional enrichment of rare inherited CNVs amongst probands with lower IQs, suggesting that many rare inherited CNVs may be specific to ASD-related phenotypes. Interestingly, when I examined the relationship of private inherited LGD SNVs (in chapter four) and these two phenotypic measures, I found a different relationship: Probands with lower IQ *were* enriched for private LGD SNVs. The difference in phenotypic effect for inherited CNVs and inherited SNVs may be explained in several ways. First, it is possible that CNV duplications, which triple gene dosage as opposed to reducing copy number (as CNV deletion or LGD mutation of a genes does), are more likely to result in social impairment, perhaps because triplication would dysregulate pathways, rather than abrogate their function. A second possibility is that a subset of private inherited LGD SNVs has a higher phenotypic impact than do CNVs; in this scenario, the observed inherited SNVs would be expected to be very young mutations, which are quickly purged from populations. A final possibility is the reverse, in that it is in fact CNVs that are more phenotypically impactful (and thus more likely to have more severe IQ-reducing effects) and are purged so quickly from populations that only relatively more benign CNVs remain as transmitted CNVs (in this scenario, the highly impactful CNVs are observed simply as *de novo* CNVs).

5.5 Defining a gradient of simplex and multiplex autism

The study of simplex autism families and a quad-based approach that includes an unaffected sibling has been instrumental in understanding both the effect of *de novo* mutations in ASD, as well as the relative risk of inherited genetic mutations for ASD. However, the ascertainment of simplex autism families is complicated by effects of *stoppage* (i.e., families are more likely to stop having more children when one child is diagnosed with ASD), and the fact that family sizes are often too small to rule out an inherited phenotype. A better understanding of the impact and phenotypic effects of inherited and *de novo* mutations will be gained by examination of larger families,

including those with multiple affected offspring. In particular, a community-based or longitudinal study of all ASD diagnoses is a logical next step in understanding the balance of influence from inherited and *de novo* genetic etiologies in the risk of ASD.

5.6 Understanding complex genetic etiologies at a family level

The exome sequencing and targeted resequencing studies of the past three years have created a wealth of new information about specific, highly penetrant genes, such as *CHD8* and *SCN2A*, and their role and risk in ASD. These studies have established that *de novo* mutation and rare variants play a dominant role in the genetic etiology of ASD and have identified mutations in highly penetrant genes that are likely causative. In this thesis, I have examined the role of inherited mutations in the context of ASD genetic etiology and identified inherited ASD risk factors and genes. Taken together, however, the identified gene candidates from both studies of *de novo* and inherited variants can only explain a small fraction of the overall heritable risk for ASD.

Identification of additional specific genes implicated in ASD can be achieved through the study of much larger cohorts (Figure 5.2) in order to establish genome-wide significance. However, these estimates assume a very high and constant penetrance and ASD risk for all *de novo* mutation of genes. Critically, if *bona fide* ASD risk genes are not fully penetrant, then even larger cohorts of samples will be required for statistical association and discovery of genes. In fact, the fraction of genes with multiple *de novo* hits that are *not* fully penetrant can be estimated by examining how many of these genes also have inherited LGD mutations in siblings. From the data generated in the reanalysis of the SSC exomes, 40/128 of genes with two or more *de novo* mutations in probands (and none in siblings) also have at least one LGD SNV mutation or rare CNV inherited in siblings, suggesting that these genes have reduced penetrance for ASD. Thus, this analysis suggests that very large cohorts will be required for the large-scale identification single ASD risk genes using a *de novo* sequencing approach.

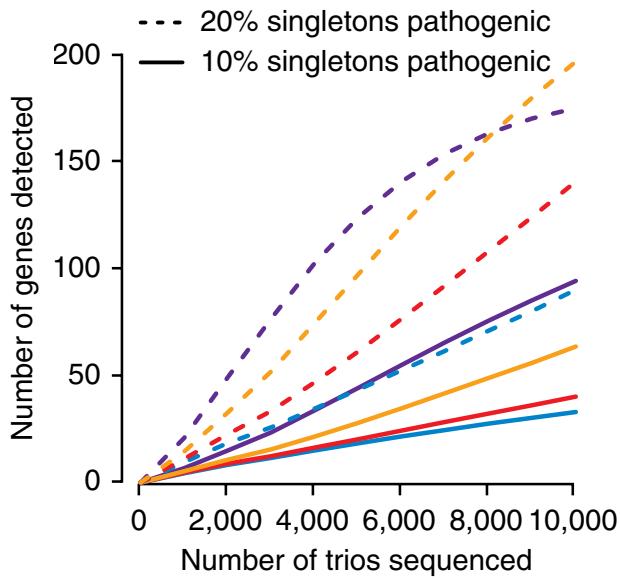


Figure 5.2: Expected hit rate (or sensitivity) of true positive genes discovered using trio sequencing studies (under a family-wise error rate of 5%; that is, each gene passes exome-wide significance of 2.6×10^{-6}). We estimate the power of trio sequencing to detect statistically significant associations for disease-associated genes, under the assumption that 10% or 20% of singleton mutations could be fully penetrant.

A critical component to identifying additional ASD genes and the underlying neurobiological pathways will be comprehensive analysis of mutations at an individual level. Such an approach will leverage the existing genotype data—including “one-off” *de novo* mutations in genes as well as inherited CNVs and SNVs—along with pathway and protein network metadata about the genes and pathways affected. Identification of interactions between genes with inherited and *de novo* mutations will be a powerful method to implicate new candidate genes and may be able to increase the fraction of cases for which a plausible genetic etiology can be identified.

As an example of such individual-level pathway identification, in chapter four I identified an ASD proband with a *de novo* missense mutation *NLGN2*, in addition to an inherited 2-exon intragenic deletion of *NRXN3*. These two genes are part of the neuroligin-neurexin interaction pathways, which directly mediate the trans-synaptic interface of neurons. While both genes have been implicated in ASD, this specific case illustrates how an incompletely penetrant inherited mutation and a *de novo* missense mutation can act in a

“synergistically” pathogenic manner. Using this framework, in conjunction with additional information about cellular- and tissue-level co-expression, will undoubtedly reveal additional ASD risk gene combinations and pathways.

5.7 Future directions

New sequencing technology and the establishment of large well-phenotyped family-based cohorts, such as the SSC, have enabled the systematic discovery of mutations that underlie the genetic etiology of ASD and ID. The fraction of explained genetic etiology is a measurable indicator of progress. In 2005, ~10% of the genetic etiology of autism was understood. Within seven years, advances in genomics technology facilitated the rapid discovery of *de novo* SNVs and CNVs leading to the discovery of disruptive genetic variants that may account for another ~25% of cases. Although the extent of locus heterogeneity in ASD and ID was initially underestimated, the development of exome sequencing and low-cost/high-throughput MIP-based resequencing has strongly implicated two dozen novel genes accounting for >3% of disease.

Many of these genes may in fact define distinct clinical “subtypes” of ASD upon detailed examination of patients with a common genetic etiology, consistent with the hypothesis that autism is an umbrella term underlying many different and distinct “autisms”. This is reminiscent of the work with CNVs, where the identification of recurrent mutations and patient follow-up led to the identification of novel syndromes and subtypes from idiopathic cases of disease (Sharp et al. 2006). There is already compelling evidence for this based on an assessment of multiple patients with *ADNP* (Helsmoortel et al. 2014), *DYRK1A* (Courcet et al. 2012) and *CHD8* mutations (Bernier et al. 2014), which appear to define microcephalic and macrocephalic subtypes, respectively. Alternatively, the “genotype-first” approach may also reveal phenotypic variability of genic mutations across a diverse array of neuropsychiatric and neurodevelopmental disorders. Similar to the CNVs of 16p11.2 and 15q13.3, which are associated with several disorders, there is evidence for this already for mutations associated with *SETBP1* (Hoischen et al. 2010) and *SCN2A*, resulting in very different outcomes. Establishment of cohorts with different

types of mutations and careful study of their phenotypes and comorbidities may reveal specific protein domains and mutation types associated with different diseases.

The knowledge of specific genes, loci, and pathways now spurs the development of functional experiments. These include using novel methods with induced pluripotent stem cells to assay specific mutations in a patient with Timothy syndrome (Yazawa et al. 2011; Paşca et al. 2011), as well as established model systems, such as mouse and zebrafish models to explore the roles of *CHD8* (Bernier et al. 2014), *DYRK1A* (Ahn et al. 2006) and *PTEN* (Backman et al. 2001) in brain volume.

Improving knowledge of ASD genetic and neurobiological etiologies will aid in the diagnosis of ASD/ID subtypes, allowing for specific recruitment for clinical trials and the development of targeted therapeutics for each subtype. This model is akin to the heterogeneity seen in other broad categories of human disorders and disease and has proven to be successful in many cases (e.g., specific therapeutics for a particular mutation in cystic fibrosis or specific forms of cancer). Integrating the genetics, neurology, and pathophysiology of these disorders holds considerable promise not only for our understanding of the biology of the human brain but also for potential treatments.

REFERENCES

- Abrahams BS, Geschwind DH. 2008. Advances in autism genetics: on the threshold of a new neurobiology. *Nature Reviews Neuroscience* **9**: 341–355.
- Ahn K-J, Jeong HK, Choi H-S, Ryoo S-R, Kim YJ, Goo J-S, Choi S-Y, Han J-S, Ha I, Song W-J. 2006. DYRK1A BAC transgenic mice show altered synaptic plasticity with learning and memory defects. *Neurobiol Dis* **22**: 463–472.
- Alkan C, Kidd JM, Marques-Bonet T, Aksay G, Antonacci F, Hormozdiari F, Kitzman JO, Baker C, Malig M, Mutlu O, et al. 2009. Personalized copy number and segmental duplication maps using next-generation sequencing. *Nat Genet* **41**: 1061–1067.
- Amir RE, Van den Veyver IB, Wan M, Tran CQ, Francke U, Zoghbi HY. 1999. Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat Genet* **23**: 185–188.
- Backman SA, Stambolic V, Suzuki A, Haight J, Elia A, Pretorius J, Tsao MS, Shannon P, Bolon B, Ivy GO, et al. 2001. Deletion of Pten in mouse brain causes seizures, ataxia and defects in soma size resembling Lhermitte-Duclos disease. *Nat Genet* **29**: 396–403.
- Bailey A, Le Couteur A, Gottesman I, Bolton P, Simonoff E, Yuzda E, Rutter M. 1995. Autism as a strongly genetic disorder: evidence from a British twin study. *Psychol Med* **25**: 63–77.
- Bailey JA, Gu Z, Clark RA, Reinert K, Samonte RV, Schwartz S, Adams MD, Myers EW, Li PW, Eichler EE. 2002. Recent segmental duplications in the human genome. *Science* **297**: 1003–1007.
- Bajpai R, Chen DA, Rada-Iglesias A, Zhang J, Xiong Y, Helms J, Chang C-P, Zhao Y, Swigut T, Wysocka J. 2010. CHD7 cooperates with PBAF to control multipotent neural crest formation. *Nature* **463**: 958–962.
- Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, Shendure J. 2011. Exome sequencing as a tool for Mendelian disease gene discovery. *Nature Reviews Genetics* **12**: 745–755.
<http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=21946919&retmode=ref&cmd=prlinks>.
- Barton A, Fendrik AJ. 2013. Sustained vs. oscillating expressions of Ngn2, Dll1 and Hes1: A model of neural differentiation of embryonic telencephalon. *Journal of Theoretical Biology* **328**: 1–8.

- Batsukh T, Pieper L, Koszucka AM, Velsen von N, Hoyer-Fender S, Elbracht M, Bergman JEH, Hoefsloot LH, Pauli S. 2010. CHD8 interacts with CHD7, a protein which is mutated in CHARGE syndrome. *Human Molecular Genetics* **19**: 2858–2866.
- Bedogni F, Hodge RD, Elsen GE, Nelson BR, Daza RAM, Beyer RP, Bammler TK, Rubenstein JLR, Hevner RF. 2010. Tbr1 regulates regional and laminar identity of postmitotic neurons in developing neocortex. *Proceedings of the National Academy of Sciences* **107**: 13129–13134.
- Ben-Shachar S, Lanpher B, German JR, Qasaymeh M, Potocki L, Nagamani SCS, Franco LM, Malphrus A, Bottenfield GW, Spence JE, et al. 2009. Microdeletion 15q13.3: a locus with incomplete penetrance for autism, mental retardation, and psychiatric disorders. *J Med Genet* **46**: 382–388.
- Bernier R, Golzio C, Xiong B, Stessman HA, Coe BP, Penn O, Witherspoon K, Gerdtz J, Baker C, Vulto-van Silfhout AT, et al. 2014. Disruptive CHD8 Mutations Define a Subtype of Autism Early in Development. *Cell* **158**: 263–276.
- Berryer MH, Hamdan FF, Klitten LL, Moller RS, Carmant L, Schwartzentruber J, Patry L, Dobrzyniecka S, Rochefort D, Neugnot-Cerioli M, et al. 2012. Mutations in SYNGAP1 Cause Intellectual Disability, Autism, and a Specific Form of Epilepsy by Inducing Haploinsufficiency. *Human Mutation* **34**: 385–394.
- Betancur C. 2011. Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. *Brain Res* **1380**: 42–77.
- Binder DK, Nagelhus EA, Ottersen OP. 2012. Aquaporin-4 and epilepsy eds. C. Steinhäuser and D. Boison. *FEBS J* **60**: 1203–1214.
- Bozdagi O, Tavassoli T, Buxbaum JD. 2013. Insulin-like growth factor-1 rescues synaptic and motor deficits in a mouse model of autism and developmental delay. *Mol Autism* **4**: 9.
- Cadigan KM. 2008. Wnt/ β -Catenin Signaling: Turning the Switch. *Developmental Cell* **14**: 322–323.
- Campbell CD, Sampas N, Tsalenko A, Sudmant PH, Kidd JM, Malig M, Vu TH, Vives L, Tsang P, Bruhn L. 2011. Population-Genetic Properties of Differentiated Human Copy-Number Polymorphisms. *The American Journal of Human Genetics* **88**: 317–332.
- Chen J, Alberts I, Li X. 2014. Dysregulation of the IGF-I/PI3K/AKT/mTOR signaling pathway in autism spectrum disorders. *International Journal of Developmental Neuroscience* **35**: 35–41.
- Chenn A, Walsh CA. 2003. Increased neuronal production, enlarged forebrains and cytoarchitectural distortions in beta-catenin overexpressing transgenic mice. *Cereb*

Cortex **13**: 599–606.

Chiang DY, Getz G, Jaffe DB, O'Kelly MJT, Zhao X, Carter SL, Russ C, Nusbaum C, Meyerson M, Lander ES. 2008. High-resolution mapping of copy-number alterations with massively parallel sequencing. *Nat Methods* **6**: 99–103.

Constantino JN, Gruber CP. *Social Responsiveness Scale*. Western Psychological Services, Los Angeles
http://portal.wpspublish.com/portal/page?_pageid=53,70492&_dad=portal&_schema=PORTAL.

Constantino JN, Todorov A, Hilton C, Law P, Zhang Y, Molloy E, Fitzgerald R, Geschwind D. 2013. Autism recurrence in half siblings: strong support for genetic mechanisms of transmission in ASD. *Mol Psychiatry* **18**: 137–138.

Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C, Williams C, Stalker H, Hamid R, Hannig V, et al. 2011. A copy number variation morbidity map of developmental delay. *Nat Genet* **43**: 838–846.

Courcet J-B, Faivre L, Malzac P, Masurel-Paulet A, Lopez E, Callier P, Lambert L, Lemesle M, Thevenon J, Gigot N, et al. 2012. The DYRK1A gene is a cause of syndromic intellectual disability with severe microcephaly and epilepsy. *J Med Genet* **49**: 731–736.

Cukier HN, Dueker ND, Slifer SH, Lee JM, Whitehead PL, Lalanne E, Leyva N, Konidari I, Gentry RC, Hulme WF, et al. 2014. Exome sequencing of extended families with autism reveals genes shared across neurodevelopmental and neuropsychiatric disorders. *Mol Autism* **5**: 1.

Darnell JC, Van Driesche SJ, Zhang C, Hung KYS, Mele A, Fraser CE, Stone EF, Chen C, Fak JJ, Chi SW, et al. 2011. FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* **146**: 247–261.

Davidson J, Goin-Kochel RP, Green-Snyder LA, Hundley RJ, Warren Z, Peters SU. 2012. Expression of the Broad Autism Phenotype in Simplex Autism Families from the Simons Simplex Collection. *J Autism Dev Disord*.

de Ligt J, Willemsen MH, van Bon BWM, Kleefstra T, Yntema HG, Kroes T, Vulto-van Silfhout AT, Koolen DA, de Vries P, Gilissen C, et al. 2012. Diagnostic exome sequencing in persons with severe intellectual disability. *N Engl J Med* **367**: 1921–1929.

de Vries BBA, Pfundt R, Leisink M, Koolen DA, Vissers LELM, Janssen IM, Reijmersdal SV, Nillesen WM, Huys EHLPG, Leeuw N de, et al. 2005. Diagnostic genome profiling in mental retardation. *Am J Hum Genet* **77**: 606–616.

Durand CM, Betancur C, Boeckers TM, Bockmann J, Chaste P, Fauchereau F, Nygren G, Rastam M, Gillberg IC, Anckarsäter H, et al. 2007. Mutations in the gene encoding

the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. *Nat Genet* **39**: 25–27.

- Endele S, Rosenberger G, Geider K, Popp B, Tamer C, Stefanova I, Milh M, Kortüm F, Fritsch A, Pientka FK, et al. 2010. Mutations in GRIN2A and GRIN2B encoding regulatory subunits of NMDA receptors cause variable neurodevelopmental phenotypes. *Nat Genet* **42**: 1021–1026.
- Fairless R, Masius H, Rohlmann A, Heupel K, Ahmad M, Reissner C, Dresbach T, Missler M. 2008. Polarized Targeting of Neurexins to Synapses Is Regulated by their C-Terminal Sequences. *J Neurosci* **28**: 12969–12981.
- Fischbach GD, Lord C. 2010. The Simons Simplex Collection: A Resource for Identification of Autism Genetic Risk Factors. *Neuron* **68**: 192–195.
- Fotaki V, Dierssen M, Alcántara S, Martínez S, Martí E, Casas C, Visa J, Soriano E, Estivill X, Arbonés ML. 2002. Dyrk1A haploinsufficiency affects viability and causes developmental delay and abnormal brain morphology in mice. *Mol Cell Biol* **22**: 6636–6647.
- Fromer M, Moran JL, Chambert K, Banks E, Bergen SE, Ruderfer DM, Handsaker RE, McCarroll SA, O'Donovan MC, Owen MJ, et al. 2012. Discovery and statistical genotyping of copy-number variation from whole-exome sequencing depth. *Am J Hum Genet* **91**: 597–607.
- Fu Y-H, Kuhl DPA, Pizzuti A, Pieretti M, Sutcliffe JS, Richards S, Verkert AJMH, Holden JJA, Fenwick RG Jr., Warren ST, et al. 1991. Variation of the CGG repeat at the fragile X site results in genetic instability: Resolution of the Sherman paradox. *Cell Reports* **67**: 1047–1058.
- Garber M, Grabherr MG, Guttman M, Trapnell C. 2011. Computational methods for transcriptome annotation and quantification using RNA-seq. *Nat Methods* **8**: 469–477.
- Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. *arXiv q-bio.GN*.
- Gilman SR, Iossifov I, Levy D, Ronemus M, Wigler M, Vitkup D. 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.
- Girirajan S, Campbell CD, Eichler EE. 2010. Human Copy Number Variation and Complex Genetic Disease. *Annu Rev Genet*.
- Girirajan S, Rosenfeld JA, Coe BP, Parikh S, Friedman N, Goldstein A, Filipink RA, McConnell JS, Angle B, Meschino WS, et al. 2012. Phenotypic heterogeneity of genomic disorders and rare copy-number variants. *N Engl J Med* **367**: 1321–1331.

- Glessner JT, Wang K, Cai G, Korvatska O, Kim CE, Wood S, Zhang H, Estes A, Brune CW, Bradfield JP, et al. 2009. Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature* **459**: 569–573.
- Guedj F, Pereira PL, Najas S, Barallobre M-J, Chabert C, Souchet B, Sebric C, Verney C, Herault Y, Arbones M, et al. 2012. DYRK1A: A master regulatory protein controlling brain growth. *Neurobiol Dis* **46**: 190–203.
- Hach F, Hormozdiari F, Alkan C, Hormozdiari F, Birol I, Eichler EE, Sahinalp SC. 2010. mrsFAST: a cache-oblivious algorithm for short-read mapping. *Nat Methods* **7**: 576–577.
- Hallmayer J, Cleveland S, Torres A, Phillips J, Cohen B, Torigoe T, Miller J, Fedele A, Collins J, Smith K, et al. 2011. Genetic heritability and shared environmental factors among twin pairs with autism. *Arch Gen Psychiatry* **68**: 1095–1102.
- Hamdan FF, Gauthier J, Spiegelman D, Noreau A, Yang Y, Pellerin S, Dobrzeniecka S, Côté M, Perreault-Linck E, Perreault-Linck E, et al. 2009. Mutations in SYNGAP1 in autosomal nonsyndromic mental retardation. *N Engl J Med* **360**: 599–605.
- Håvik B, Le Hellard S, Rietschel M, Lybæk H, Djurovic S, Mattheisen M, Mühleisen TW, Degenhardt F, Priebe L, Maier W, et al. 2011. The complement control-related genes CSMD1 and CSMD2 associate to schizophrenia. *Biol Psychiatry* **70**: 35–42.
- Helbig I, Mefford HC, Sharp AJ, Guipponi M, Fichera M, Franke A, Muhle H, de Kovel C, Baker C, Spiczak von S, et al. 2009. 15q13.3 microdeletions increase risk of idiopathic generalized epilepsy. *Nat Genet* **41**: 160–162.
- Helsmoortel, C., Vulto-van Silfhout, A. T., Coe, B. P., Vandeweyer, G., Rooms, L., van den Ende, J., et al. (2014). A SWI/SNF-related autism syndrome caused by de novo mutations in ADNP. *Nature Genet*, **46**: 380–384.
- Hoischen A, van Bon BWM, Gilissen C, Arts P, van Lier B, Steehouwer M, de Vries P, de Reuver R, Wieskamp N, Mortier G, et al. 2010. De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. *Nat Genet* **42**: 483–485.
- Holder JL Jr., Lotze TE, Bacino C, Cheung SW. 2012. A child with an inherited 0.31 Mb microdeletion of chromosome 14q32.33: Further delineation of a critical region for the 14q32 deletion syndrome. *Am J Med Genet* **158A**: 1962–1966.
- Hormozdiari F, Alkan C, Eichler EE, Sahinalp SC. 2009. Combinatorial algorithms for structural variation detection in high-throughput sequenced genomes. *Genome Research* **19**: 1270–1278.
- Hsueh Y-P, Wang T-F, Yang F-C, Sheng M. 2000. Nuclear translocation and transcription regulation by the membrane-associated guanylate kinase CASK/LIN-2. *Nature* **404**: 298–302.

- Huang Z. 2005. The origin recognition core complex regulates dendrite and spine development in postmitotic neurons. *The Journal of Cell Biology* **170**: 527–535.
- International Schizophrenia Consortium. 2008. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* **455**: 237–241.
- Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J, Yamrom B, Lee Y-H, Narzisi G, Leotta A, et al. 2012. De novo gene disruptions in children on the autistic spectrum. *Neuron* **74**: 285–299.
- Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, Doerks T, Julien P, Roth A, Simonovic M, et al. 2009. STRING 8--a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Research* **37**: D412–6.
- Kamiya K, Kaneda M, Sugawara T, Mazaki E, Okamura N, Montal M, Makita N, Tanaka M, Fukushima K, Fujiwara T, et al. 2004. A nonsense mutation of the sodium channel gene SCN2A in a patient with intractable epilepsy and mental decline. *J Neurosci* **24**: 2690–2698.
- Karakoc E, Alkan C, O'Roak BJ, Dennis MY, Vives L, Mark K, Rieder MJ, Nickerson DA, Eichler EE. 2011. Detection of structural variants and indels within exome data. *Nat Methods* **9**: 176–178.
- Kidd JM, Cooper GM, Donahue WF, Hayden HS, Samps N, Graves T, Hansen N, Teague B, Alkan C, Antonacci F, et al. 2008. Mapping and sequencing of structural variation from eight human genomes. *Nature* **453**: 56–64.
- Kim H-G, Kishikawa S, Higgins AW, Seong I-S, Donovan DJ, Shen Y, Lally E, Weiss LA, Najm J, Kutsche K, et al. 2008. Disruption of Neurexin 1 Associated with Autism Spectrum Disorder. *The American Journal of Human Genetics* **82**: 199–207.
- Klei L, Sanders SJ, Murtha MT, Hus V, Lowe JK, Willsey AJ, Moreno-De-Luca D, Yu TW, Fombonne E, Geschwind D, et al. 2012. Common genetic variants, acting additively, are a major source of risk for autism. *Mol Autism* **3**: 9.
- Koolen DA, Kramer JM, Neveling K, Nillesen WM, Moore-Barton HL, Elmslie FV, Toutain A, Amiel J, Malan V, Tsai AC-H, et al. 2012. Mutations in the chromatin modifier gene KANSL1 cause the 17q21.31 microdeletion syndrome. *Nat Genet* **44**: 639–641.
- Korbel JO, Abyzov A, Mu XJ, Carriero N, Cayting P, Zhang Z, Snyder M, Gerstein MB. 2009. PEMer: a computational framework with simulation-based error models for inferring genomic structural variants from massive paired-end sequencing data. *Genome Biology* **10**: R23.
- Korbel JO, Urban AE, Gruber F, Du J, Royce TE, Starr P, Zhong G, Emanuel B, Weissman SM, Snyder M, et al. 2007. Systematic prediction and validation of breakpoints associated with copy-number variants in the human genome.

- Krumm N, O'Roak BJ, Karakoc E, Mohajeri K, Nelson B, Vives L, Jacquemont S, Munson J, Bernier R, Eichler EE. 2013. Transmission Disequilibrium of Small CNVs in Simplex Autism. *Am J Hum Genet* **93**: 595–606.
- Krumm N, O'Roak BJ, Shendure J, Eichler EE. 2014. A de novo convergence of autism genetics and molecular neuroscience. *Trends Neurosci* **37**: 95–105.
- Krumm N, Sudmant PH, Ko A, O'Roak BJ, Malig M, Coe BP, NHLBI Exome Sequencing Project, Quinlan AR, Nickerson DA, Eichler EE. 2012. Copy number variation detection and genotyping from exome sequence data. *Genome Research* **22**: 1525–1532.
- Kumar RA, Marshall CR, Badner JA, Babatz TD, Mukamel Z, Aldinger KA, Sudi J, Brune CW, Goh G, KaraMohamed S, et al. 2009. Association and Mutation Analyses of 16p11.2 Autism Candidate Genes. *PLoS ONE* **4**: e4582.
- Lepagnol-Bestel A-M, Zvara A, Maussion G, Quignon F, Ngimbous B, Ramoz N, Imbeaud S, Loe-Mie Y, Benihoud K, Agier N, et al. 2009. DYRK1A interacts with the REST/NRSF-SWI/SNF chromatin remodelling complex to deregulate gene clusters involved in the neuronal phenotypic traits of Down syndrome. *Human Molecular Genetics* **18**: 1405–1414.
- Levy D, Ronemus M, Yamrom B, Lee Y-H, Leotta A, Kendall J, Marks S, Lakshmi B, Pai D, Ye K, et al. 2011. Rare De Novo and Transmitted Copy-Number Variation in Autistic Spectrum Disorders. *Neuron* **70**: 886–897.
- Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
- Lichtenstein P, Carlstrom E, Rastam M, Gillberg C, Anckarsater H. 2010. The Genetics of Autism Spectrum Disorders and Related Neuropsychiatric Disorders in Childhood. *American Journal of Psychiatry* **167**: 1357–1363.
- Litterman N, Ikeuchi Y, Gallardo G, O'Connell BC, Sowa ME, Gygi SP, Harper JW, Bonni A. 2011. An OBSL1-Cul7Fbxw8 ubiquitin ligase signaling mechanism regulates Golgi morphology and dendrite patterning. ed. P. Scheiffele. *PLoS Biol* **9**: e1001060.
- Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, Hasz R, Walters G, Garcia F, Young N, et al. 2013. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**: 580–585.
- Losh M, Childress D, Lam K, Piven J. 2008. Defining key features of the broad autism phenotype: a comparison across parents of multiple- and single-incidence autism families. *Am J Med Genet B Neuropsychiatr Genet* **147B**: 424–433.

- Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J, Shago M, Moessner R, Pinto D, Ren Y, et al. 2008. Structural variation of chromosomes in autism spectrum disorder. *Am J Hum Genet* **82**: 477–488.
- Morita, M., Ler, L. W., Fabian, M. R., Siddiqui, N., Mullin, M., Henderson, V. C., et al. (2012). A novel 4EHP-GIGYF2 translational repressor complex is essential for mammalian development. *Molecular and Cellular Biology*, 32(17), 3585–3593. doi:10.1128/MCB.00455-12.
- Matsuura T, Sutcliffe JS, Fang P, Galjaard RJ, Jiang YH, Benton CS, Rommens JM, Beaudet AL. 1997. De novo truncating mutations in E6-AP ubiquitin-protein ligase gene (UBE3A) in Angelman syndrome. *Nat Genet* **15**: 74–77.
- Mazur-Kolecka B, Golabek A, Kida E, Rabe A, Hwang Y-W, Adayev T, Wegiel J, Flory M, Kaczmarek W, Marchi E, et al. 2012. Effect of DYRK1A activity inhibition on development of neuronal progenitors isolated from Ts65Dn mice. *J Neurosci Res* **90**: 999–1010.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20**: 1297–1303.
- Miller DT, Shen Y, Weiss LA, Korn J, Anselm I, Bridgemohan C, Cox GF, Dickinson H, Gentile J, Harris DJ, et al. 2009. Microdeletion/duplication at 15q13.2q13.3 among individuals with features of autism and other neuropsychiatric disorders. *J Med Genet* **46**: 242–248.
- Moessner R, Marshall CR, Sutcliffe JS, Skaug J, Pinto D, Vincent J, Zwaigenbaum L, Fernandez B, Roberts W, Szatmari P, et al. 2007. Contribution of SHANK3 mutations to autism spectrum disorder. *Am J Hum Genet* **81**: 1289–1297.
- Moller RS, Kübart S, Hoeltzenbein M, Heye B, Vogel I, Hansen CP, Menzel C, Ullmann R, Tommerup N, Ropers H-H, et al. 2008. Truncation of the Down syndrome candidate gene DYRK1A in two unrelated patients with microcephaly. *Am J Hum Genet* **82**: 1165–1170.
- Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5**: 621–628.
- Mutz K-O, Heilkenbrinker A, Lönne M, Walter J-G, Stahl F. 2013. Transcriptome analysis using next-generation sequencing. *Curr Opin Biotechnol* **24**: 22–30.
- Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A, Lin C-F, Stevens C, Wang L-S, Makarov V, et al. 2012. Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* **485**: 242–245.
- Ng D, Pitcher GM, Szilard RK, Sertié A, Kanisek M, Clapcote SJ, Lipina T, Kalia LV,

- Joo D, McKerlie C, et al. 2009a. Neto1 Is a Novel CUB-Domain NMDA Receptor-Interacting Protein Required for Synaptic Plasticity and Learning. *PLoS Biol* **7**: e1000041.
- Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, et al. 2009b. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet* **42**: 30–35.
- Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE, et al. 2009c. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **461**: 272–276.
- Nishiyama M, Oshikawa K, Tsukada Y-I, Nakagawa T, Iemura S-I, Natsume T, Fan Y, Kikuchi A, Skoultschi AI, Nakayama KI. 2009. CHD8 suppresses p53-mediated apoptosis through histone H1 recruitment during early embryogenesis. *Nat Cell Biol* **11**: 172–182.
- Nishiyama M, Skoultschi AI, Nakayama KI. 2012. Histone H1 recruitment by CHD8 is essential for suppression of the Wnt- β -catenin signaling pathway. *Mol Cell Biol* **32**: 501–512.
- O’Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, Girirajan S, Karakoc E, Mackenzie AP, Ng SB, Baker C, et al. 2011. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nat Genet* **43**: 585–589.
- O’Roak BJ, Vives L, Fu W, Egertson JD, Stanaway IB, Phelps IG, Carvill G, Kumar A, Lee C, Ankenman K, et al. 2012a. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* **338**: 1619–1622.
- O’Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, Levy R, Ko A, Lee C, Smith JD, et al. 2012b. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* **485**: 246–250.
- Ogiwara I, Ito K, Sawaishi Y, Osaka H, Mazaki E, Inoue I, Montal M, Hashikawa T, Shike T, Fujiwara T, et al. 2009. De novo mutations of voltage-gated sodium channel α II gene SCN2A in intractable epilepsies. *Neurology* **73**: 1046–1053.
- Paşca SP, Portmann T, Voineagu I, Yazawa M, Shcheglovitov A, Pasca AM, Cord B, Palmer TD, Chikahisa S, Nishino S, et al. 2011. Using iPSC-derived neurons to uncover cellular phenotypes associated with Timothy syndrome. *Nat Med* **17**: 1657–1662.
- Peiffer DA, Le JM, Steemers FJ, Chang W, Jenniges T, Garcia F, Haden K, Li J, Shaw CA, Belmont J, et al. 2006. High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. *Genome Research* **16**: 1136–1148.
- Lorenz, P., Dietmann, S., Wilhelm, T., Koczan, D., Autran, S., Gad, S., et al. (2010). The

ancient mammalian KRAB zinc finger gene cluster on human chromosome 8q24.3 illustrates principles of C2H2 zinc finger evolution associated with unique expression profiles in human tissues. *BMC Genomics*, *11*(1), 206. doi:10.1186/1471-2164-11-206

- Petrovski S, Wang Q, Heinzen EL, Allen AS, Goldstein DB. 2013. Genic Intolerance to Functional Variation and the Interpretation of Personal Genomes. *PLoS Genetics* **9**: e1003709.
- Pinkel D, Segraves R, Sudar D, Clark S, Poole I, Kowbel D, Collins C, Kuo WL, Chen C, Zhai Y, et al. 1998. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet* **20**: 207–211.
- Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R, Conroy J, Magalhaes TR, Correia C, Abrahams BS, et al. 2010. Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* **466**: 368–372.
- Poultney CS, Goldberg AP, Drapeau E, Kou Y, Harony-Nicolas H, Kajiwarra Y, De Rubeis S, Durand S, Stevens C, Rehnström K, et al. 2013. Identification of Small Exonic CNV from Whole-Exome Sequence Data and Application to Autism Spectrum Disorder. *The American Journal of Human Genetics* **93**: 607–619.
- Rauch A, Wieczorek D, Graf E, Wieland T, Ende S, Schwarzmayer T, Albrecht B, Bartholdi D, Beygo J, Di Donato N, et al. 2012. Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet* **380**: 1674–1682.
- Ronemus M, Iossifov I, Levy D, Wigler M. 2014. The role of de novo mutations in the genetics of autism spectrum disorders. *Nature Reviews Genetics* **15**: 133–141.
- Rowntree RK, Harris A. 2003. The Phenotypic Consequences of CFTR Mutations. *Ann Human Genet* **67**: 471–485.
- Salinas PC, Zou Y. 2008. Wnt Signaling in Neural Circuit Assembly. *Annu Rev Neurosci* **31**: 339–358.
- Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, Moreno-De-Luca D, Chu SH, Moreau MP, Gupta AR, Thomson SA, et al. 2011. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron* **70**: 863–885.
- Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, Ercan-Sencicek AG, DiLullo NM, Parikshak NN, Stein JL, et al. 2012. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485**: 237–241.
- Santen GWE, Aten E, Sun Y, Almomani R, Gilissen C, Nielsen M, Kant SG, Snoeck IN, Peeters EAJ, Hilhorst-Hofstee Y, et al. 2012. Mutations in SWI/SNF chromatin

remodeling complex gene ARID1B cause Coffin-Siris syndrome. *Nat Genet* **44**: 379–380.

- Santos Dos C, Essioux L, Teinturier C, Tauber M, Goffin V, Bougnères P. 2004. A common polymorphism of the growth hormone receptor is associated with increased responsiveness to growth hormone. *Nat Genet* **36**: 720–724.
- Sathirapongsasuti JF, Lee H, Horst BAJ, Brunner G, Cochran AJ, Binder S, Quackenbush J, Nelson SF. 2011. Exome sequencing-based copy-number variation and loss of heterozygosity detection: ExomeCNV. *Bioinformatics* **27**: 2648–2654.
- Scharpf RB, Irizarry RA, Ritchie ME, Carvalho B, Ruczinski I. 2011. Using the R Package crlmm for Genotyping and Copy Number Estimation. *J Stat Softw* **40**: 1–32.
- Scholz R, Berberich S, Rathgeber L, Kollek A, Köhr G, Kornau H-C. 2010. AMPA Receptor Signaling through BRAG2 and Arf6 Critical for Long-Term Synaptic Depression. *Neuron* **66**: 768–780.
- Schuurs-Hoeijmakers JHM, Geraghty MT, Kamsteeg E-J, Ben-Salem S, de Bot ST, Nijhof B, van de Vondervoort IIGM, van der Graaf M, Nobau AC, Otte-Höller I, et al. 2012. Mutations in DDHD2, Encoding an Intracellular Phospholipase A1, Cause a Recessive Form of Complex Hereditary Spastic Paraplegia. *The American Journal of Human Genetics* **91**: 1073–1081.
- Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, Yamrom B, Yoon S, Krasnitz A, Kendall J, et al. 2007. Strong Association of De Novo Copy Number Mutations with Autism. *Science* **316**: 445–449.
- Sharp AJ, Hansen S, Selzer RR, Cheng Z, Regan R, Hurst JA, Stewart H, Price SM, Blair E, Hennekam RC, et al. 2006. Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. *Nat Genet* **38**: 1038–1042.
- Sharp AJ, Mefford HC, Li K, Baker C, Skinner C, Stevenson RE, Schroer RJ, Novara F, De Gregori M, Ciccone R, et al. 2008. A recurrent 15q13.3 microdeletion syndrome associated with mental retardation and seizures. *Nat Genet* **40**: 322–328.
- Song WJ, Sternberg LR, Kasten-Sportès C, Keuren ML, Chung SH, Slack AC, Miller DE, Glover TW, Chiang PW, Lou L, et al. 1996. Isolation of human and murine homologues of the Drosophila minibrain gene: human homologue maps to 21q22.2 in the Down syndrome "critical region". *Genomics* **38**: 331–339.
- Stefansson H, Rujescu D, Cichon S, Pietiläinen OPH, Ingason A, Steinberg S, Fossdal R, Sigurdsson E, Sigmundsson T, Buizer-Voskamp JE, et al. 2008. Large recurrent microdeletions associated with schizophrenia. *Nature* **455**: 232–236.
- Steffenburg S, Gillberg C, Hellgren L, Andersson L, Gillberg IC, Jakobsson G, Bohman M. 1989. A twin study of autism in Denmark, Finland, Iceland, Norway and Sweden.

J Child Psychol Psychiatry **30**: 405–416.

Stessman HA, Bernier R, Eichler EE. 2014. A genotype-first approach to defining the subtypes of a complex disease. *Cell* **156**: 872–877.

Su AI. 2004. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proceedings of the National Academy of Sciences* **101**: 6062–6067.

Sudmant PH, Kitzman JO, Antonacci F, Alkan C, Malig M, Tsalenko A, Sampas N, Bruhn L, Shendure J, 1000 Genomes Project, et al. 2010. Diversity of human copy number variation and multicopy genes. *Science* **330**: 641–646.

Talkowski ME, Rosenfeld JA, Blumenthal I, Pillalamarri V, Chiang C, Heilbut A, Ernst C, Hanscom C, Rossin E, Lindgren AM, et al. 2012. Sequencing chromosomal abnormalities reveals neurodevelopmental loci that confer risk across diagnostic boundaries. *Cell* **149**: 525–537.

Tejedor F, Zhu XR, Kaltenbach E, Ackermann A, Baumann A, Canal I, Heisenberg M, Fischbach KF, Pongs O. 1995. minibrain: a new protein kinase family involved in postembryonic neurogenesis in *Drosophila*. *Neuron* **14**: 287–301.

Thompson BA, Tremblay V, Lin G, Bochar DA. 2008. CHD8 is an ATP-dependent chromatin remodeling factor that regulates beta-catenin target genes. *Mol Cell Biol* **28**: 3894–3904.

van Bon BWM, Hoischen A, Hehir-Kwa J, de Brouwer APM, Ruivenkamp C, Gijsbers ACJ, Marcelis CL, de Leeuw N, Veltman JA, Brunner HG, et al. 2011. Intragenic deletion in *DYRK1A* leads to mental retardation and primary microcephaly. *Clin Genet* **79**: 296–299.

Venkatraman ES, Olshen AB. 2007. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics* **23**: 657–663.

Vissers LELM, de Ligt J, Gilissen C, Janssen I, Steehouwer M, de Vries P, van Lier B, Arts P, Wieskamp N, del Rosario M, et al. 2010. A de novo paradigm for mental retardation. *Nat Genet* **42**: 1109–1112.

Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, Horvath S, Mill J, Cantor RM, Blencowe BJ, Geschwind DH. 2011. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* **474**: 380–384.

Wada K, Saigoh K, Wang Y-L, Suh J-G, Yamanishi T, Sakai Y, Kiyosawa H, Harada T, Ichihara N, Wakana S, et al. 1999. Intragenic deletion in the gene encoding ubiquitin carboxy-terminal hydrolase in *gad* mice. *Nat Genet* **23**: 47–51.

Wali A, Ali G, John P, Lee K, Chishti MS, Leal SM, Ahmad W. 2007. Mapping of a Gene for Alopecia with Mental Retardation Syndrome (APMR3) on Chromosome 18q11.2-q12.2. *Ann Human Genet* **71**: 570–577.

- Walsh CA, Morrow EM, Rubenstein JLR. 2008. Autism and Brain Development. *Cell* **135**: 396–400.
- Walsh KM, Bracken MB. 2011. Copy number variation in the dosage-sensitive 16p11.2 interval accounts for only a small proportion of autism incidence: A systematic review and meta-analysis. *Genet Med* **13**: 377–384.
- Wang ET, Sandberg R, Luo S, Khrebtkova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB. 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**: 470–476.
- Wang T-F, Ding C-N, Wang G-S, Luo S-C, Lin Y-L, Ruan Y, Hevner R, Rubenstein JLR, Hsueh Y-P. 2004. Identification of Tbr-1/CASK complex target genes in neurons. *J Neurochem* **91**: 1483–1492.
- Wobst H, Förster S, Laurini C, Sekulla A, Dreiseidler M, Höhfeld J, Schmitz B, Diestel S. 2012. UCHL1 regulates ubiquitination and recycling of the neural cell adhesion molecule NCAM. *FEBS J* **279**: 4398–4409.
- Yabut O, Domogauer J, D'Arcangelo G. Dyrk1A Overexpression Inhibits Proliferation and Induces Premature Neuronal Differentiation of Neural Progenitor Cells. *jneurosci.org*.
- Yazawa M, Hsueh B, Jia X, Pasca AM, Bernstein JA, Hallmayer J, Dolmetsch RE. 2011. Using induced pluripotent stem cells to investigate cardiac phenotypes in Timothy syndrome. *Nature* **471**: 230–234.
- Yoon S, Xuan Z, Makarov V, Ye K, Sebat J. 2009. Sensitive and accurate detection of copy number variants using read depth of coverage. *Genome Research* **19**: 1586–1592.

Web and Software Resources

CoNIFER (Copy Number Inference from Exome Reads): source code, tutorials and sample can be downloaded from <http://conifer.sf.net> and additional pipeline implementations are available at <https://github.com/nkrumm/conifer-tools>

MIM (Mendelian Inheritance in Man) identifiers can be accessed via OMIM (Online Mendelian Inheritance in Man): <http://omim.org>

mrsFAST: source code and binaries are available at <http://mrsfast.sf.net>

NDAR (National Database of Autism Research): <http://ndar.nih.gov> Data from Chapter 4 available at: <https://ndar.nih.gov/study.html?id=334>

Appendix A (Chapter 1)

Details of StringDB network generation:

In order to create the PPI network in Figure 3, we started with the *de novo* mutations published in each of the six exome studies [1-6] and limited these to events found in probands and intersecting exons or canonical splice sites. The network in Figure 3 was created using all genes with *de novo* truncating variants (defined as nonsense variants, frameshifting variants or variants likely affecting mRNA splicing) as well as six additional genes (*DLG4*, *GRIN2A*, *CASK*, *PSEN1*, *CHD7*, *NLGN1*) in which only missense variants have been observed thus far, but which have important neurobiological roles and/or disease association. In all, we included 158 genes, of which 157 could be identified in the StringDB database (*LTN1* was not found in any human interactions in StringDB).

Data from the StringDB interaction database version 9.05 [7] was used to create the edges of the PPI. We strictly limited our interactions to only human (organism ID 9606) interactions which were based on experimental evidence and an overall combined interaction confidence score of 400 or more. We did *not* include interactions solely based on any of the other StringDB interaction types, such as *in silico* text-mining, co-expression, etc. Overall, we included 85,678 interactions and 12,113 nodes in our analysis.

In order to create the network displayed in Figure 3, we took these steps:

1. Intersected the 157 identified genes based on the criteria above with the human, experimentally-validated StringDB interactions. These form the central red (truncating mutations) and blue nodes (selected missense mutations) in Figure 3, and are connected using thick black lines.
2. Found the two largest connected components. We observed that these were connected via the DLGAP1 protein (see main text for discussion) and added

this node as a unfilled (white) node with dashed lines.

3. In addition, we surmised that our set of truncating mutations was likely incomplete, and that many ASD/ID genes may be excluded from the central network simply due to the fact that rare variants in these genes have not yet been discovered. Thus, we “grew” or expanded the network by allowing genes with truncating mutations to be included as “peripheral” nodes if they were within a distance of two (i.e., one intervening node) of the central network. These nodes are drawn as a lighter shade of red and have finely dashed edges. For this analysis, we excluded three proteins (SUMO1, SUMO2 and UBC) which had highly non-specific interactions in StringDB (sumosylation and ubiquitination).
4. We indicated which mutations have only been observed in studies of ID by using half-filled circles. The reciprocal (ASD-only) situation is not indicated due to the fact that there have been nearly ten-fold more ASD exomes sequenced than ID exomes.
5. Lastly, we scaled the sizes of the nodes based the number of times mutations in cases had been observed in each gene (including the mutations from the MIP resequencing data).

Estimating PPI significance:

In order to test if the PPI network of de novo mutations found in the six reviewed exome studies was significantly distinct from randomly formed networks of similar size, we performed two simulation studies. These two simulations were based on random sampling from the complete set of known PPI interactions (i.e., from StringDB) or from random permutation of the existing network. Both simulations were designed to take into account the highly variable degree distribution of interaction networks-- that is, some nodes are highly connected “hub nodes” while other proteins are scantily connected, if at all. The results of the simulations are described in Table S1, and each is described in more detail below.

Stratified node resampling:

For each iteration of the simulation, we randomly selected a stratified (based on degree distribution for the nodes with mutations) set of nodes from the complete StringDB interaction network (limited to interactions with “experimental” evidence and a minimum interaction score of 400). This ensured that the nodes we picked were similar in connectivity and that representation of “hub” nodes and “outlier” nodes was equivalent to that of the actual network. A new PPI graph was generated from each set of stratified random nodes, and the structural characteristics of these graphs were compiled into a null distribution. We primarily examined the average clustering coefficient and the total number of edges of the permuted graphs and compared these to the characteristics of the actual PPI networks. P-values were derived using the empirical distributions from 10,000 iterations of the simulation.

Edge swapping simulation:

In this simulation, we did not alter the set of nodes included in each PPI network, but instead permuted the edges found within the PPI network, thus preserving the degree distribution of the network. Specifically, in each iteration of the simulation, a random sampling of edges (where the number of sampled edges was equal to the total number of edges in the PPI network) in the network were swapped with another eligible edge:



After randomly swapping edges, we re-computed the average clustering coefficient, size of largest connected component and number of edges for the subgraph of the genes (nodes) with observed mutations and computed the empirical p-value as above. Due to the increased complexity and running time of this simulation, we performed only 1,000 iterations.

Table S1: Summary table of PPI network simulations

	Probands				Siblings			
	# of nodes	# Edges	Average Clust. Coeff.	Largest Conn. Comp.	# of nodes	# Edges	Average Clust. Coeff.	Largest Conn. Comp.
<i>Missense + Truncating</i>	775	$p < 10e-5$	$p = 0.009$	$p < 10e-5$	355	$p = 0.10$	$p = 0.36$	$p = 0.10$
		$p < 10e-4$	$p < 10e-4$	$p = 0.02$		$p = 0.13$	$p = 0.052$	$p = 0.36$
<i>Missense only</i>	640	$p < 10e-5$	$p = 0.002$	$p < 10e-5$	317	$p = 0.06$	$p = 0.26$	$p = 0.07$
		$p < 10e-4$	$p < 10e-4$	$p = 0.006$		$p = 0.11$	$p = 0.048$	$p = 0.27$
<i>Synon. only</i>	254	$p = 0.062$	$p = 0.69$	$p = 0.04$	133	$p = 0.28$	$p = 0.52$	$p = 0.02$
		$p = 0.08$	$p = 0.52$	$p = 0.07$		$p = 0.315$	$p = 0.5$	$p = 0.07$
<i>Truncating only</i>	151	$p = 0.59$	$p = 0.56$	$p = 0.89$	39	$p = 0.64$	$p = 0.50$	$p = 0.65$
		$p = 0.53$	$p = 0.51$	$p = 0.86$		$p = 0.61$	$p = 0.50$	$p = 0.61$
<i>Missense + Trunc. (ASD Only)</i>	667	$p < 10e-5$	$p = 0.003$	$p = 0.0002$	346	$p = 0.31$	$p = 0.48$	$p = 0.74$
		$p < 10e-4$	$p < 10e-4$	$p = 0.035$		$p = 0.39$	$p = 0.1$	$p = 0.9$

Top row p-values are from stratified node resampling simulation

Bottom row p-values are from the edge-swap simulation

Nominally significant ($p < 0.05$) values highlighted in **bold**

Details of Hidden Species simulation in Figure 1:

In order to estimate the number of genes implicated in ASD under a *de novo*/rare variant model, we used mutations in probands from the four ASD exome studies and a reformulation of the “unseen species problem” (see [8] for review; [9] for application to *de novo* CNVs discovered in autism), where genes with severe *de novo* SNPs in probands are considered “observed species”, and binned by their frequency of appearance (i.e., “singletons”, “doubletons”, etc.). For each category (truncating, truncating+missense), we find the distribution of the number of recurrently mutated genes (i.e., the bins and bin counts of a histogram function). All genes with more than one mutation are included, as is a fraction of the “singleton” mutations (those with only one observed mutation in the four studies). The recurrence counts are shown below:

Table S2: Recurrence of de novo mutations in 4 ASD studies

<i>Recurrence in four exome studies</i>	Truncating only	Truncating + Missense	Truncating + Severe missense
6	–	1	1
5	–	1	0
4	1	3	1
3	1	7	4
2	7	59	26
1	137	946	490

Given these frequencies and frequency counts, we estimated the total number of genes implicated in autism (the total number of species) using the Chao and Lee estimator implemented in the R package SPECIES [10]. The “Percentage of de novo singleton events considered pathogenic” refers to the fraction of the singletons (recurrence = 1) included in the frequency counts.

References:

- 1 O'Roak, B.J. *et al.* (2012) Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* 485, 246–250
- 2 Sanders, S.J. *et al.* (2012) De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* 485, 237–241
- 3 Iossifov, I. *et al.* (2012) De novo gene disruptions in children on the autistic spectrum. *Neuron* 74, 285–299
- 4 Neale, B.M. *et al.* (2012) Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* 485, 242–245
- 5 Rauch, A. *et al.* (2012) Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet* 380, 1674–1682
- 6 de Ligt, J. *et al.* (2012) Diagnostic exome sequencing in persons with severe intellectual disability. *N Engl J Med* 367, 1921–1929
- 7 Franceschini, A. *et al.* (2012) STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Research* 41, D808–D815
- 8 BungeFitzpatrick (1993) Estimating the Number of Species: A Review. *Journal of the American Statistical Association* 88, 364–374
- 9 Sanders, S.J. *et al.* (2011) Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron* 70, 863–885
- 10 Wang, J.-P. 01-Apr-(2011), SPECIES: An R Package for Species Richness Estimation. *Journal of Statistical Software*. [Online]. Available: <http://www.jstatsoft.org/>. [Accessed: 30-Aug-2011]

Appendix B (Chapter 2)

Library construction and exome capture:

All exome samples were prepared by subjecting 2 ug of genomic DNA to a series of shotgun library construction steps, including fragmentation through acoustic sonication (Covaris), end-polishing and A-tailing, ligation of sequencing adaptors, and PCR amplification. Following library construction, 1 μ g of shotgun library is hybridized to biotinylated capture probes for 72 hours and then recovered via streptavidin beads. Unbound DNA is washed away, and the captured DNA is PCR amplified for sequencing.

Sequence data processing and alignment:

Raw sequenced reads (from FASTQ files) were first split into 36bp chunks (in order to avoid interference from indels), and mapped using the mrsFAST (v.2.3.0.2) aligner. Up to two mismatches were allowed per read. To reduce computational overhead, we created a concatenated exome index, consisting of the targeted exons (see below), plus 300bp flanking sequence from the hg19 (NCBI build 37) human reference genome, masked with RepeatMasker and Tandem Repeat Finder. After mapping to this concatenated “exome”, we translated mapped coordinates back to hg19 genome coordinates for further processing.

Exome probe definitions:

For the mrsFAST-based alignments, we developed a probe set (i.e., target regions) by intersecting target definitions of the Roche Nimblegen EZ Exome SeqCap Version 2 (from http://www.nimblegen.com/downloads/annotation/ez_exome_v2/SeqCapEZ_Exome_v2.0_Design_Annotation_files.zip) exome capture kit with RefSeq exons (excluding UTR regions). In addition, we included 4,857 non-exonic targeted regions from the SeqCap Version 2 target definition list. This resulted in 194,080

target probes (available at <http://conifer.sourceforge.net>)

Initial exon-level normalization:

We calculated RPKM values for the 194,080 target probes individually. The RPKM normalization is given by

$$\text{RPKM} = 10^9 * \text{Read Starts} / \text{Total Mapped Reads} * \text{Target Size (bp)}$$

where the number of Read Starts is defined as the number of reads starting within the target boundaries, and the Total Mapped Reads corresponds to the number of unique reads which had at least one mapping. This initial RPKM normalization step adjusts our read-depth estimates for target (exon) size as well as the overall sequencing coverage in the experiment. To reduce erroneous signal from failed or improperly targeted probes, we excluded 3,964 targets which had a median RPKM < 1 in the 533 ESP samples.

Next, to control for probe-to-probe differences in capture efficiency, we standardized the RPKM values using a z-transformation. The median and standard deviation of each exon were derived from RPKM values of the 533 ESP exomes. The formula for the zRPKM value is:

$$\text{zRPKM} = (\text{RPKM}_{\text{exon, sample}} - \text{Median}_{\text{exon}}) / \text{StdDev}_{\text{exon}}$$

Removing systematic bias between batches:

A previous analysis of exome read-depth values from ~1,700 ESP exomes using principal components analysis (PCA) revealed several strong components, some of which were attributed to “batch” effects (unpublished, Sara Ng and Jay Shendure). We hypothesized that these strong components do not correspond to biological signal, but rather to differences in capture protocol, efficiency and sequencing bias. Using singular value decomposition, a mathematical analog of

PCA, we decompose the exon-by-sample (X) data matrix into three matrices:

$$X = USV^T$$

In order to remove the strongest k components, we set $S_1 \dots S_k$ to zero to form S' , and then recalculate X as the dot product of U , S' and V^T . For computation efficiency, each chromosome is normalized individually across the population. We used an implementation of SVD in the `scipy.stats` package available for the python programming language.

Discovery of rare CNVs

For discovery of rare CNVs, we removed between 12 and 15 (k) singular values, a number which we empirically adjusted based on the inflection point of the “scree plot” (Fig S2), as well as by manual inspection of the final normalized data. To reduce the false positive rate of discovery for rare CNVs, we applied a 15-exon centrally-weighted moving average across exons. We set discovery thresholds at -1.5 or +1.5 for rare deletions and duplications, respectively, and required at least three exome probes to exceed the threshold. To account for the fact that smoothing shrinks the apparent size of discovered events, regions which exceeded this threshold were slightly expanded until the sample’s smoothed value crossed within two standard deviations surrounding the population mean of the smoothed values (Fig S1b).

Sample-level quality control:

We excluded ESP exomes from the final background distribution if our algorithm predicted more than 10 calls, as we noted that these samples had a greatly increased total call count (up to 111 calls/sample), and that the calls were largely false positives. This resulted in the exclusion of a total of 80 of 613 initial exomes (87% pass rate) ESP exomes from the background distribution, leaving our final set of 533 exomes. No exomes from the HapMap cohort (range: 1-7 calls per

individual) or the autism cohort (range: 0-14 calls per individual) were excluded.

Genotyping CNPs:

For genotyping copy number polymorphic (CNP) regions of the genome, as well as assessing the copy-number of multi-copy genes, we developed a slightly modified approach. Starting from zRPKM values, we again applied the SVD transformation, but opted to remove only five components, in order to prevent the SVD algorithm from remove *bona fide* signal from the regions of interest. We genotype each individual by determining the average, resulting in the “SVD-ZRPKM value”.

Whole Genome Copy Number Correlations:

To estimate the absolute copy number at CNP loci, read-depth from independent whole-genome sequencing (as previously described in (Sudmant et al., 2010)) was used. Briefly, regions of known copy-number were used to create a copy-number standard curve, and the absolute copy number of tiling 1kb windows across the genome was estimated. For genotyping, the median of the 1kb window estimates was used.

Because we wanted to assess a correlation between exome and whole-genome based methods, we only included loci in the final set if the whole-genome copy number estimate indicated that the locus was polymorphic among the seven HapMap samples tested. We defined a locus to be polymorphic if the absolute range of copy numbers amongst the HapMap samples was greater than 1. Finally, we defined the median copy number of each locus as the median of the absolute copy number estimates among the seven HapMap samples.

Absolute copy number estimation using population frequency information:

To convert relative SVD-ZRPKM values into absolute copy numbers, we used an unsupervised clustering algorithm to cluster SVD-ZRPKM genotype values, and

then leveraged genotypes from 43 CNPs in a large set of HapMap samples from (Campbell et al., 2011) to match clusters to absolute copy number.

Unsupervised clustering was done using a mean-shift algorithm implemented in the python package SciKits.learn. The mean-shift algorithm is similar to k-means clustering, but does not require *a priori* information regarding the number of clusters. After clustering, we automatically merged clusters together if their centers were not spaced linearly on the x-axis, as we found that this marginally improved the clustering for some loci. Finally, we fit the most common copy-number state(s) for each locus from (Campbell et al., 2011) to the largest cluster(s) identified by the exome-based SVD-ZRPKM values by maximizing the r^2 value between the two vectors (from each data source) of copy-number states. In other words, we attempted to match the frequencies of each copy number state identified by (Campbell et al., 2011) to consecutive clusters identified by our clustering method. To determine an absolute copy number genotype of a CNP locus for a HapMap sample, we simply determined to which cluster the sample belonged and the matched absolute copy number for that cluster.

Sensitivity call set for HapMap Samples:

To assess sensitivity, we started with CNV calls from the discovery experiment from Conrad and colleagues (Conrad et al., 2010) as a gold standard. This list contained at first 6919 calls for the 5 overlapping hapmap samples in our set. Of these, 486 overlapped at least 3 exome probes (required by our discovery algorithm). Because segmental duplications are prone to array-CGH reference and detection bias, we removed 416 calls for which 50% of the underlying exome probes were in segmental duplications. Finally, we removed 20 calls found in somatically rearranged regions:

chr6:32386993-32787910	HLA
chr6:31226231-31328167	HLA
chr14:105994256-107283087	Ig Heavy chain
chr22:22380820-23265082	Ig light chain lambda
chr7:141975722-142519580	T-cell receptor beta subunit

This resulted in 50 calls. For each call, we reviewed several data sources: 1) Illumina i1M or 650Y (for NA15510) SNP array LogR intensities and B-allele frequency, 2) whole genome copy number estimates (from (Sudmant et al., 2010), but not available for NA15510), 3) fosmid-based calls from (Kidd et al., 2008) and 4) SVD-ZRPKM signal across ESP and HapMap samples. We manually curated the 50 calls into four categories: Rare CNVs (5 total), CNPs or CNP-like (36 events), events in high-diversity regions of the genome (6 events; primarily Olfactory receptors and zinc-finger genes), and false positives in the Conrad et al. set (3 calls). False positives had no corroborating evidence in any other data set, and were not counted towards the sensitivity estimates.

Discovery of rare CNVs in ASD trios:

Using the input set of 366 ASD cohort individuals (122 probands) with 366 randomly picked ESP samples, and removing 15 components, our algorithm made a total of 1,043 calls among the 366 individuals in the ASD cohort (with 369 calls in probands), with each sample having between 0 and 14 calls; overall 340 individuals had at least one call. Merging all overlapping calls in the ASD resulted in 282 CNVRs.

As the exome capture reaction targets many genes present in duplicated regions of the genome, and as many exons share homologous sequence, a significant proportion of our calls in probands are due to changes in the copy number of these genes due to independent assortment of parental haplotypes. Starting with the 369 calls made in the 117 probands, we filtered calls to enrich for “rare”

CNVs. Calls which had greater than 50% reciprocal overlap (as determined by the fraction of underlying exome probes within the call also in segmental duplications) with segmental duplications were removed (163/369, or 44%). Next, we calculated the median copy number of calls based on whole-genome read-depth copy-number estimates from ~660 genomes (Sudmant et al., 2010), and additionally filtered 13 calls (3.5%) with more than 3+ copies population-wide (as events stemming from these segmentally-duplicated or higher-copy regions of the genome are likely due to the independent assortment of parental haplotypes, and not “true” rare CNVs). Additionally, we manually curated the calls to remove calls within regions undergoing somatic rearrangement (two calls; one at the *IGH* locus and one in the *HLA* locus), and merged adjacent or overlapping calls. These steps left 191 calls among 97 probands, and these calls were primarily found in non-duplicated genes and diploid regions of the genome. We categorized each call into one of three bins: de novo, inherited or copy-number polymorphic (Table S3).

Description of CNVs in ASD trios:

We found eight putative *de novo* events (Table S2). For six of these, we were able to corroborate the event using available Illumina SNP microarray data as well as targeted array-CGH experimental data (Sanders et al. 2011, O’Roak et al., submitted). The other two de novo events were each driven by increases in SVD-ZRPKM values for the eighth exon of *FAF2* gene. Although we were able to confirm an excess of reads mapping to this exon by manual inspection of the mapped reads (from both mrsFAST and BWA alignments), we were not able to experimentally validate these duplications using a quantitative real-time PCR assay targeting the eighth exon itself (data not shown).

Next, we looked for inherited events using our exome read-depth analysis and found that 128 events in the probands were inherited from either the probands mother or father. For 117 of these events, the SVD-RPKM values of both the

proband and the parent exceeded the detection threshold (± 1.5); however, for 11 of these calls, the SVD-RPKM values between proband and parents was just below the deletion or duplication threshold required for calling, and inheritance status was determined by manual inspection.

Inspection of the SVD-RPKM values for remaining 55 events (14 loci; see Table S#) revealed that these events strongly resemble copy-number polymorphic sites or contained processed pseudogenes. Such events are likely due to increases or decreases in copy number from the independent assortment of parental alleles; furthermore, changes in processed pseudogenes at other genomic loci can change the apparent copy number of annotated genes in the exome-capture reaction. As we had explicitly attempted to filter out such sites, we investigated these sites further. The observed signals in five of the loci, *PRKRA* (18 events), *RNF145* (3 events), and *CDC27* (2 events), *HNRNPA1* (1 event), and *TDG* (1 event) are likely driven by processed pseudogenes. Most of the remaining loci (*DAZL*, *BTNL8/BTNL3*, *CLPS*, *OR4*, *SIGLEC14*, and *KRT34*) were previously identified to be copy-number polymorphic by Conrad et al. (2010). Finally, the SVD-RPKM values of the last event, a duplication of exons in *KRT8* and *KRT18* are in-line with the signature of CNPs or highly duplicated exons.

Comparison of mrsFAST- and BWA-based read-depth estimation

BWA-based mappings were generated using the default settings for BWA (0.5.6) and post-processed with a pipeline developed specifically for SNP and single nucleotide variant (SNV) discovery. Reads which had more than one high-quality mappings were removed from the alignment and a minimum mapping quality (MAPQ) of 30 was required of all reads. The same method for generating RPKM values from BWA alignments was used as was for mrsFAST-based alignments. We calculated RPKM values for the same 194,080 intervals used elsewhere in this report, and again excluded targets with a median RPKM < 1 , a total of 7,117 probes in this experiment.

To make up the sample set for the comparison experiment, we combined 492 ESP samples, for which we had both mrsFAST and BWA-based mapping information, with the 8 HapMap samples. We noticed the the overall variance (as determined by the scree plot) in the BWA-based mapping was lower, and opted to remove only 6 components of variance. For the mrsFAST-based mappings, we removed the usual first 12 components. All other processing steps were done in the same fashion as elsewhere in this paper.

The signal-to-noise ratio for calls was calculated using the formula

$$\text{SNR} = |\mu_{\text{call}}| / \sigma_{\text{chromosome}}$$

where μ_{call} is the mean of the SVD-ZRPKM values for the exons within a call, and $\sigma_{\text{chromosome}}$ is the standard deviation of all the SVD-ZRPKM values of the call's chromosome. We calculated the SNR for the seven rare validated calls from table S1 for both mrsFAST-based and BWA-based SVD-ZRPKM values (Table S6). Six of seven rare CNVs showed improved SNR using the mrsFAST-based mappings, with a median improvement of 58% over BWA (mean 38% improvement).

Comparison to ExomeCNV algorithm:

We compared our algorithm to the previously published ExomeCNV (Sathirapongsasuti et al., 2011) in order to better understand the strengths and weaknesses of each. ExomeCNV is designed to detect copy number aberration in the context of cancer, a special case of copy number variation which requires additional parameters to be defined (e.g., the rate of admixture/contamination of tumor and normal), and which must be able to handle samples for which a large fraction of the genome is not diploid. Accordingly, ExomeCNV is designed around

a digital comparative hybridization algorithm, which requires that both the test and reference are as closely matched as possible (e.g., tumor-normal pairs of exomes from the same capture and sequence), and includes many features to better characterize cancer exomes. In contrast, ours is designed to discover genic deletions and duplications of exonic regions independently in each sample by first eliminating systematic noise using singular value decomposition.

We compared the ability of both algorithms to detect germline variation in DNA samples extensively analyzed and validated as part of other studies. To assess the sensitivity and specificity of both algorithms, we used the five HapMap samples for which exome sequence data had been generated and where high-density microarray analyses had been performed previously (Conrad et al., 2010). We set NA19240 as the reference sample, and used ExomeCNV to call CNVs on the remaining four samples (NA12878, NA15510, NA18517, and NA19129). Similar to the authors own use of the NovaAlign alignment package, we used the available BWA alignments for this comparison, and used the same 194,080 probes to generate an interval coverage file using the GATK (version 1.3.8) software package. We left all ExomeCNV parameters at their default values: sensitivity and specificity were set at 0.9999 for exons (maximizing specificity) and 0.99 for calls (“auc” option), and the admixture rate was set at a conservative 0.5 (despite the fact that we did not expect any biological admixture, we found that keeping this setting reduced the number of false positive calls).

Among the four test samples, ExomeCNV predicted 450 CNVs, of which only 63 (14%) overlapped with calls in the Conrad et al. call set by more than 10% reciprocal overlap. In contrast, our algorithm found 24 calls among these four samples, of which 21 (87.5%) overlapped the Conrad et al. set. While both programs were able to find all of the five rare CNVs (Table S3), we note that ExomeCNV predicted 16 CNVs larger than 500kb, which did not have any overlap with the high resolution Conrad et al. set of calls. This low specificity

would make it very difficult to find “true positives” in the ExomeCNV output, even when filtering for large CNVs only.

Using exon-level log-ratio output from ExomeCNV, we next compared how sensitive it was to changes in copy-number of duplicated genes. Across the 62 CNP loci genotyped by our algorithm (Table S4), ExomeCNV was able to generate LogR values for 51 loci (82%). Example correlations and a comparison between ExomeCNV and our algorithm are shown for four loci in Figure S8a. Across all loci, when compared to the log-ratio values to the whole-genome estimate for each locus, the median r^2 across these loci was 0.57 (c.f. this work’s algorithm $r^2 = 0.92$). As with the BWA alignment comparison, the genotyping dynamic range of ExomeCNV was severely limited, and the LogR values from ExomeCNV correlated only poorly with the corresponding whole-genome estimates of absolute copy number for loci with median copy number greater than seven (Figure S8c).

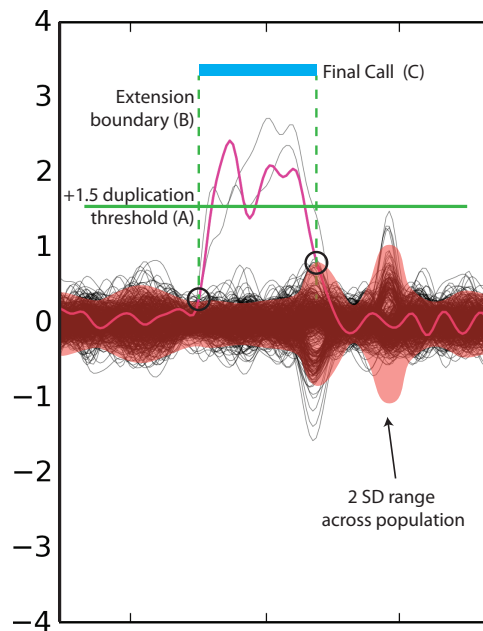
Finally, although the authors of ExomeCNV recognize that their algorithm depends on sample-to-sample consistency, large cohorts of tens to hundreds of exomes cannot be expected to maintain such consistency. Crucially, our algorithm allows for the comparison of samples from different cohorts, and even different iterations of the exome capture reaction itself. To demonstrate this, we examined two ESP samples from two different experimental cohorts (but stemming from the same study, and using the same capture kit version, library preparation steps and sequencing machines). The output from ExomeCNV for chromosome 20 is shown in the top left panel of Figure S7. When we counted the fraction of exome probes which ExomeCNV predicted as copy-number variant, we found that a biologically implausible 96.6% of the exome was detected as changed from diploid copy number (Figure S7, top right panel). In contrast, when we picked an ESP sample from the same experimental batch (and which was closely matched based on the variance we observed using the SVD

decomposition) as the reference, ExomeCNV reported only 0.4% of exome probes as non-diploid (Figure S7, bottom panel). When we applied our algorithm (this work) at a very sensitive setting (± 1 SVD-ZRPKM threshold), we found only that for the same samples, only 0.06% and 0.15% of the exons were altered from diploid. This comparison highlights the strength of singular value decomposition for eliminating batch effects and systematic noise that may arise from exome capture experiments.

References:

- Campbell, C. D., Sampas, N., Tsalenko, A., Sudmant, P. H., Kidd, J. M., Malig, M., Vu, T. H., et al. (2011). Population-Genetic Properties of Differentiated Human Copy-Number Polymorphisms. *The American Journal of Human Genetics*, 88(3), 317–332. doi:10.1016/j.ajhg.2011.02.004
- Conrad, D. F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., Aerts, J., et al. (2010). Origins and functional impact of copy number variation in the human genome *Nature*, 464(7289), 704–712. doi:10.1038/nature08516
- Kidd, J. M., Cooper, G. M., Donahue, W. F., Hayden, H. S., Sampas, N., Graves, T., Hansen, N., et al. (2008). Mapping and sequencing of structural variation from eight human genomes *Nature*, 453(7191), 56–64. doi:10.1038/nature06862
- Sudmant, P. H., Kitzman, J. O., Antonacci, F., Alkan, C., Malig, M., Tsalenko, A., Sampas, N., et al. (2010). Diversity of human copy number variation and multicopy genes *Science*, 330(6004), 641–646. doi:10.1126/science.1197005
- Sathirapongsasuti, J. F., Lee, H., Horst, B. A. J., Brunner, G., Cochran, A. J., Binder, S., Quackenbush, J., et al. (2011). Exome sequencing-based copy-number variation and loss of heterozygosity detection: ExomeCNV. *Bioinformatics*, 27(19), 2648–2654. doi:10.1093/bioinformatics/btr462

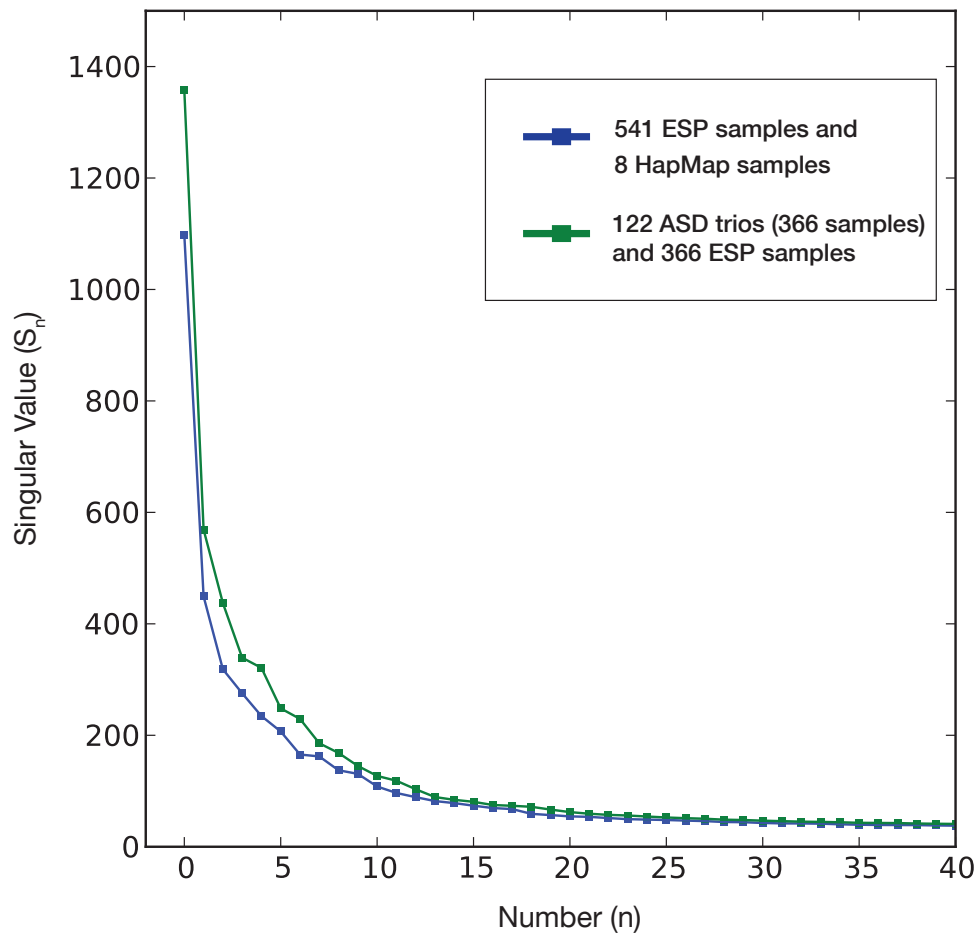
Figure S1: Threshold call overview:



S1: Threshold algorithm

To discover rare CNVs, we found smoothed SVD-ZRPKM values which crossed a threshold (A) of +1.5 or -1.5 for duplications and deletions, respectively. To account for the fact that our smoothed values shrink the apparent size of the call, we extended calls such that the final call (C) better represented the extend of the actual CNV. To do this, we extended calls from the initial supra-threshold event until the smoothed SVD-ZRPKM values dipped below ± 2 standard deviations surrounding the population median (red highlight) of the SVD-ZRPKM values (marked in figure by line [B], and by black circles).

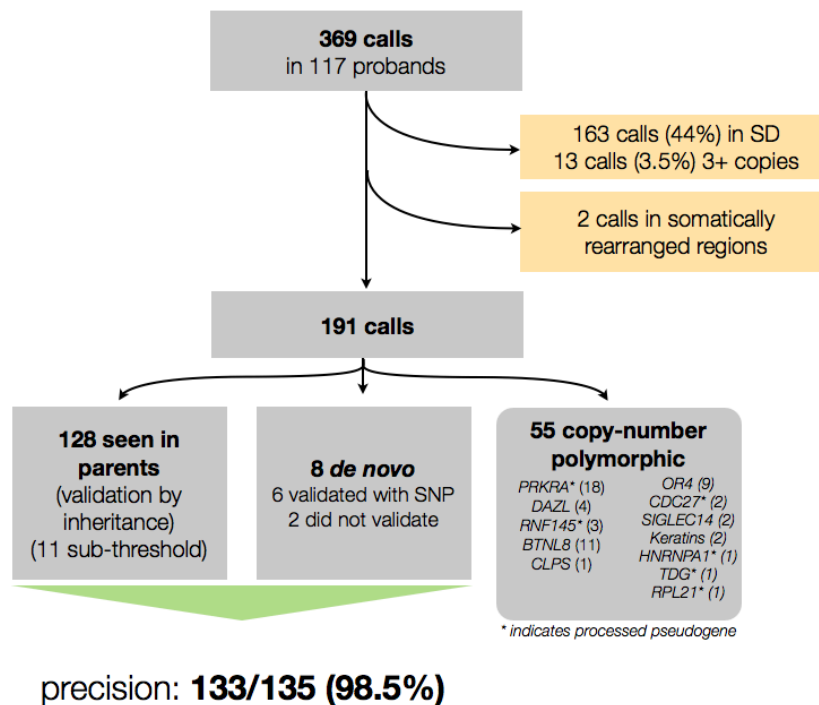
Figure S2: Scree Plot



S2: Scree plot

This scree plot shows the first 40 singular values (S_n) from the HapMap (blue) ASD trio (green) samples. The relative contributed variance of each singular value is proportional to its strength indicated on the y-axis.

Figure S3: Filtering of calls in 122 ASD trios



S3: Filtering of calls in 122 ASD trios:

Starting with 369 detected calls in 122 ASD probands, we applied a set of filters to restrict calls to unique/diploid regions of the genome in order to estimate the precision of our method. Calls which had greater than 50% reciprocal overlap (as determined by the fraction of underlying exome probes within the call also in segmental duplications) with segmental duplications were removed (163/369, or 44%). Next, we calculated the median copy number of calls based on whole-genome read-depth copy-number estimates from ~660 genomes {Sudmant:2010kp}, and additionally filtered 13 calls (3.5%) with more than 3+ copies population-wide (as events stemming from these segmentally-duplicated or higher-copy regions of the genome are likely due to the independent assortment of parental haplotypes, and not “true” rare CNVs). Additionally, we manually curated the calls to remove calls within regions undergoing somatic rearrangement (two calls; one at the *IGH* locus and one in the *HLA* locus), and merged adjacent or overlapping calls. We classified the 128/191 remaining calls as inherited (see text for details), eight as de novo, and 55 as copy-number polymorphic. We estimated the precision of our method for the detection of rare CNVs from our accuracy with de novo CNVs (6/8) and the number of inherited events (128 in total), resulting in a 98.5% precision.

Figure S4: Filtering of calls from Conrad et al. (2010) array-CGH experiment:

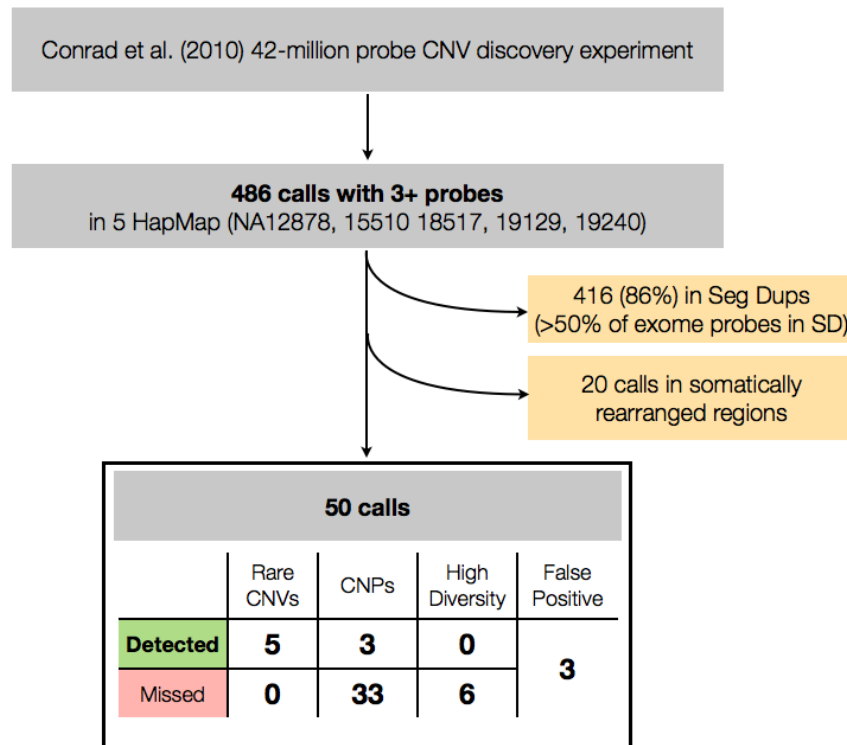


Figure S4: Filtering of calls from Conrad et al. (2010) array-CGH experiment:

We estimated the sensitivity of our method using array comparative genomic hybridization calls from Conrad et al. (2010) as a gold standard. Starting with calls from the 42-million probe CNV discovery experiment in {Conrad:2010ja}, there were 486 calls with at least three exome probes in the five HapMap samples for which we had exome sequences. Calls which had greater than 50% reciprocal overlap (as determined by the fraction of exome probes within the call also in segmental duplications) with segmental duplications were removed; additionally, we removed 20 calls in somatically rearranged regions. We manually inspected the remaining 50 calls (Table S2) to assess sensitivity of the method. Five events were rare and all five were detected by the ± 1.5 SVD-ZRPKM threshold. There were 36 CNPs, of which only three cross the threshold for rare CNVs. Six of the remaining events were either located in high diversity regions of the genome. Finally, we noted that three of the events were very likely false positive events in the Conrad dataset, as they were not corroborated by Illumina 1M SNP microarray data, nor were they found by a fosmid mapping approach{Kidd:2008bo}.

Figure S5: BWA and mrsFAST comparison – genome view

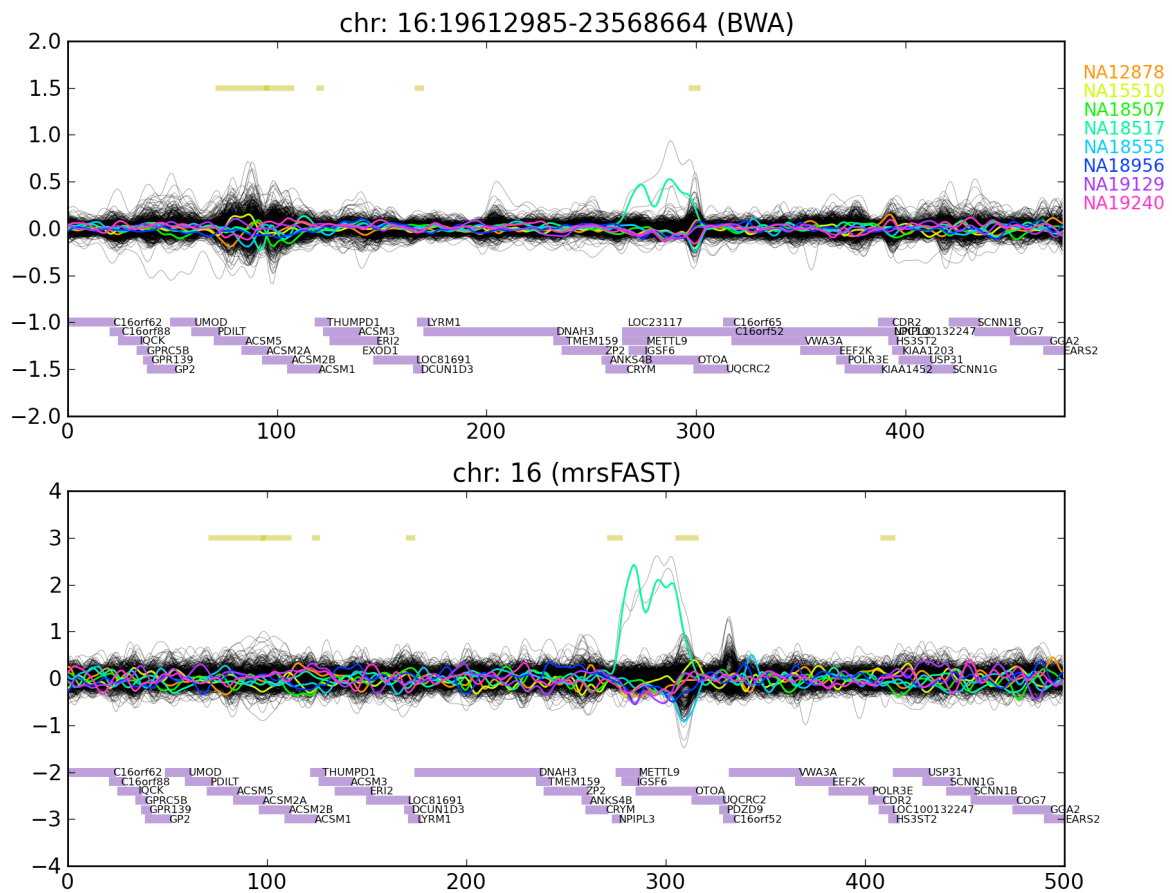


Figure S5:

Visual comparison of BWA and mrsFAST-based mappings on a stretch of chromosome 16. We found that across the seven validated rare CNVs from table S1, the SVD-ZRPKM values derived from BWA mappings had a 57% lower signal-to-noise ratio, as noted by the decreased signal of NA18517 at the *METTL9/OTOA* locus for BWA-based mappings. (Y-axes have different scales to account for the lower standard deviation seen in the BWA-based SVD-ZRPKM values.)

Figure S6a: BWA and mrsFAST comparison – genotyping accuracy

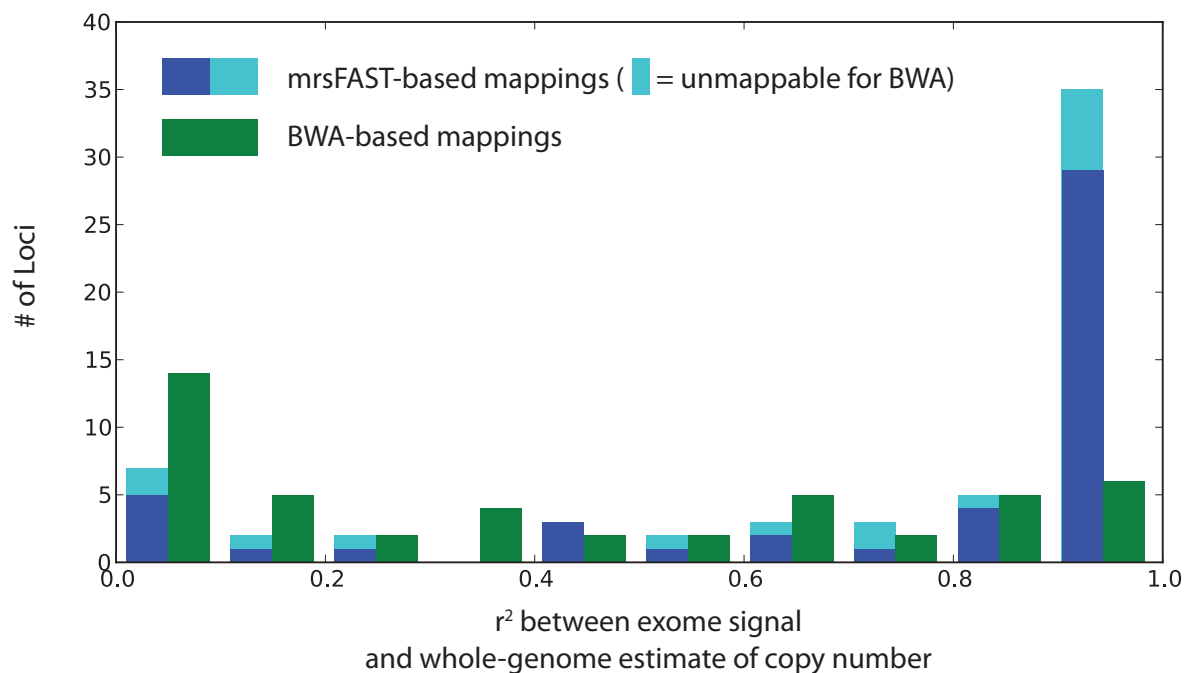


Figure S6a:

Comparison of correlations coefficients of SVD-ZRPKM to whole-genome copy number estimate across 62 CNP loci between BWA- and mrsFAST-based mapping strategies. The median r^2 for the BWA-based experiment is 0.62 (green bars), while for mrsFAST the median r^2 is 0.92 (blue bars). Moreover, for 15 loci, the BWA-based mappings did not have sufficient read-coverage in the loci to be genotyped, making them intractable to BWA-based read-depth genotyping.

Figure S6b: BWA and mrsFAST comparison – by median copy number

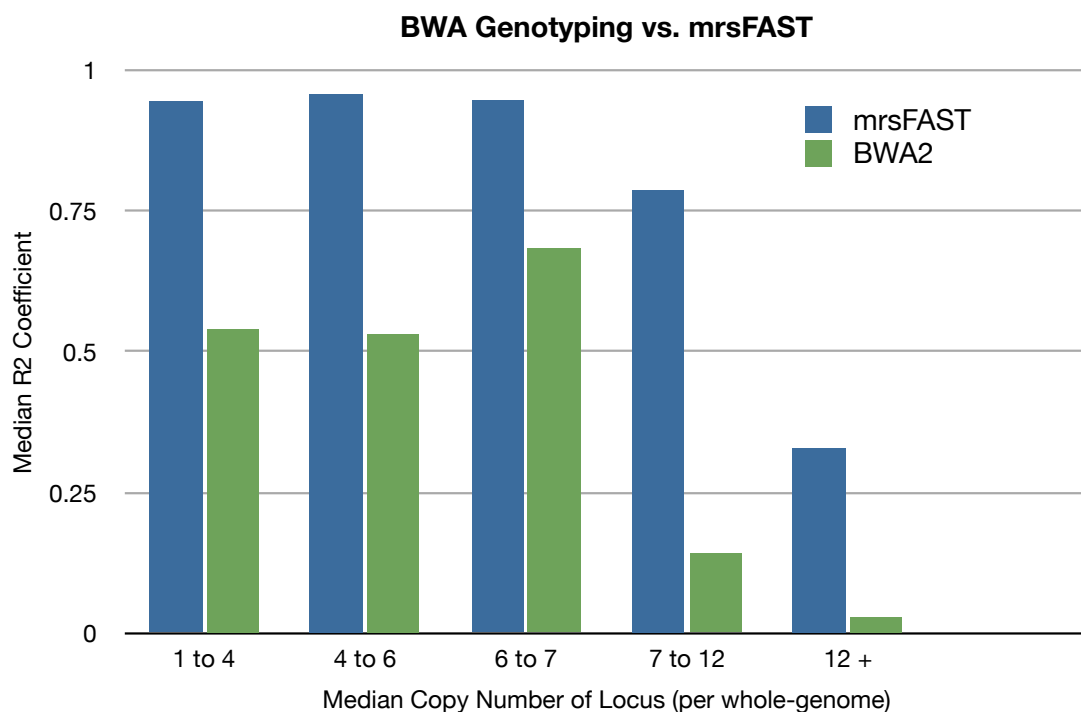


Figure S6b:

Comparison of BWA-based and mrsFAST-based alignments for genotyping of 62 loci, binned by median copy number of each locus. We calculated the median copy number of the 62 loci based on whole-genome read-depth copy-number estimates from ~660 genomes. We note that mrsFAST-based mapping significantly improves the correlation between the SVD-ZRPKM genotyping scores and whole-genome absolute copy number, especially for loci with a median copy number between 7 and 12.

Figure S6c: BWA and mrsFAST comparison – *LRRC37A3* locus

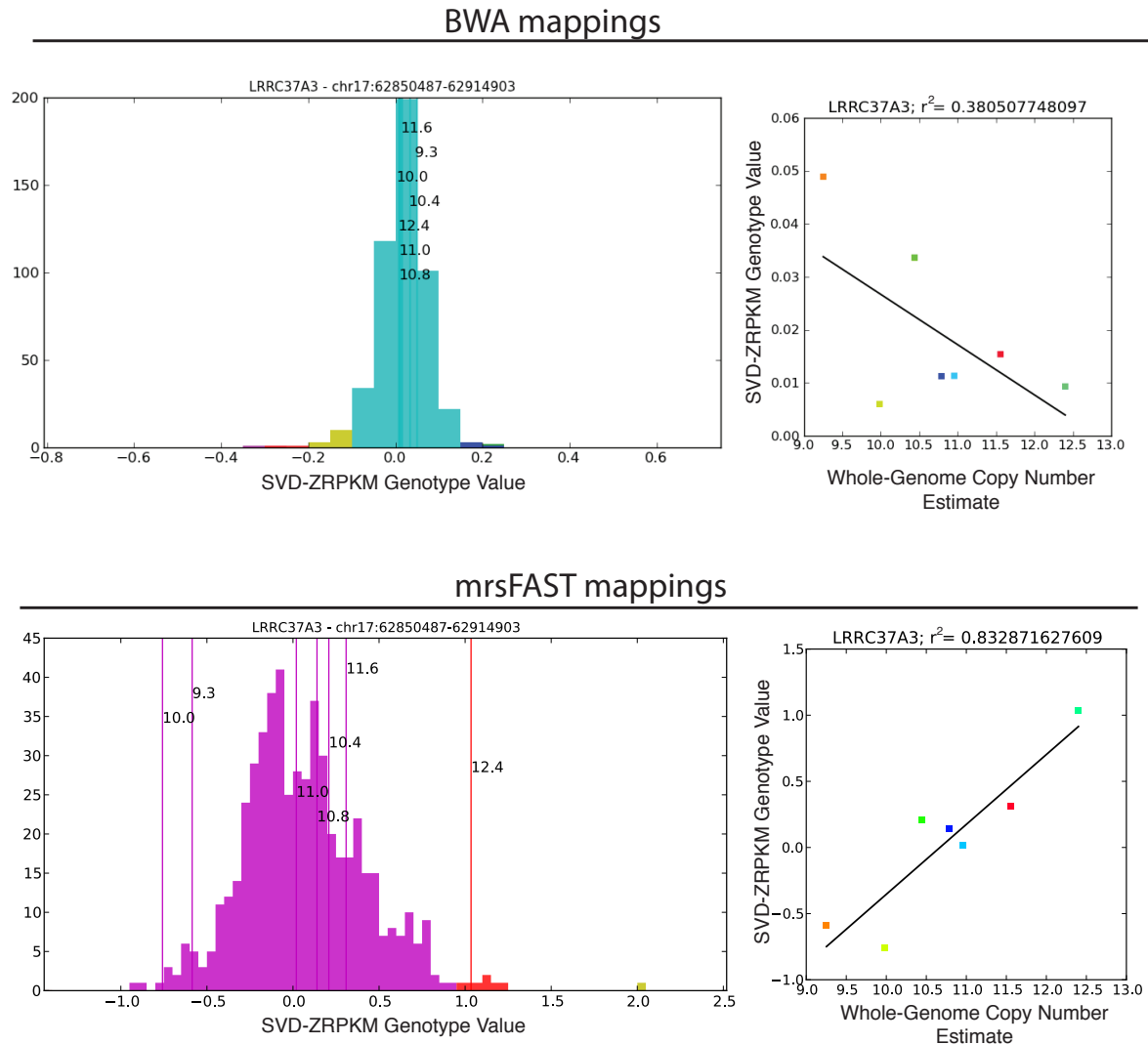


Figure S6c:

Example CNP locus (*LRRC37A3*) representative of difficulty for BWA-based genotyping of loci with median population copy number greater than seven.

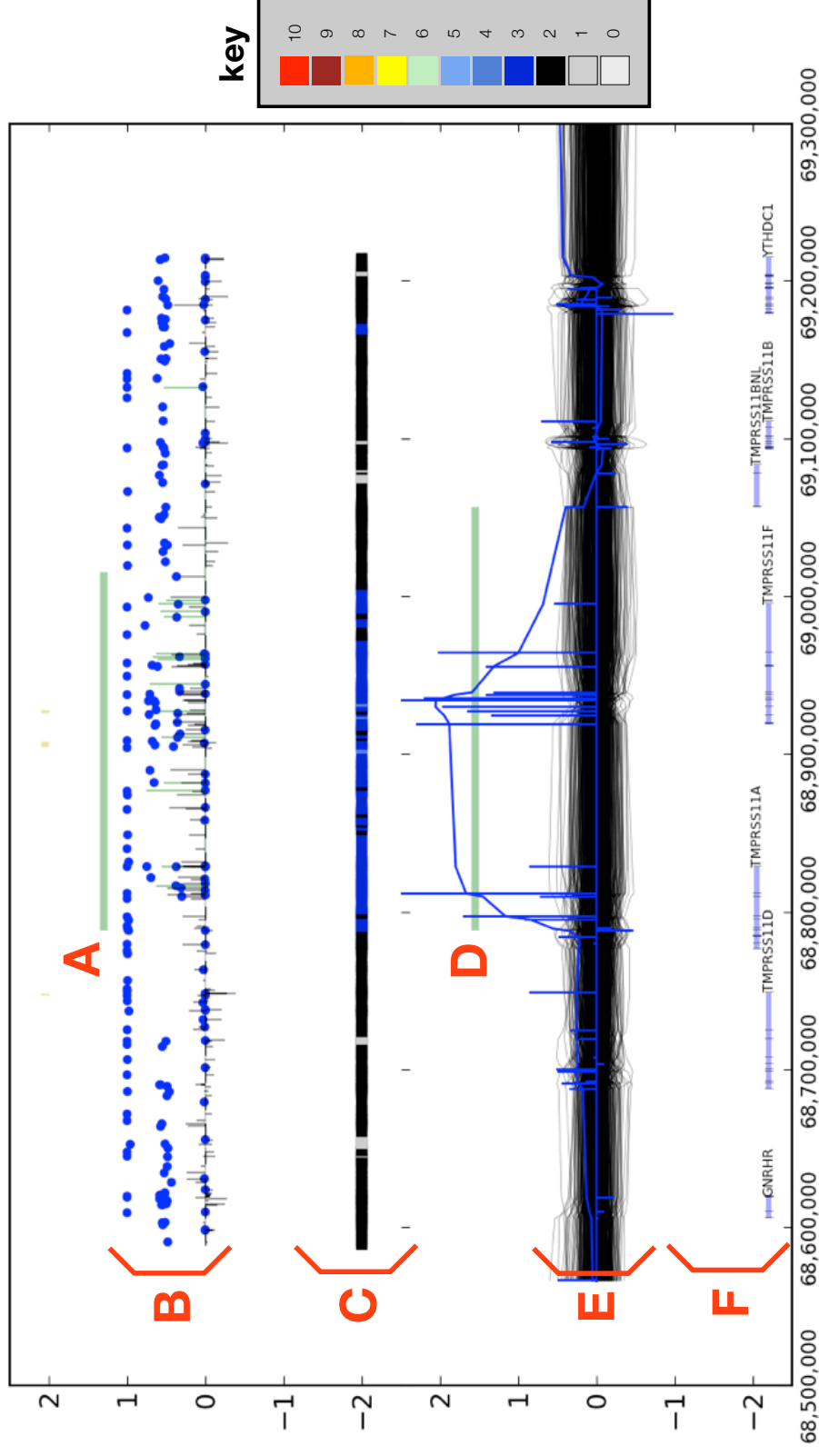
(top left): Histogram showing SVD-ZRPKM genotype values of 8 HapMap samples (indicated by horizontal lines) and 492 ESP samples. Annotated numbers on the histogram indicate the absolute copy number, as estimated from whole genome sequencing of HapMap samples.

(top right): Correlation between SVD-ZRPKM values and whole-genome derived absolute copy number for 7 HapMap samples. The poor resolution of BWA-based mappings for this locus contribute to a poor correlation and low accuracy.

(bottom left, right): the same locus for mrsFAST-based mappings. Both the histogram and the scatter plot show markedly increased resolution for distinguishing copy number states and improved SVD-ZRPKM to absolute copy-number correlation.

Figure(s) S7

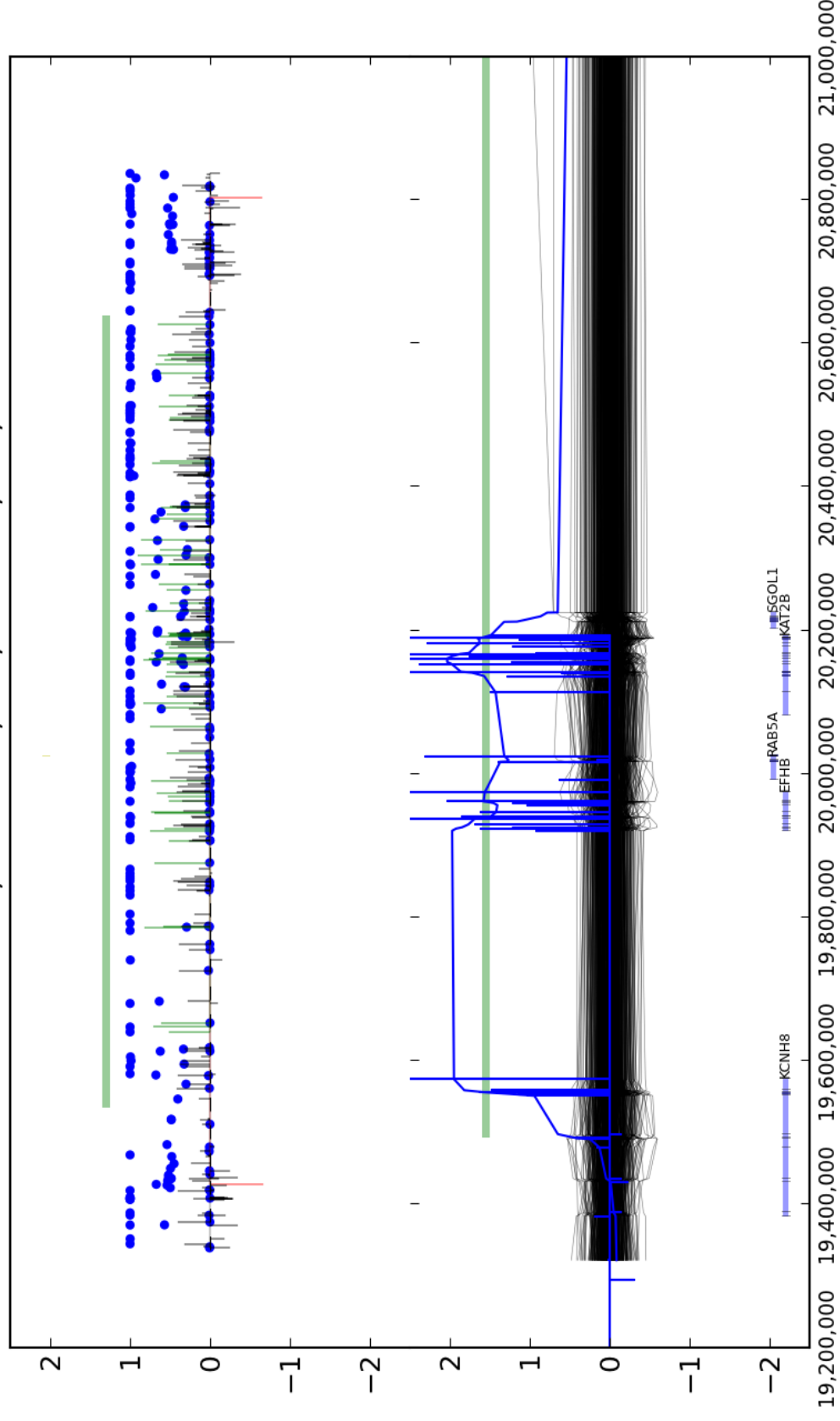
sample, chr: start – stop (hg19)



- A)** CNV call from Conrad et al. (2010)
- B)** SNP-array data (black lines - LogR; blue dots B-allele frequency)
- C)** Whole-Genome read depth from Sudmant et al. (2010) – see **key** at right
- D)** Exome-based CNV call
- E)** SVD-ZRPKM values (blue line: sample with call; black lines: 533 ESP samples)
- F)** Refseq Genes

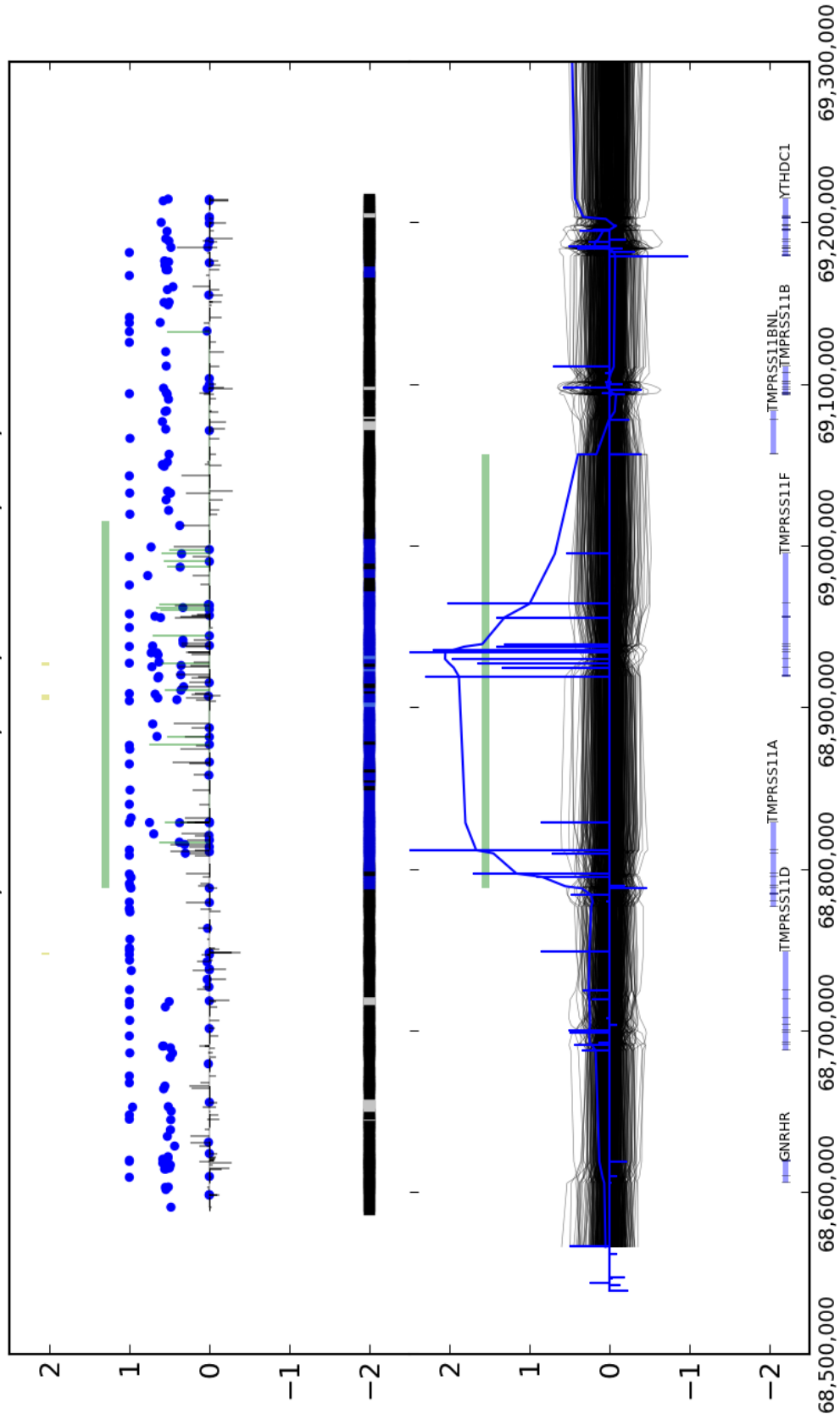
Rare Duplication

NA15510, chr3: 19,535,653 - 20,638,501



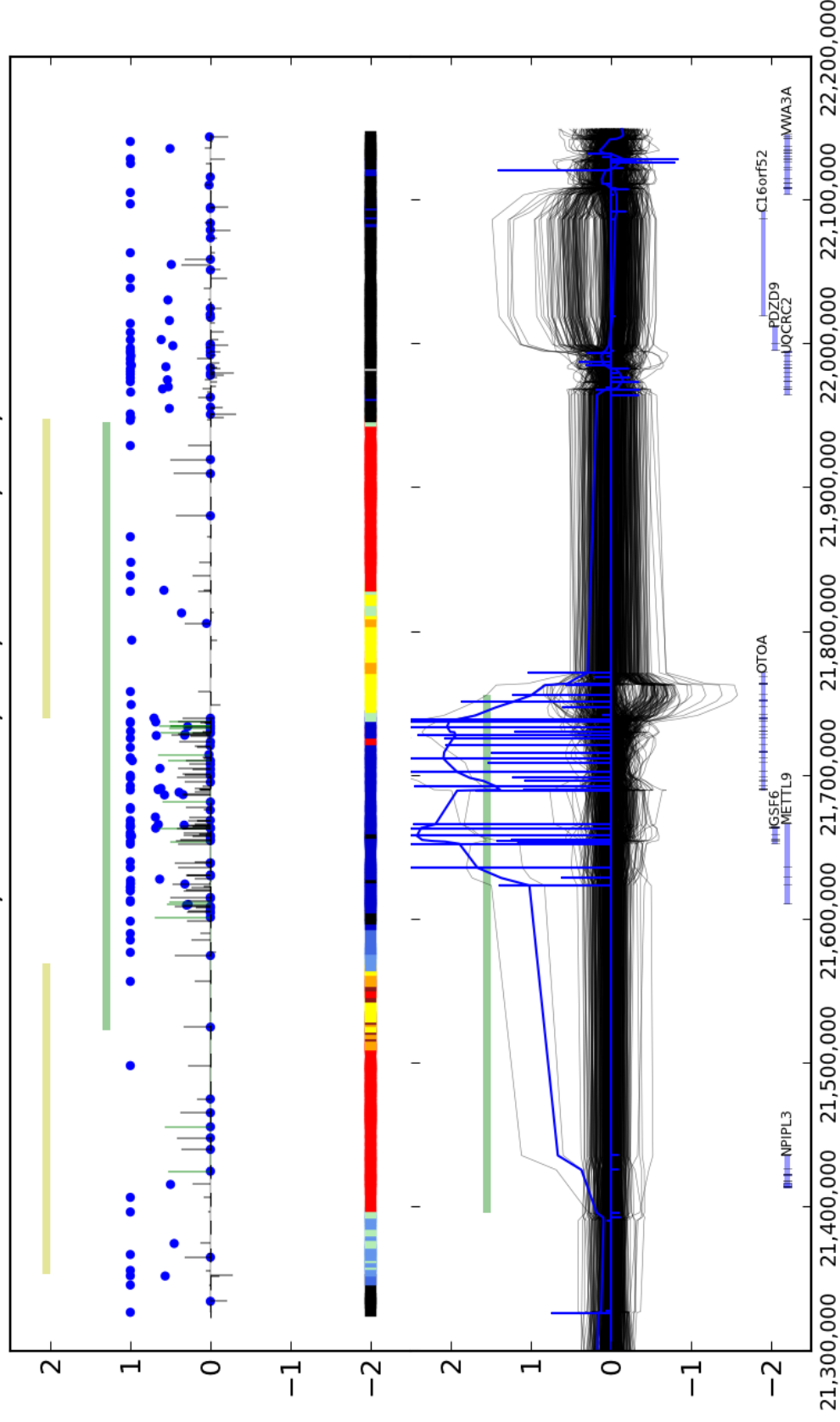
Rare Duplication

NA18517, chr4: 68,788,730 - 69,016,101



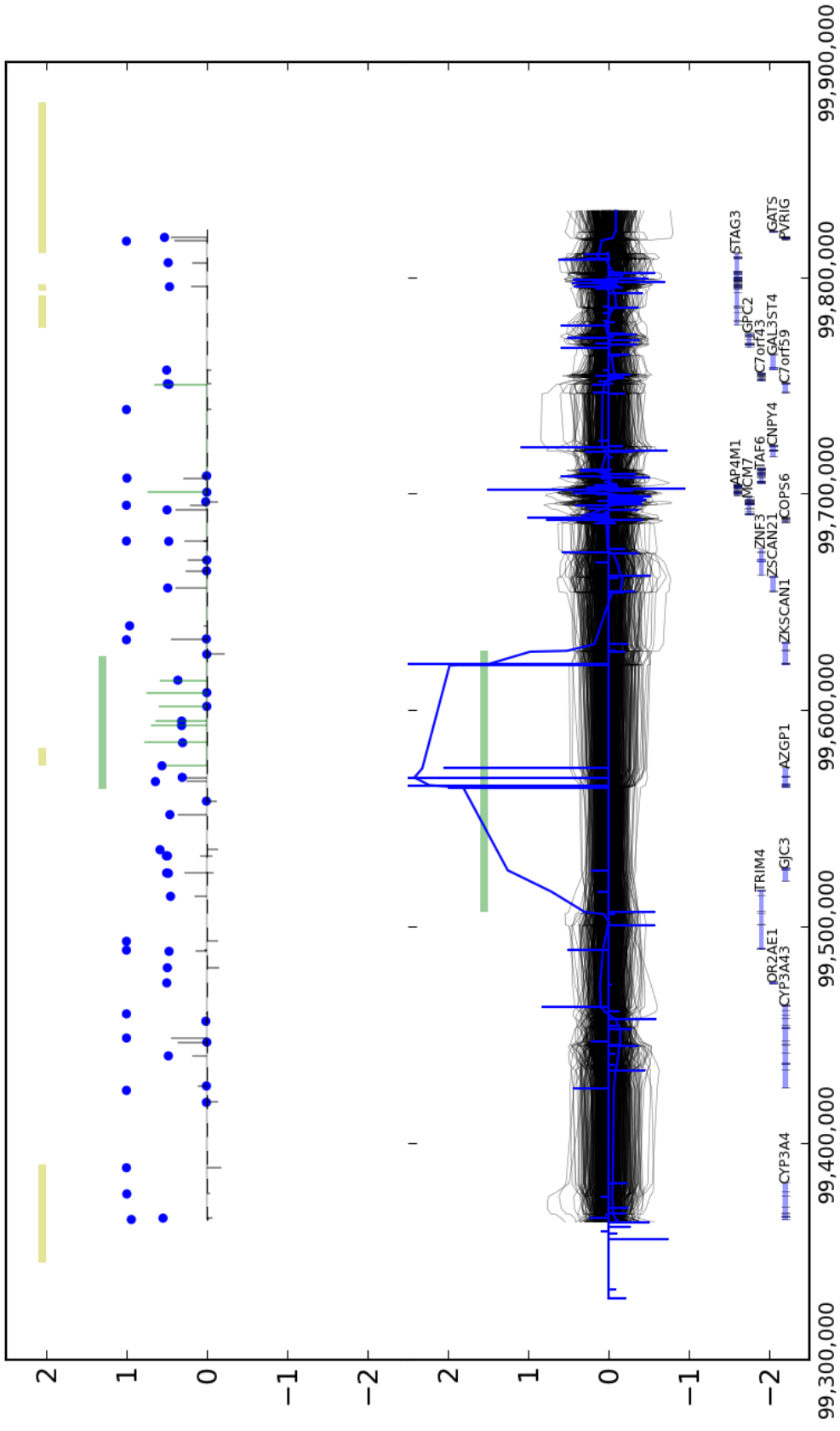
Rare Duplication

NA18517, chr16: 21,523,044 - 21,946,347



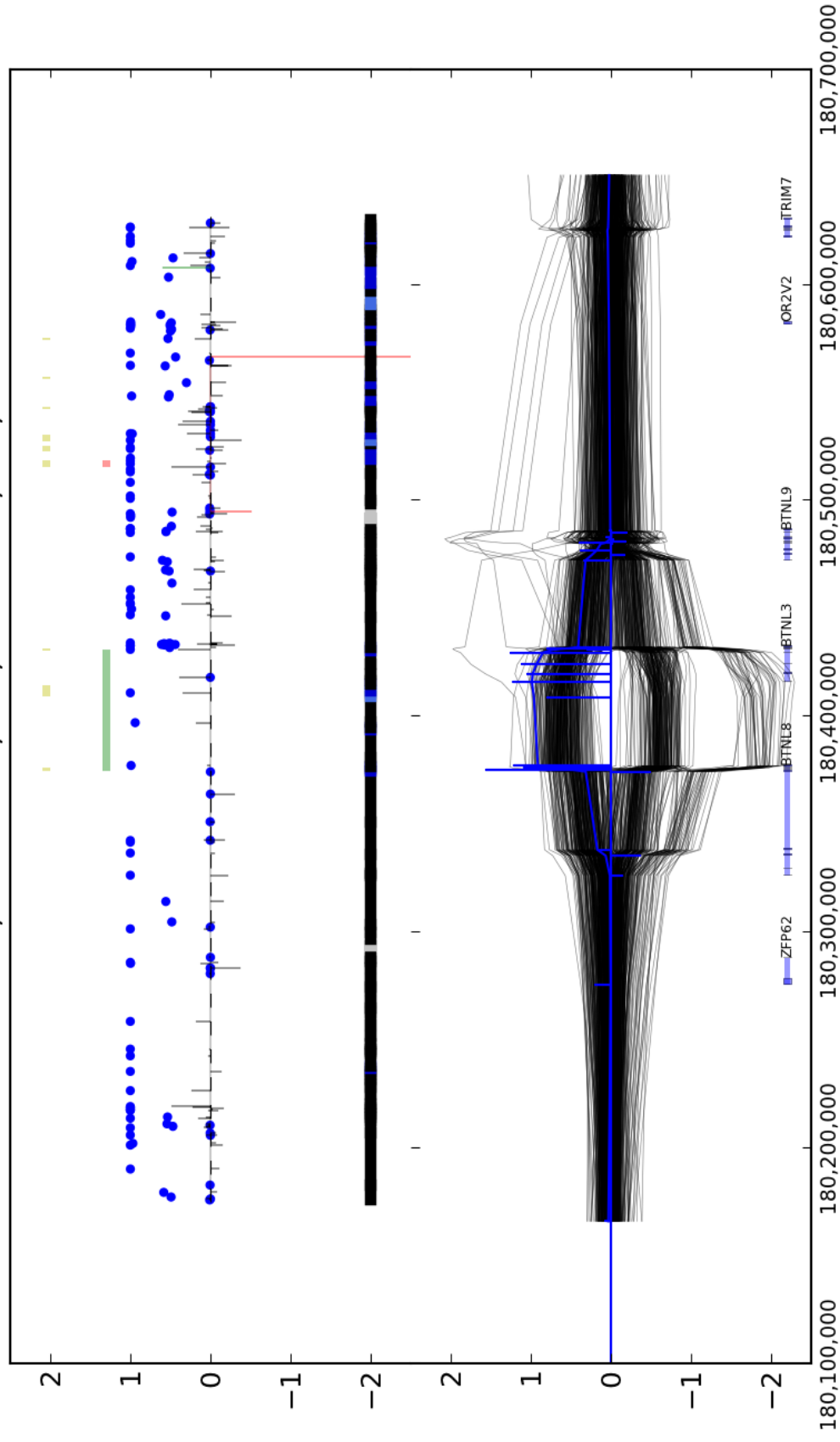
Rare Duplication

NA15510, chr7: 99,564,133 - 99,625,411



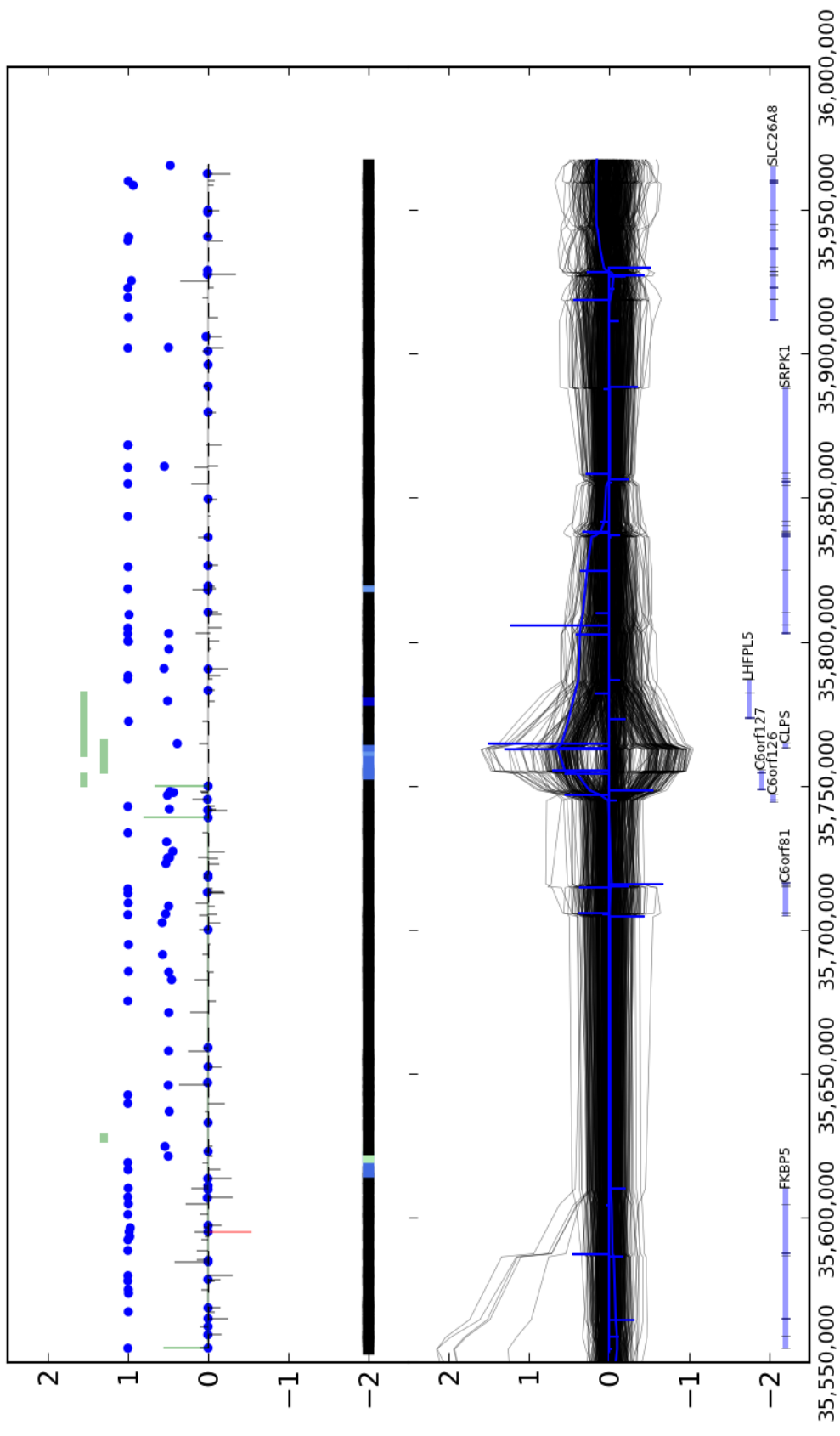
CNP

NA18517, chr5: 180,374,610 - 180,431,110

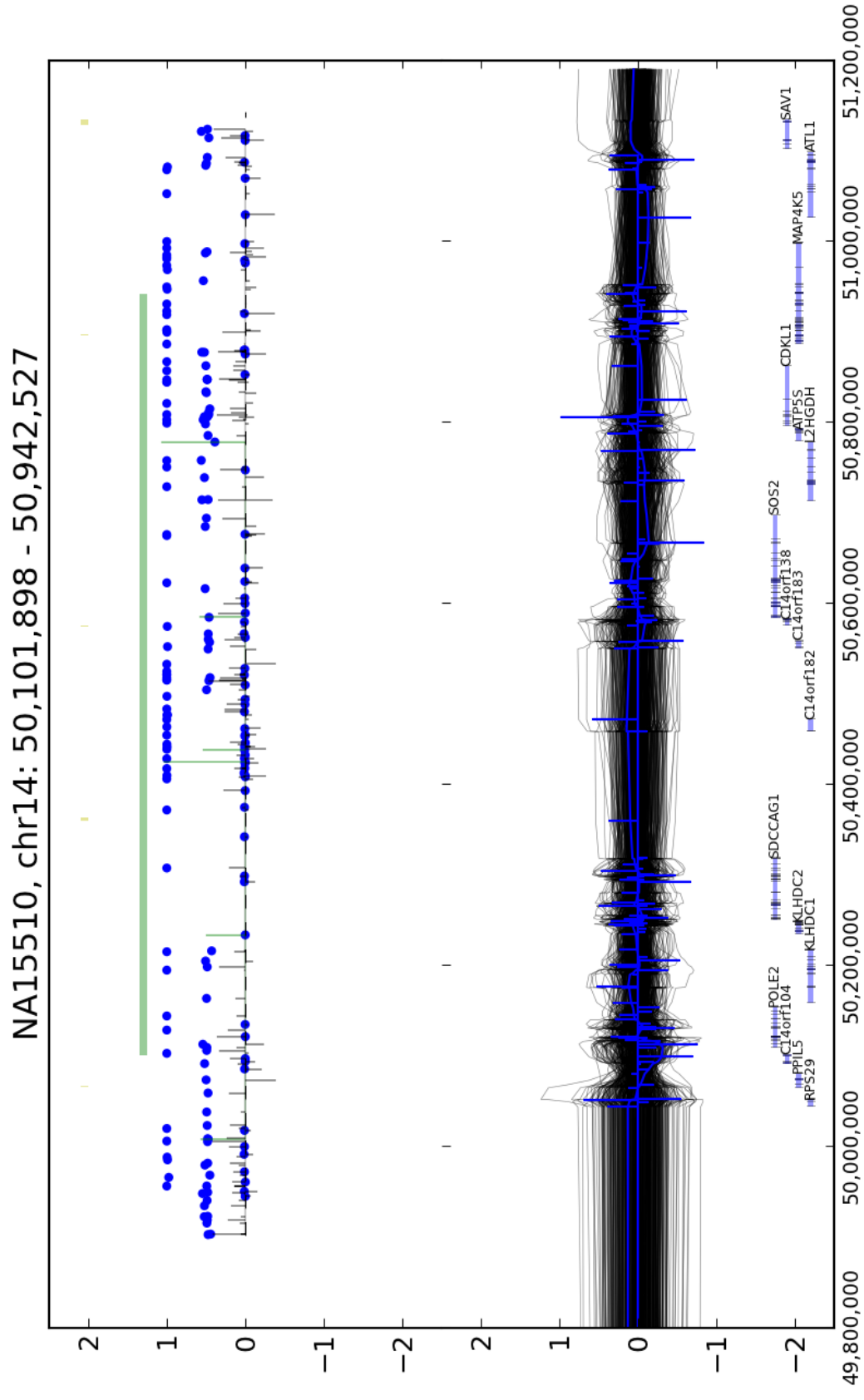


CNP

NA19240, chr6: 35,754,736 - 35,766,415

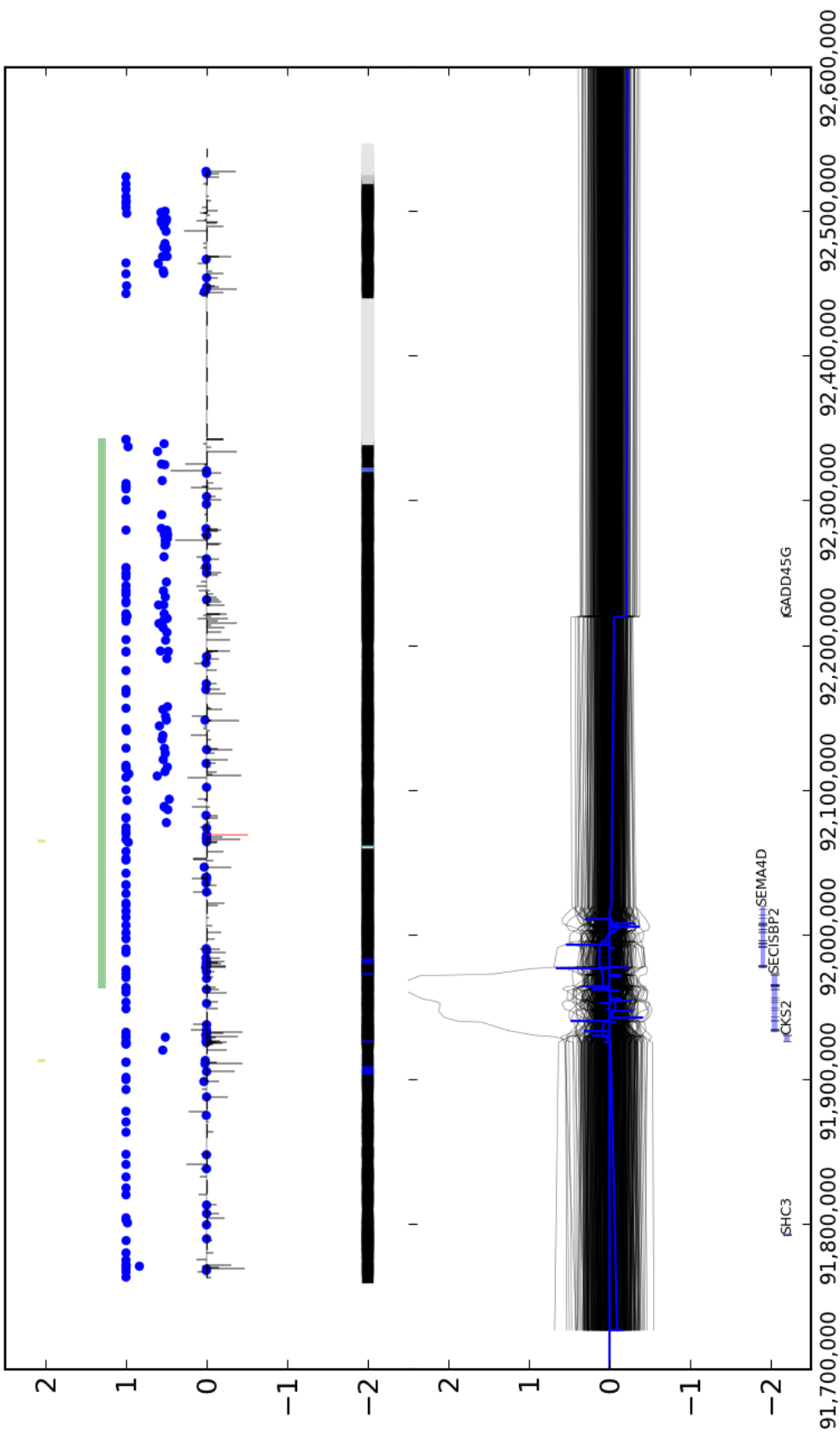


False Positive



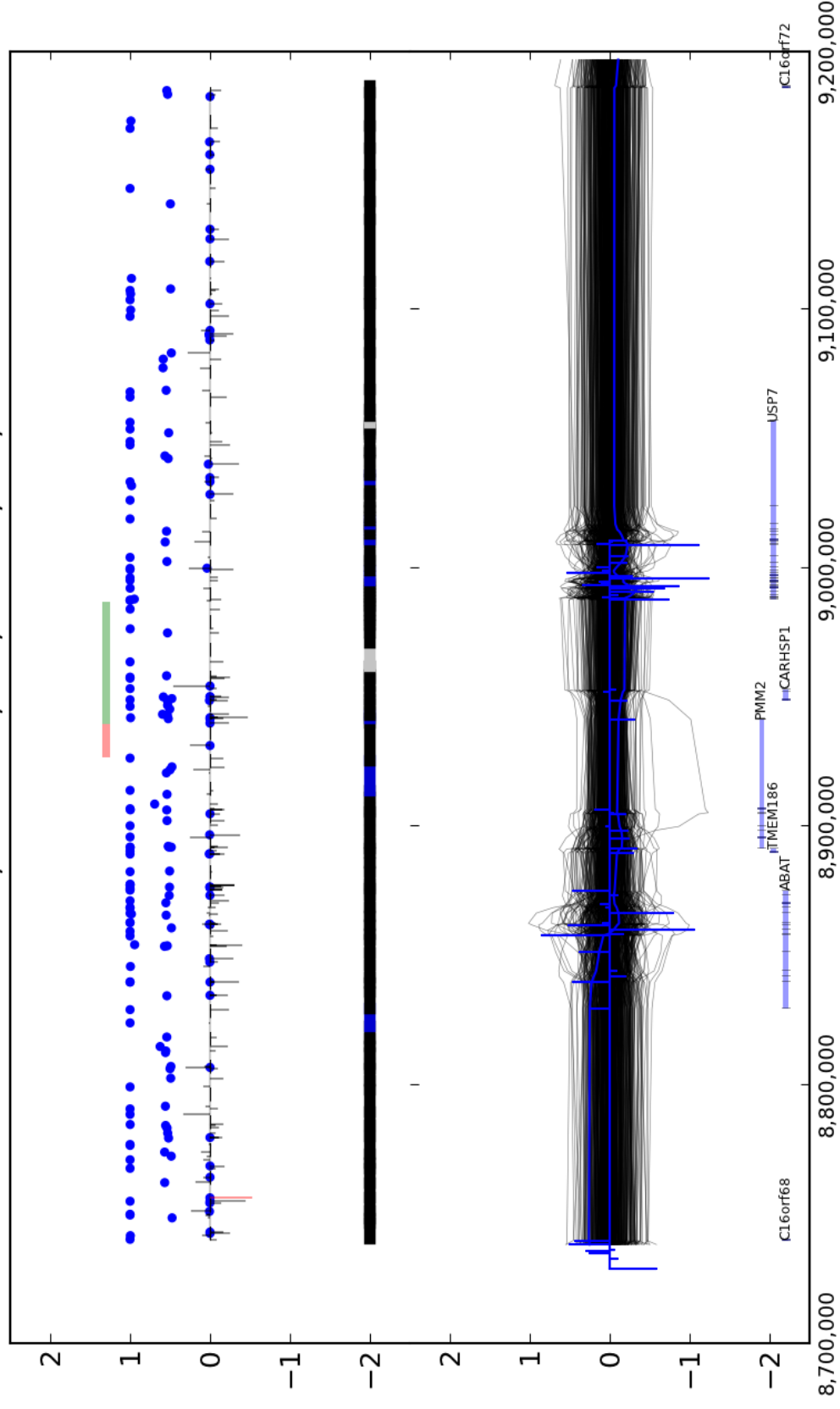
False Positive

NA12878, chr9: 91,963,403 - 92,343,382

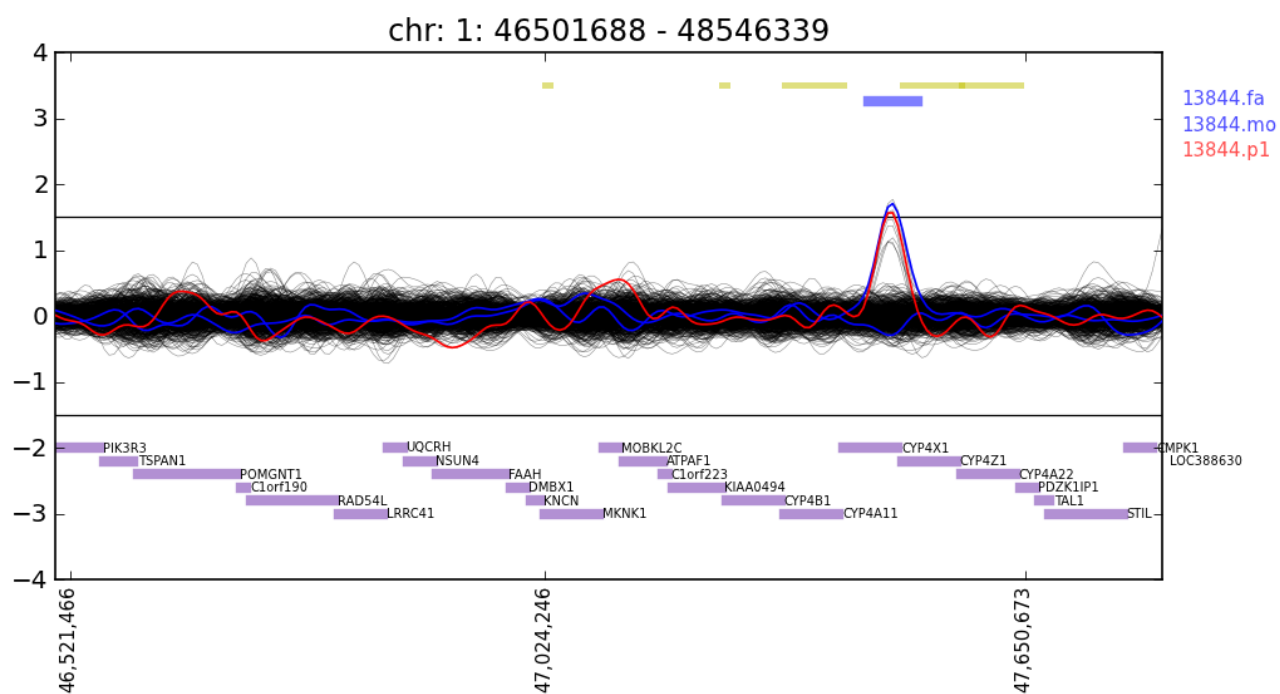
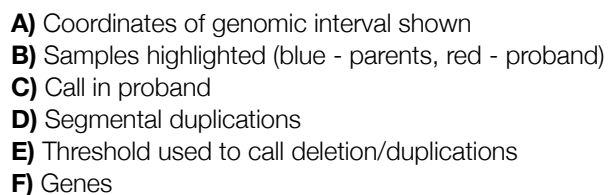


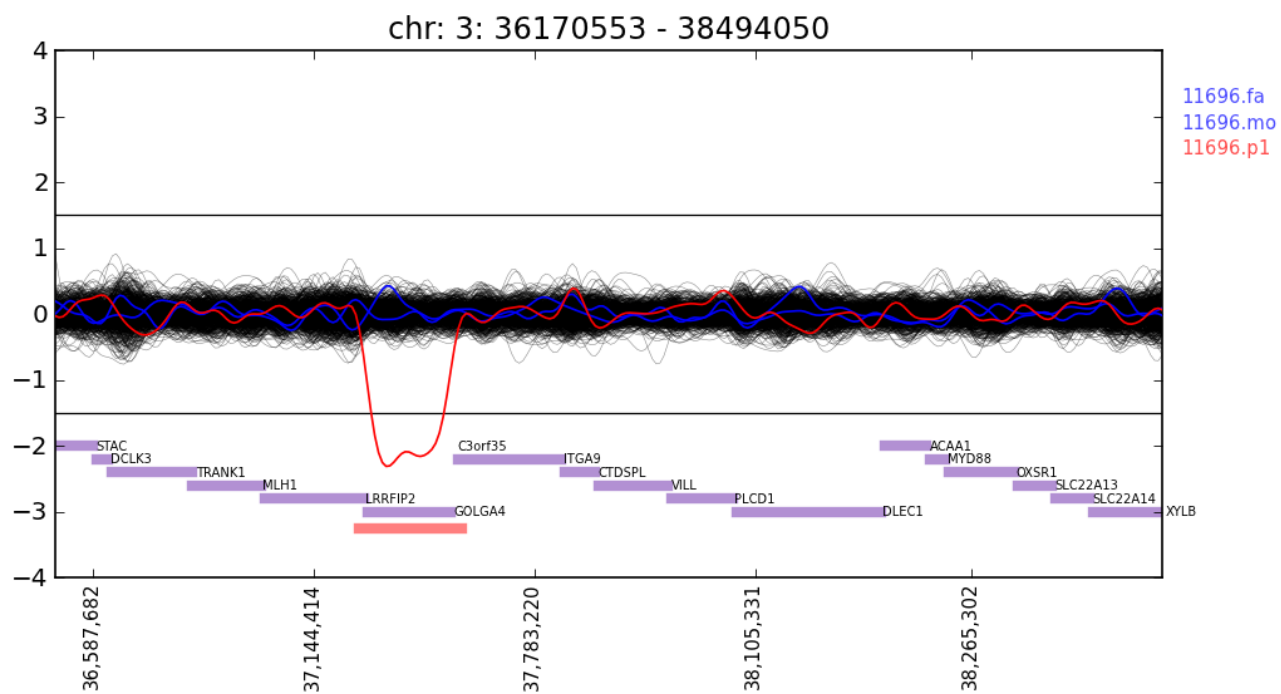
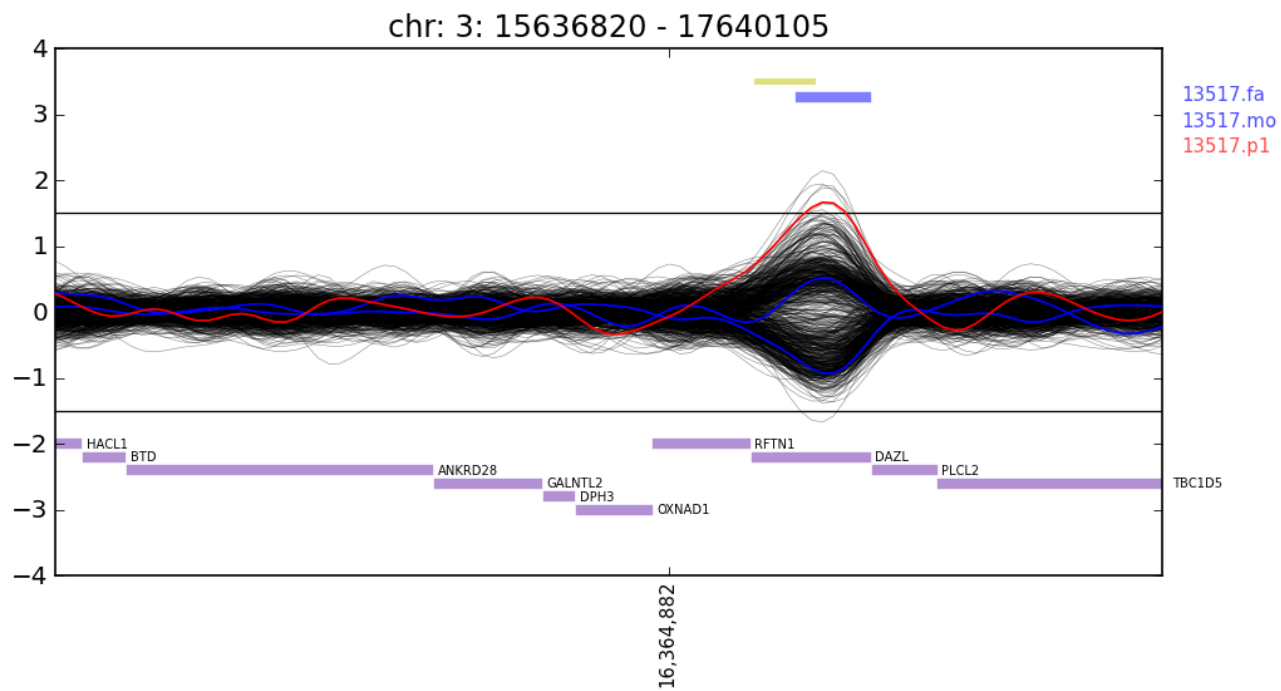
False Positive

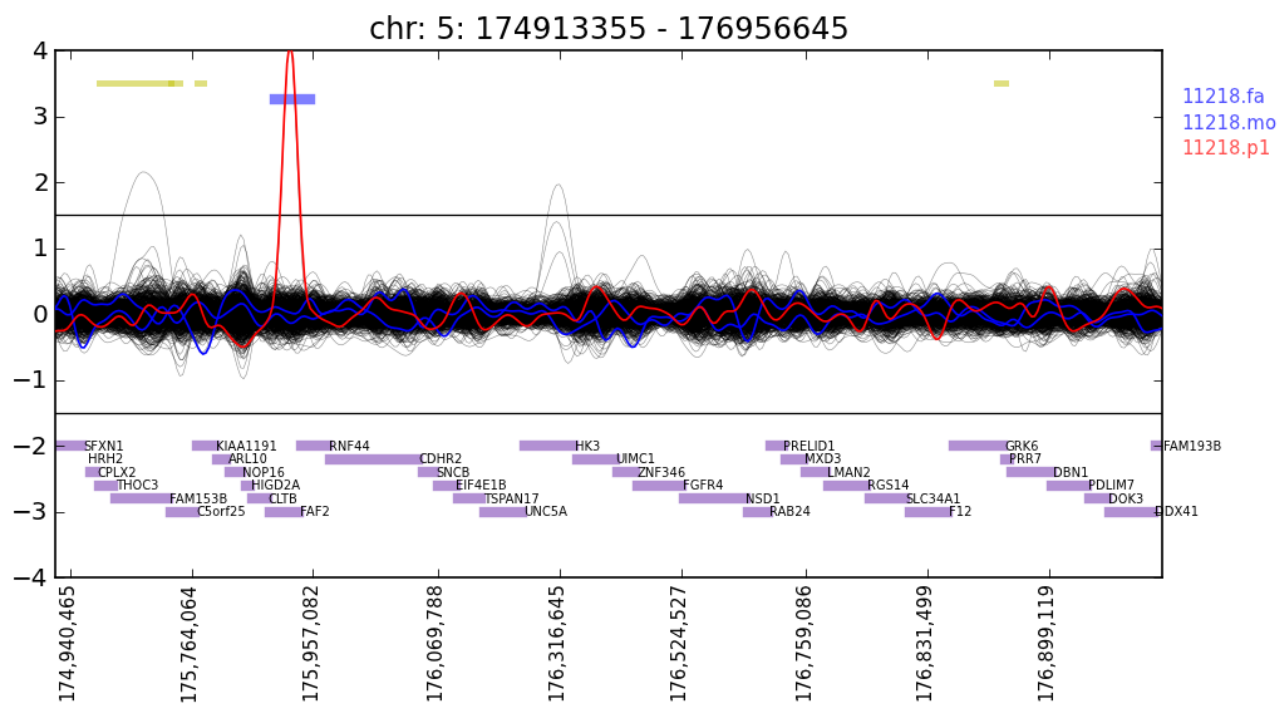
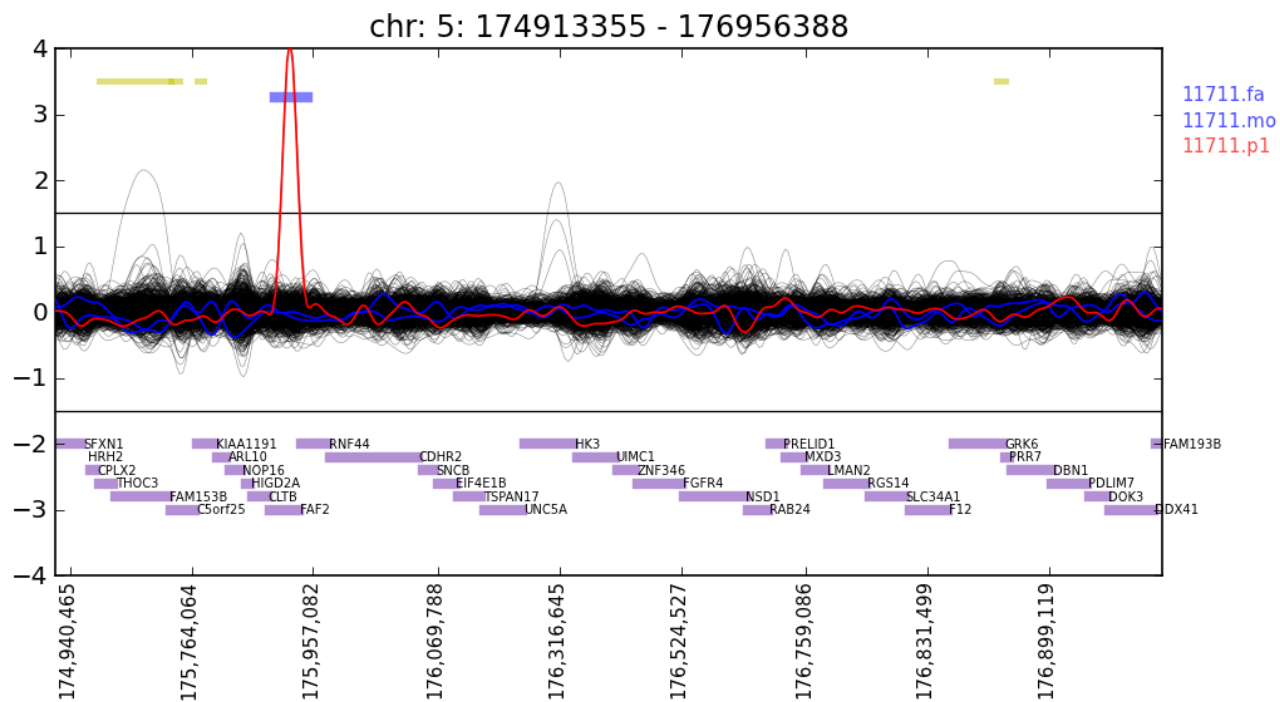
NA19129, chr16: 8,939,807 - 8,987,025

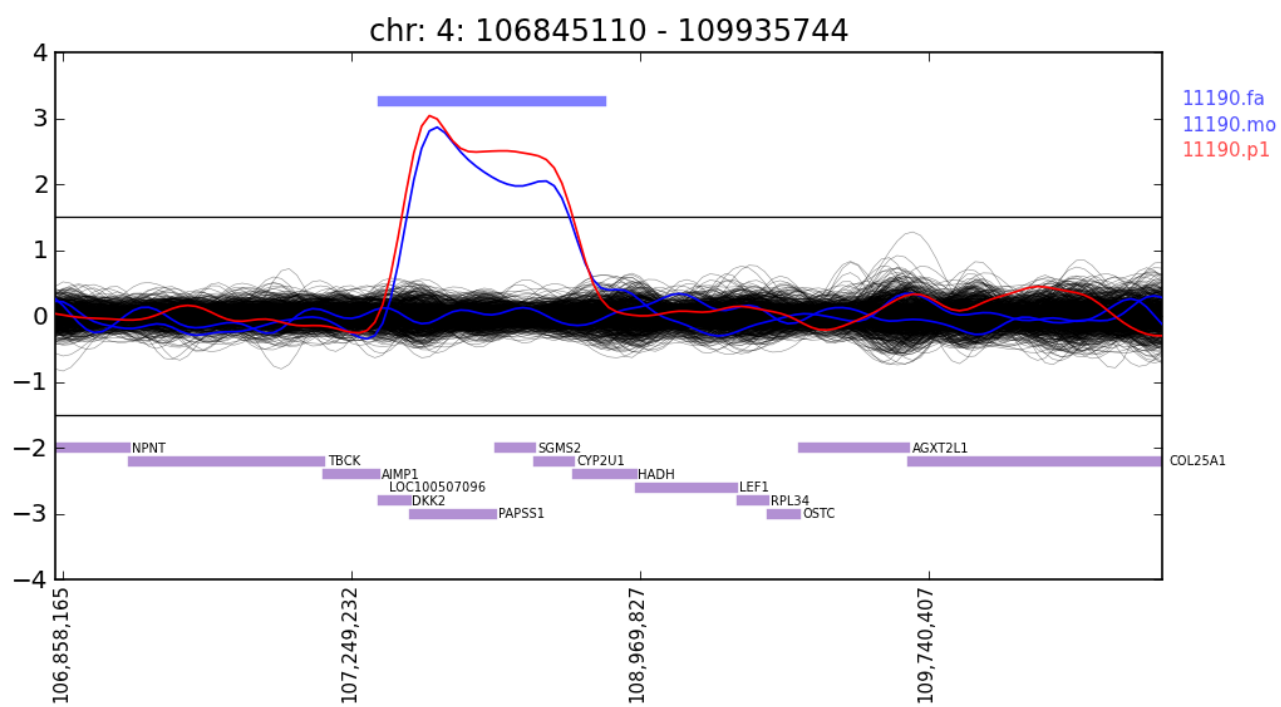
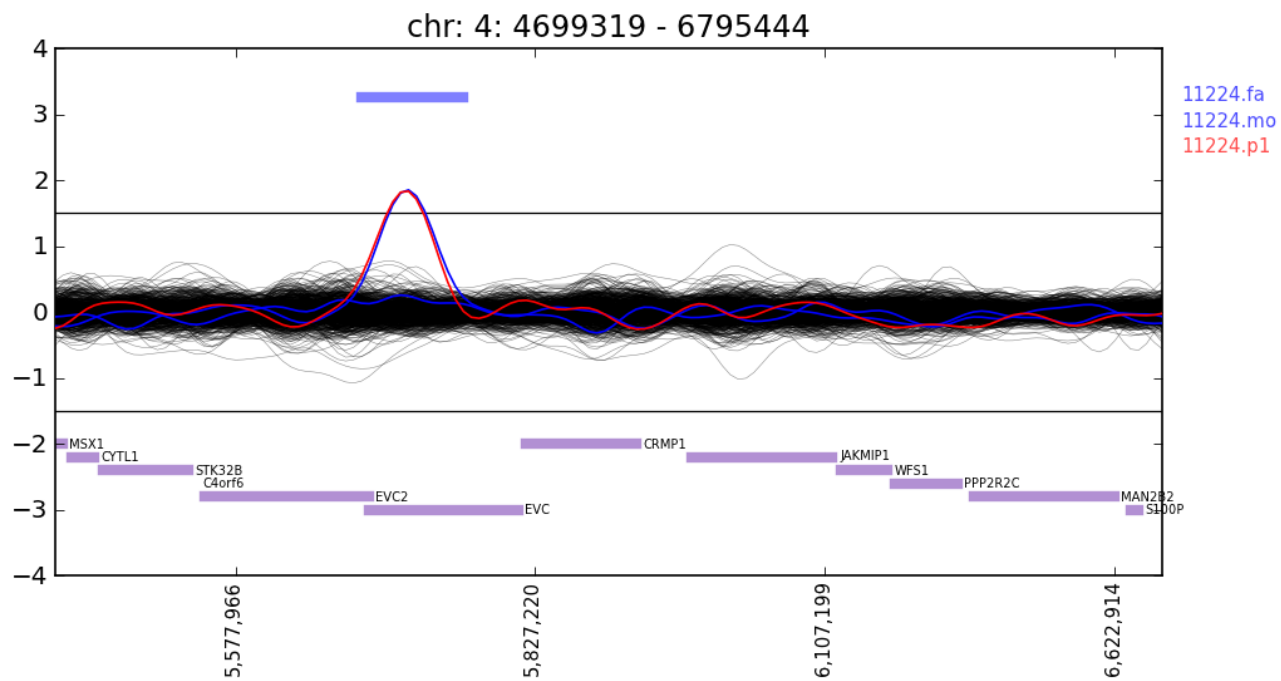


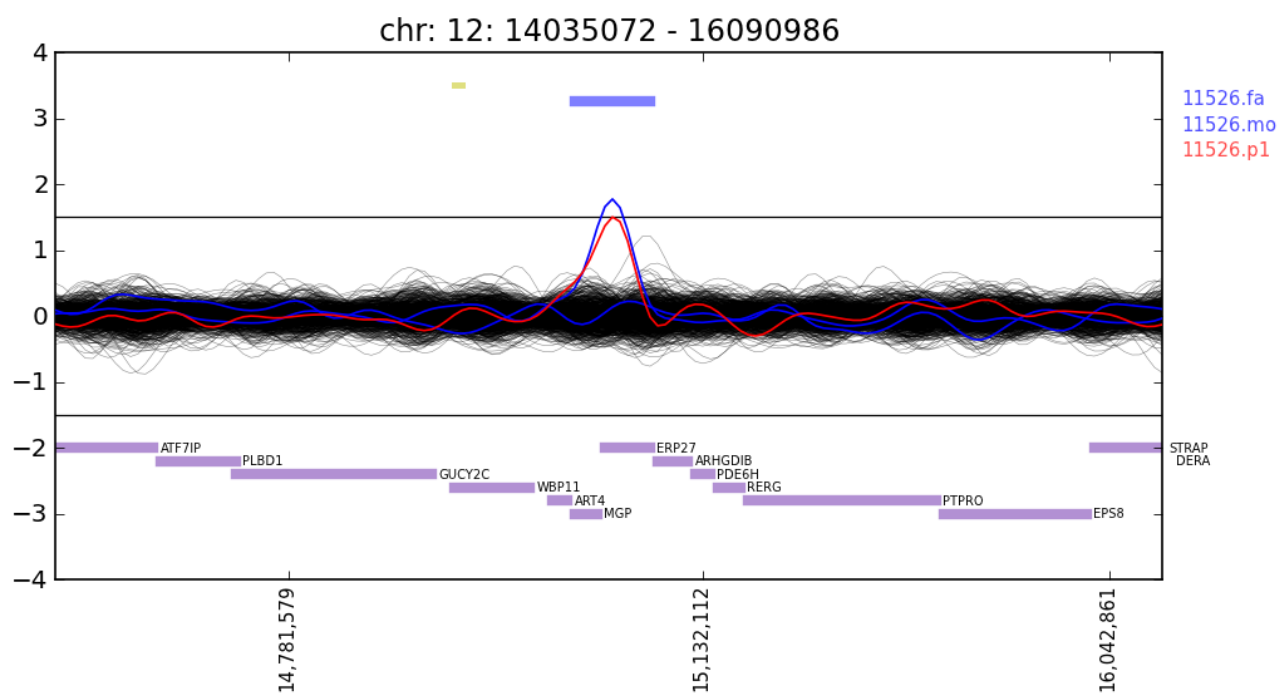
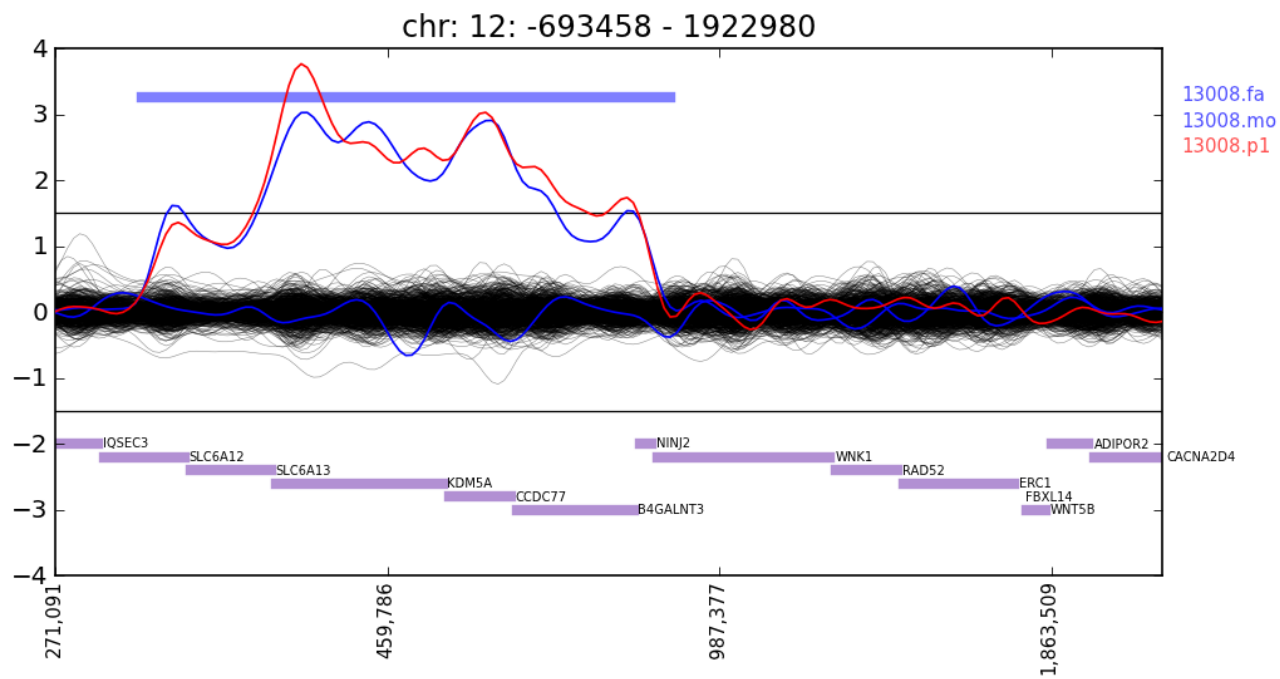
A chr: 16: 20763689 - 23538986

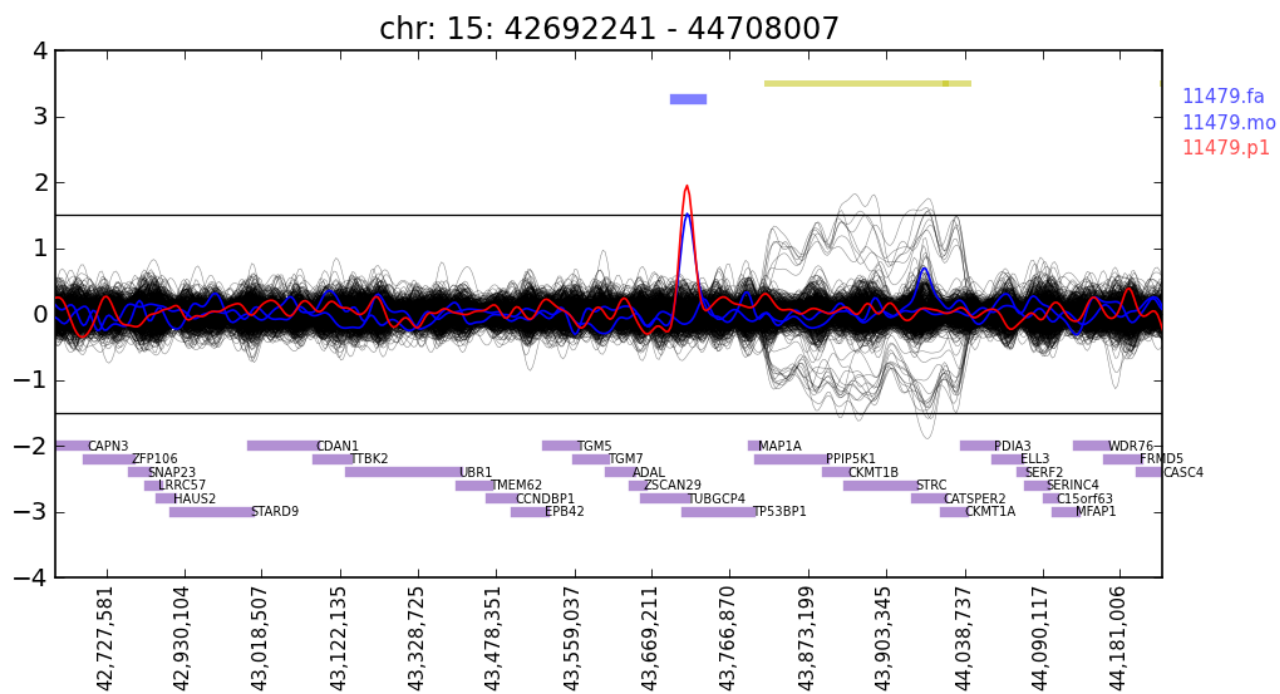
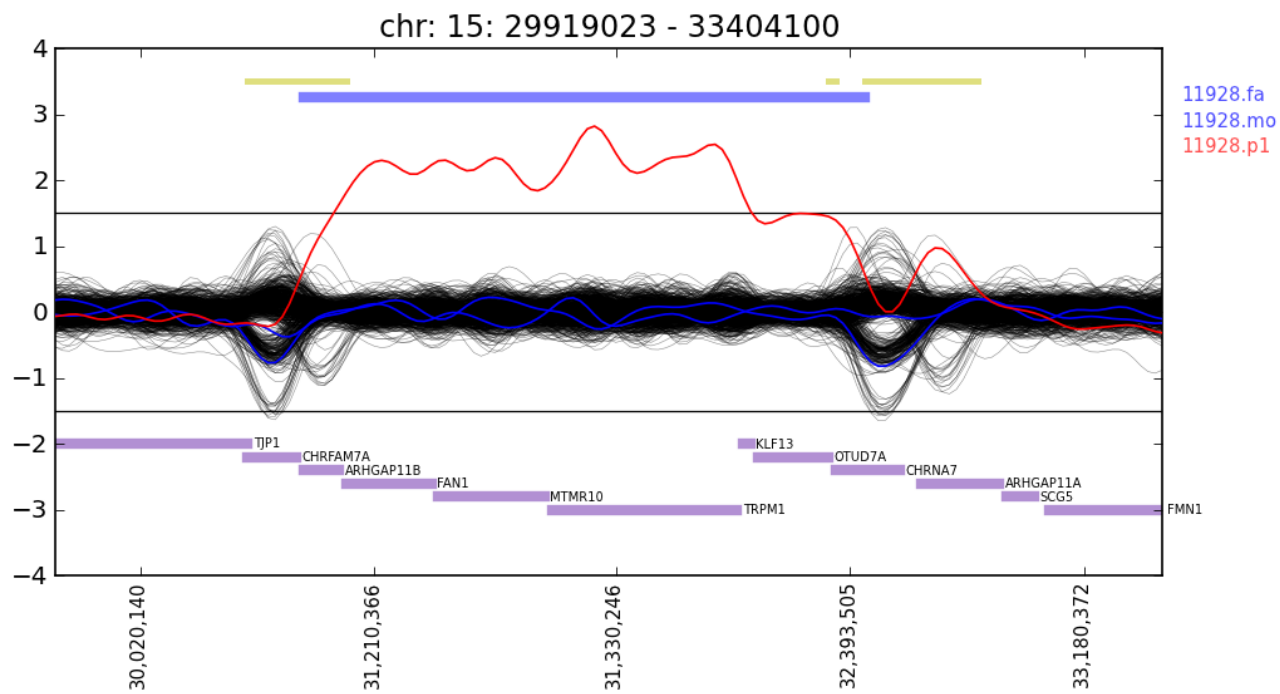


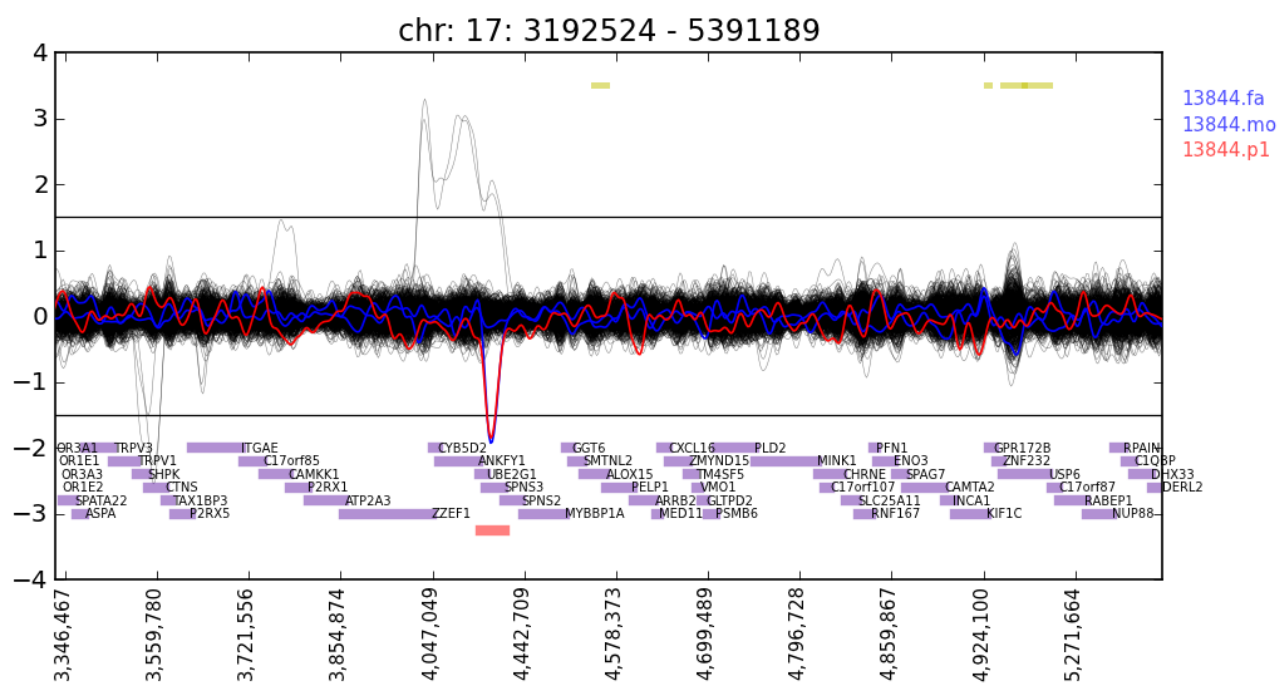
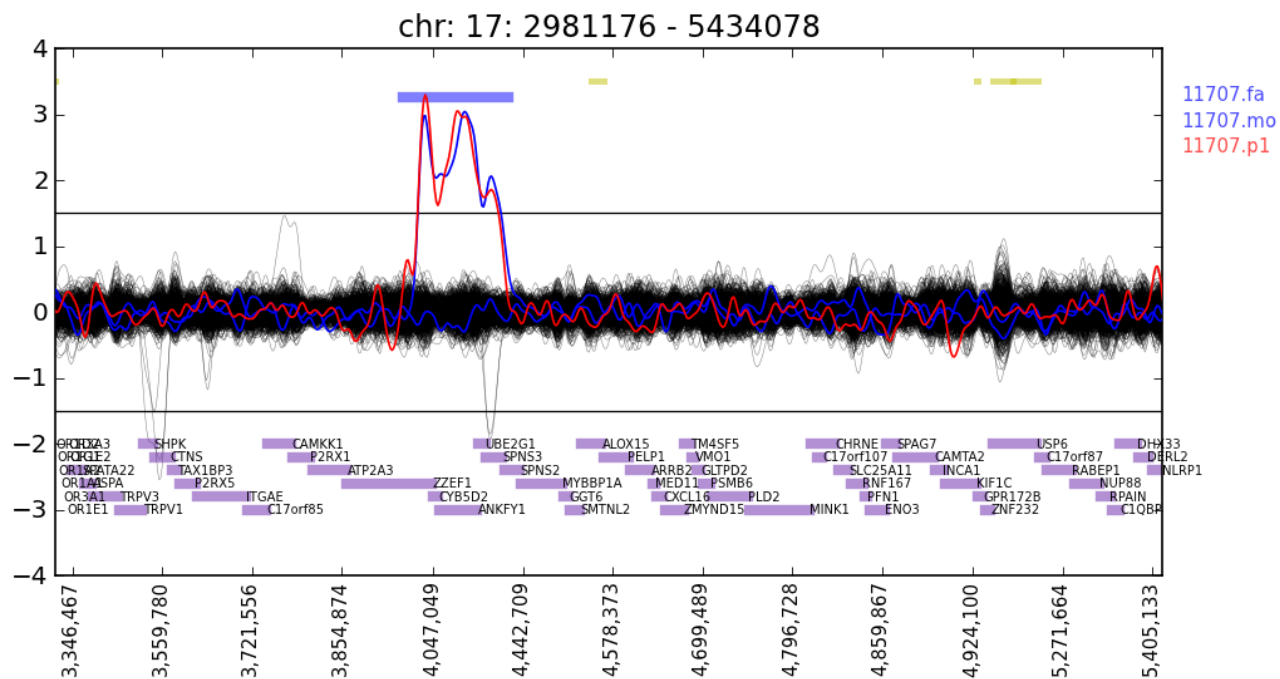


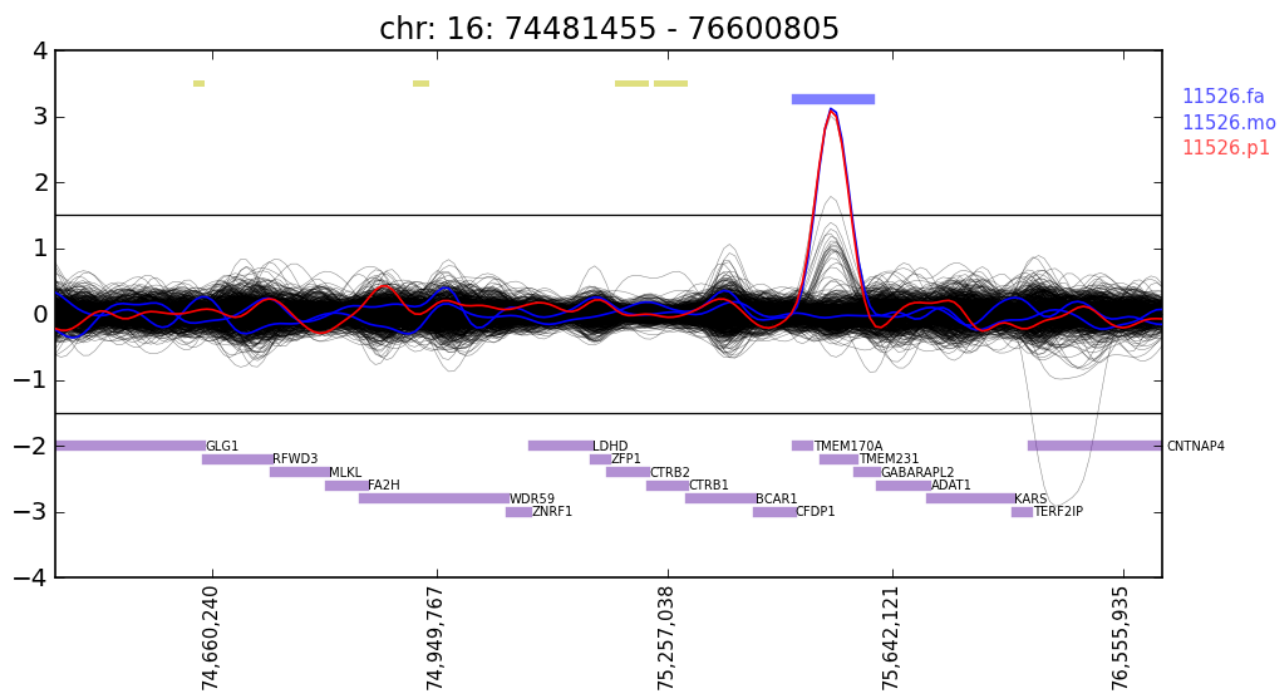
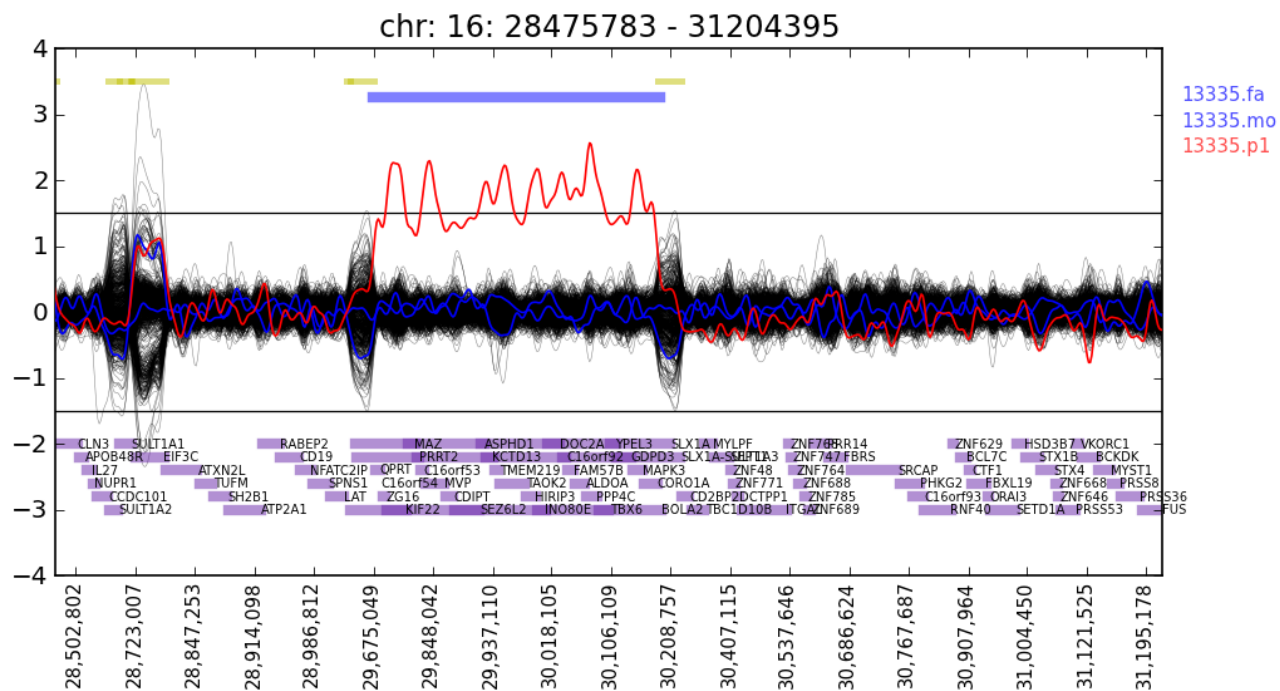












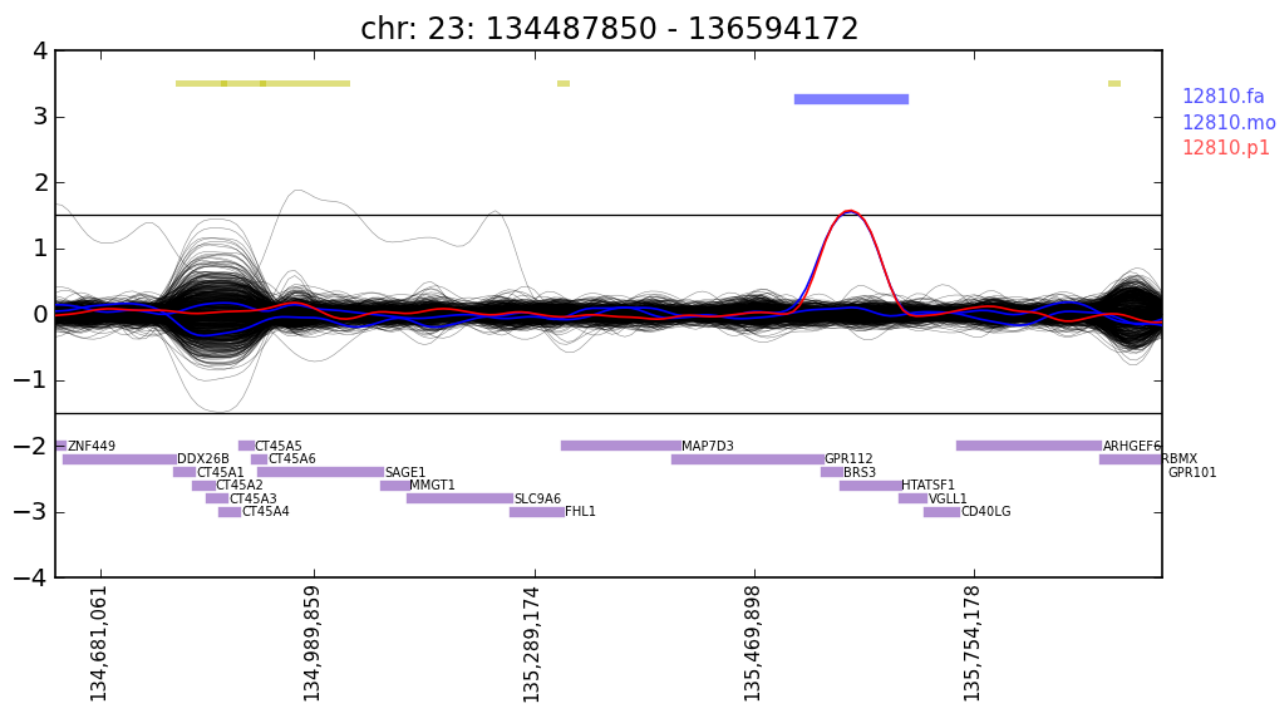
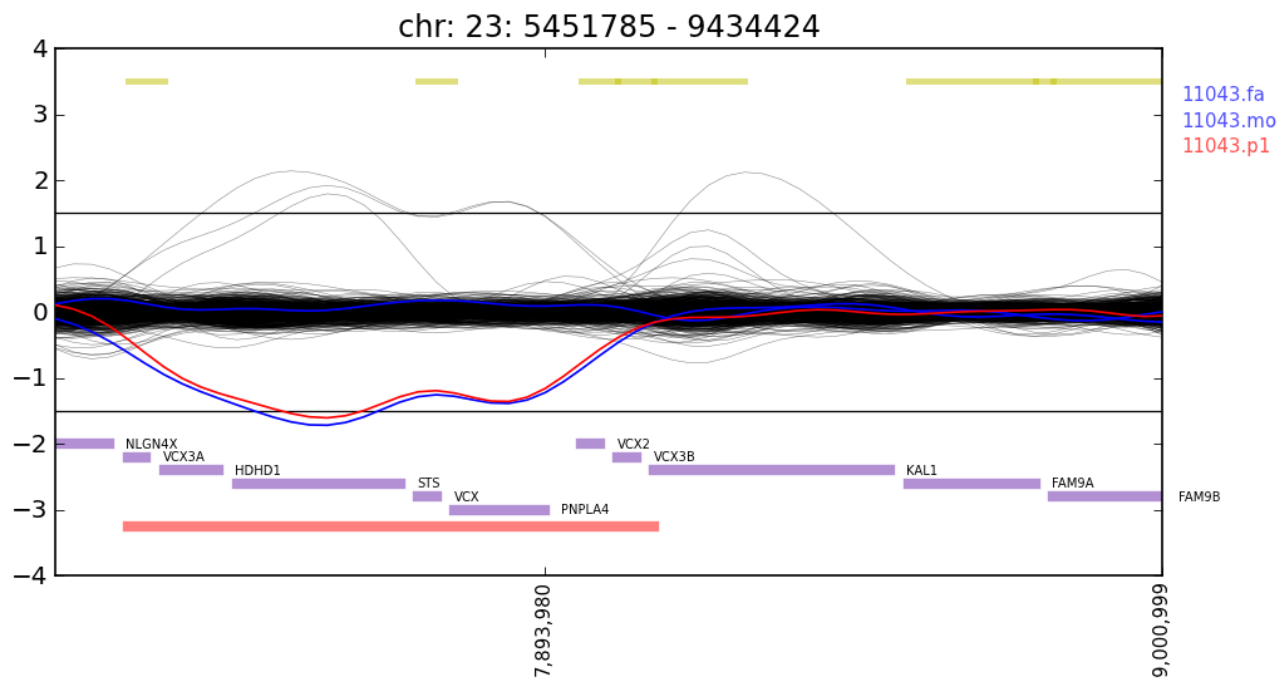


Figure S9: ExomeCNV results for two references from different cohorts

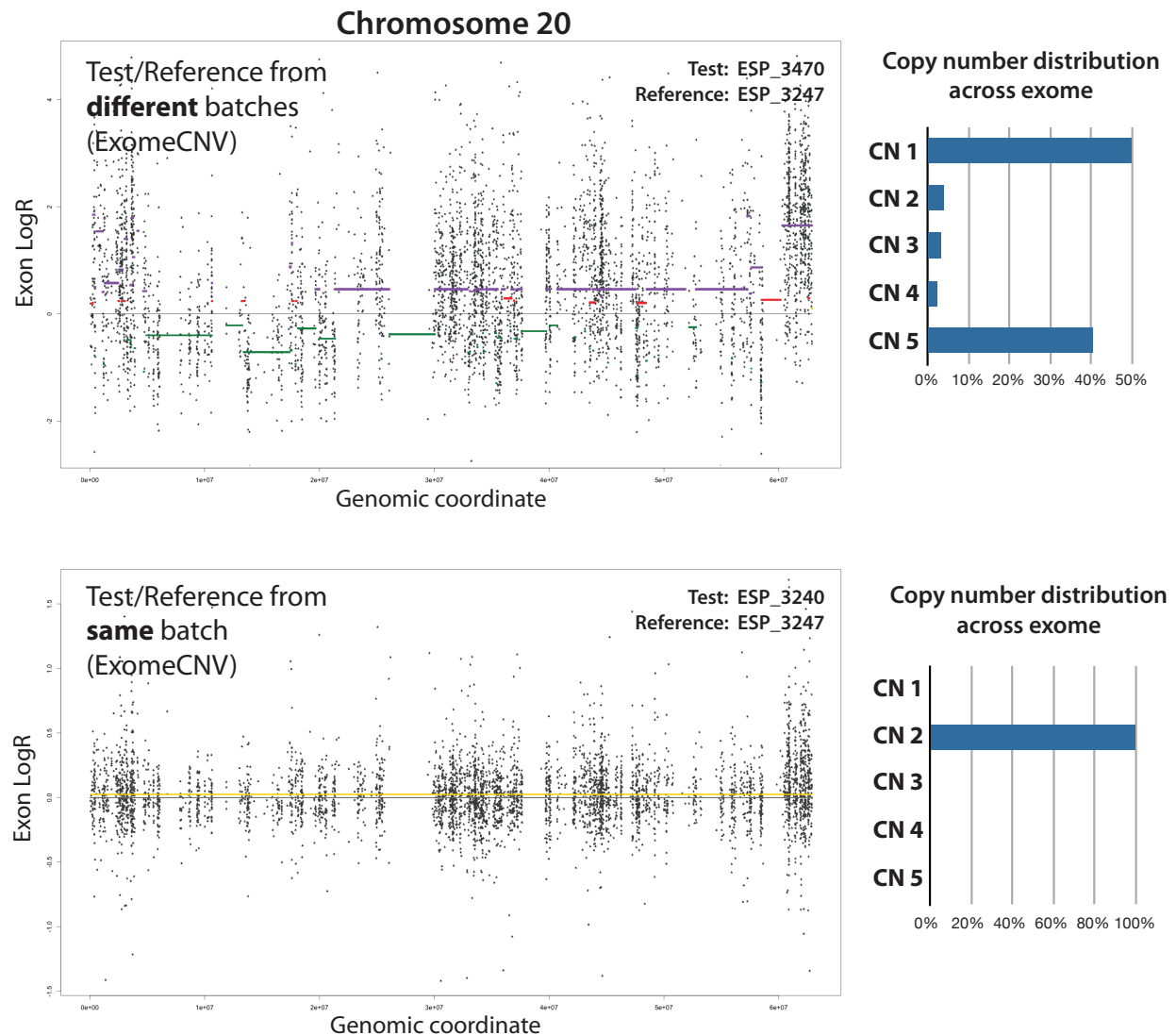


Figure S9: ExomeCNV results for two references from different cohorts

Top: Comparison of two ESP exomes from differing cohorts. Plot shows ExomeCNV LogR output for chromosome 20 and colored bars indicate location of altered copy number. A biologically implausible fraction of the exome (96.6%) is marked as non-diploid (bar chart, top right).

Bottom: Using the SVD algorithm, we matched the same reference (ESP_3247) to a sample from the same cohort/experimental batch. Accordingly, ExomeCNV was less influenced by systematic noise stemming from the exome capture, and marked a much more realistic 99.6% of the exome as diploid.

Figure S10: ExomeCNV and CoNIFER genotyping comparison summary

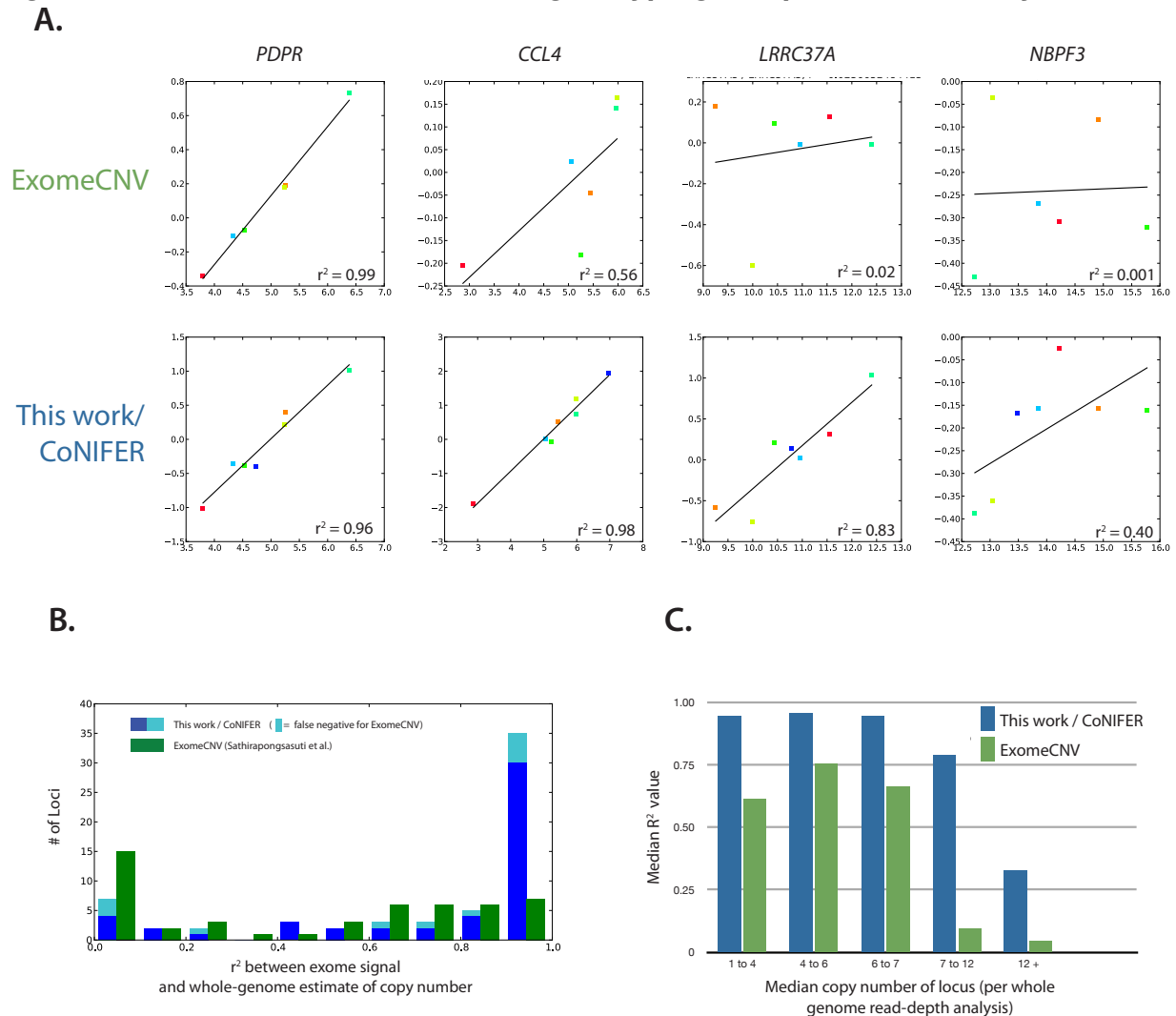


Figure S10: ExomeCNV and CoNIFER genotyping comparison summary

a) Comparison of genotyping correlation between ExomeCNV LogR value (y-axis; top row) and SVD-ZRPKM value (y-axis, bottom row) vs. absolute copy number established by whole-genome read-depth (x-axis, both rows; Sudmant et al., 2010) for four selected loci. b) Distribution of r^2 values across 62 genotyped CNP loci: green bars represent ExomeCNV results (median $r^2 = 0.57$); dark blue bars are the same loci assayed using this work's algorithm (median $r^2 = 0.92$), while light blue bars represent loci which could not be assayed using ExomeCNV (11 loci). c) Median r^2 correlations for ExomeCNV and our algorithm, binned by the median copy number of each CNP locus.

Table S1: Precision of HapMap Calls

Sample	hg19 Coordinates (chr - start - stop)			Call Type	Reciprocal Overlap (%)	Annotation
NA15510	9	108,380,239	109,692,970	duplication	0%	Rare
NA19240	12	133,659,688	133,727,740	duplication	15%	Rare
NA15510	7	99,507,187	99,627,998	duplication	51%	Rare
NA18517	16	21,396,577	21,756,357	duplication	55%	Rare
NA15510	3	19,492,646	21,465,556	duplication	56%	Rare
NA19129	6	29,910,533	30,043,566	deletion	83%	Rare
NA18517	4	68,788,472	69,057,034	duplication	85%	Rare
NA15510	1	155,227,075	155,264,543	duplication	95%	Rare
NA19240	17	20,355,663	20,356,920	duplication	3%	CNP
NA12878	19	46,543,414	46,663,725	deletion	5%	CNP
NA19240	15	20,649,180	20,667,952	duplication	5%	CNP
NA19240	17	34,416,017	34,432,705	duplication	6%	CNP
NA19240	17	34,624,770	34,641,537	duplication	6%	CNP
NA19240	17	34,523,196	34,539,971	duplication	6%	CNP
NA19240	17	18,392,254	18,395,859	duplication	7%	CNP
NA19129	22	21,063,583	21,067,678	duplication	7%	CNP
NA19240	15	43,891,024	43,903,182	deletion	9%	CNP
NA18517	5	69,716,795	69,741,679	duplication	10%	CNP
NA19240	17	25,970,550	25,976,008	duplication	10%	CNP
NA19129	8	6,835,291	7,092,762	duplication	17%	CNP
NA12878	3	195,452,668	195,474,251	deletion	22%	CNP
NA18517	5	70,297,524	70,357,135	duplication	22%	CNP
NA19240	17	39,535,207	39,551,622	duplication	24%	CNP
NA18517	5	68,862,415	68,886,194	duplication	25%	CNP
NA12878	22	22,328,728	23,249,131	deletion	30%	CNP
NA18517	6	35,748,887	35,787,224	duplication	31%	CNP
NA19240	6	29,643,162	30,041,255	duplication	32%	CNP
NA12878	17	34,624,258	34,748,588	deletion	36%	CNP
NA18517	2	89,049,497	89,292,723	duplication	37%	CNP
NA12878	17	34,499,213	34,583,568	deletion	41%	CNP
NA15510	19	33,481,419	33,503,630	duplication	43%	CNP
NA19240	14	106,385,019	107,114,522	duplication	47%	CNP
NA19129	19	41,313,388	41,388,115	duplication	47%	CNP
NA15510	7	100,319,584	100,336,236	deletion	50%	CNP
NA19240	10	124,362,280	124,377,856	duplication	50%	CNP
NA19240	22	22,676,730	22,937,826	duplication	50%	CNP
NA19240	7	100,319,584	100,336,236	deletion	50%	CNP
NA15510	15	22,077,863	22,739,445	duplication	50%	CNP
NA19240	16	72,097,154	72,109,118	duplication	52%	CNP
NA19240	1	110,173,564	110,260,048	duplication	56%	CNP
NA19240	15	22,082,360	22,490,341	duplication	56%	CNP
NA15510	22	23,010,695	23,155,078	duplication	57%	CNP
NA19129	2	89,246,519	90,260,548	deletion	58%	CNP
NA15510	5	180,335,585	180,431,762	deletion	59%	CNP
NA12878	22	24,325,587	24,431,968	duplication	62%	CNP
NA15510	14	106,235,274	106,347,756	deletion	62%	CNP
NA15510	19	41,313,713	41,388,115	deletion	64%	CNP
NA18517	16	72,092,152	72,120,701	duplication	68%	CNP

Sample	hg19 Coordinates (chr - start - stop)			Call Type	Reciprocal Overlap (%)	Annotation
NA19240	19	33,481,419	33,517,538	duplication	70%	CNP
NA18517	19	43,514,149	43,857,918	duplication	74%	CNP
NA19129	22	21,827,328	21,905,265	duplication	76%	CNP
NA19240	3	196,528,797	196,547,438	duplication	76%	CNP
NA19240	14	20,215,586	20,404,761	duplication	77%	CNP
NA18517	14	106,235,274	106,332,116	duplication	81%	CNP
NA19240	22	23,029,156	23,235,998	duplication	83%	CNP
NA19129	1	155,178,595	155,218,094	duplication	92%	CNP
NA19129	10	124,340,381	124,358,613	duplication	95%	CNP
NA12878	5	180,377,236	180,430,876	duplication	98%	CNP
NA18517	1	161,479,609	161,643,863	deletion	100%	CNP
NA12878	1	25,627,436	25,633,220	deletion	-	CNP
NA12878	4	69,403,342	69,434,202	deletion	-	CNP
NA12878	7	141,754,553	141,759,766	deletion	-	CNP
NA12878	11	60,971,578	61,018,753	deletion	-	CNP
NA12878	14	106,233,850	106,332,116	deletion	-	CNP
NA12878	17	18,390,960	18,396,173	deletion	-	CNP
NA12878	17	20,354,793	20,361,697	deletion	-	CNP
NA12878	17	25,969,263	25,976,008	deletion	-	CNP
NA12878	17	34,398,310	34,495,524	deletion	-	CNP
NA15510	1	25,617,131	25,643,570	deletion	-	CNP
NA15510	1	25,712,201	25,729,237	deletion	-	CNP
NA15510	19	39,329,036	39,334,554	deletion	-	CNP
NA18517	22	24,373,137	24,431,968	deletion	-	CNP
NA12878	9	39,102,493	39,149,974	duplication	-	CNP
NA15510	3	196,509,517	196,539,722	duplication	-	CNP
NA18517	19	43,236,935	43,439,921	duplication	-	CNP
NA19129	22	24,572,078	24,581,897	duplication	-	CNP
NA19240	6	31,969,272	31,970,032	duplication	3%	HLA locus
NA19240	6	31,999,819	32,003,054	duplication	13%	HLA locus
NA12878	6	31,948,780	31,961,254	deletion	19%	HLA locus
NA12878	6	31,978,929	31,996,054	deletion	26%	HLA locus
NA19240	6	31,982,819	31,996,054	duplication	27%	HLA locus
NA19240	6	31,949,884	31,961,254	duplication	31%	HLA locus
NA19129	6	31,948,780	31,961,254	deletion	-	HLA locus
NA19129	6	31,969,272	31,970,317	deletion	-	HLA locus
NA19129	6	31,982,622	31,995,709	deletion	-	HLA locus
NA19129	6	31,999,327	32,003,054	deletion	-	HLA locus

Table S2: Sensitivity vs. Conrad *et al.* aCGH Calls

Conrad CNVR	Sample	hg19 Coordinates (chr - start - stop)		Call Type	# exome probes	Genes	Annotation	Discovered?
CNVR1313.1	NA15510	3	19,535,653	20,638,501	duplication	47	KCNH8, EFHB, RAB5A, KAT2B, SGOL1	Rare Yes
CNVR1952.1	NA18517	4	68,788,730	69,016,101	duplication	17	TMPRSS11A	Rare Yes
CNVR3507.1	NA15510	7	99,564,133	99,625,411	duplication	6	AZGP1, ZKSCAN1	Rare Yes
CNVR5791.2	NA19240	12	133,717,202	133,779,425	duplication	8	ZNF140, ZNF10, ZNF268	Rare Yes
CNVR6668.5	NA18517	16	21,523,044	21,946,347	duplication	38	METTL9, GSF6, OTOA	Rare Yes
CNVR4393.2	NA12878	9	91,963,403	92,343,382	duplication	27		Conrad False Positive --
CNVR6152.2	NA15510	14	50,101,898	50,942,527	duplication	161		Conrad False Positive --
CNVR6631.1	NA19129	16	8,939,807	8,987,025	duplication	4		Conrad False Positive --
CNVR2861.1	NA18517	6	35,754,791	35,766,680	duplication	4	CLPS	CNP Yes
CNVR3509.1	NA19240	7	100,327,862	100,337,886	deletion	6	EPO	CNP Yes
CNVR8114.1	NA12878	22	24,344,211	24,404,564	duplication	7	GSTT1	CNP Yes
CNVR339.1	NA15510	1	144,950,054	145,080,140	duplication	7	PDE4DIP	CNP No
CNVR339.1	NA12878	1	144,948,283	145,080,140	duplication	7	PDE4DIP	CNP No
CNVR339.1	NA18517	1	144,956,694	145,083,994	duplication	4	PDE4DIP	CNP No
CNVR759.1	NA19129	2	38,956,285	38,972,493	deletion	5	GALM	CNP No
CNVR759.1	NA12878	2	38,955,877	38,972,493	deletion	5	GALM	CNP No
CNVR759.1	NA19240	2	38,955,946	38,972,830	deletion	5	GALM	CNP No
CNVR759.1	NA18517	2	38,956,285	38,972,937	deletion	5	GALM	CNP No
CNVR2719.1	NA18517	5	180,374,610	180,431,110	duplication	11	BTNL8/3	CNP No
CNVR2719.1	NA12878	5	180,376,223	180,430,715	duplication	8	BTNL8/3	CNP No
CNVR2728.1	NA19240	6	257,100	382,983	duplication	8	DUSP22	CNP No
CNVR2728.1	NA12878	6	255,650	382,508	duplication	8	DUSP22	CNP No
CNVR2728.1	NA19129	6	254,458	382,453	duplication	8	DUSP22	CNP No
CNVR2728.1	NA18517	6	257,309	384,408	duplication	8	DUSP22	CNP No
CNVR2861.1	NA19129	6	35,754,996	35,766,570	duplication	4	CLPS	CNP No
CNVR2861.1	NA19240	6	35,754,736	35,766,415	duplication	5	CLPS	CNP No
CNVR2861.1	NA12878	6	35,754,591	35,766,415	duplication	5	CLPS	CNP No
CNVR4912.1	NA18517	10	124,360,402	124,376,587	deletion	8	DMBT1	CNP No
CNVR4912.1	NA12878	10	124,360,512	124,376,427	deletion	7	DMBT1	CNP No
CNVR4912.2	NA18517	10	124,342,529	124,351,752	duplication	6	DMBT1	CNP No
CNVR4912.3	NA12878	10	124,342,337	124,360,459	duplication	14	DMBT1	CNP No
CNVR4912.3	NA19240	10	124,342,556	124,360,682	duplication	15	DMBT1	CNP No

Conrad CNVR	Sample	hg19 Coordinates (chr - start - stop)		Call Type	# exome probes	Genes	Annotation	Discovered?	
CNVR4912.3	NA19129	10	124,341,208	124,358,673	duplication	15	DMBT1	CNP	No
CNVR5179.1	NA19129	11	55,366,154	55,452,992	deletion	6	OR4	High Diversity	No
CNVR5179.1	NA12878	11	55,365,742	55,453,061	deletion	6	OR4	High Diversity	No
CNVR6072.4	NA19240	14	20,177,270	20,422,582	duplication	6	OR4	High Diversity	No
CNVR6072.5	NA18517	14	20,289,680	20,424,616	deletion	4	OR4	High Diversity	No
CNVR7095.1	NA18517	17	39,382,871	39,395,430	deletion	3	KRTAP	High Diversity	No
CNVR7097.1	NA19240	17	39,507,055	39,525,624	deletion	6	KRT34	CNP	No
CNVR7098.1	NA19240	17	39,532,301	39,539,205	duplication	7	KRT34	CNP	No
CNVR7673.1	NA12878	19	46,622,831	46,628,261	deletion	3	IGFL3	CNP	No
CNVR7702.1	NA19129	19	52,131,804	52,148,913	deletion	8	SIGLEC14	CNP	No
CNVR7708.1	NA19240	19	53,322,989	53,361,358	duplication	3	ZNF468	High Diversity	No
CNVR7763.1	NA19129	20	1,552,963	1,595,689	deletion	4	SIRBP1	CNP	No
CNVR7763.1	NA18517	20	1,552,963	1,595,689	deletion	4	SIRBP1	CNP	No
CNVR7763.1	NA19240	20	1,552,963	1,595,689	deletion	4	SIRBP1	CNP	No
CNVR8114.1	NA19129	22	24,344,595	24,404,495	duplication	7	GSTT1	CNP	No
CNVR8114.7	NA19240	22	24,364,568	24,404,701	duplication	7	GSTT1	CNP	No
CNVR8114.7	NA15510	22	24,371,095	24,404,715	duplication	7	GSTT1	CNP	No
CNVR4841.3	NA19129	10	89,189,706	89,275,944	duplication	4		CNP	No

Table S3(a): Rare CNVs in 122 ASD probands

Sample	hg19 Coordinates (chr - start - stop)			Inheritance	SNP Microarray
11696.p1	chr3	37,170,553	37,494,050	de novo	Validated
11711.p1	chr5	175,913,355	175,956,388	de novo	Did Not Validate
11218.p1	chr5	175,913,355	175,956,645	de novo	Did Not Validate
12581.p1	chr9	140,671,069	141,015,333	de novo	Validated
13726.p1	chr11	56,510,303	60,235,941	de novo	Validated
11928.p1	chr15	30,919,023	32,404,100	de novo	Validated
13335.p1	chr16	29,475,783	30,204,395	de novo	Validated
13815.p1	chr16	75,681,780	76,532,591	de novo	Validated
13844.p1	chr1	47,501,688	47,546,339	Matching event in parent	No Data
12810.p1	chr1	86,965,336	87,043,755	Matching event in parent	Validated
11715.p1	chr1	185,089,514	185,137,530	Matching event in parent	Validated
12667.p1	chr1	185,092,988	185,144,245	Matching event in parent	Validated
11895.p1	chr1	206,241,532	206,557,431	Matching event in parent	Validated
12130.p1	chr1	206,241,532	206,557,431	Matching event in parent	Validated
11707.p1	chr1	207,307,748	207,640,257	Matching event in parent	Validated
11064.p1	chr2	33,622,199	36,691,798	Matching event in parent	Validated
11472.p1	chr2	44,508,525	44,549,039	Matching event in parent	Validated
11895.p1	chr2	86,276,282	86,677,085	Matching event in parent	Validated
11023.p1	chr2	198,285,151	198,593,302	Matching event in parent	Validated
11023.p1	chr2	209,027,927	209,104,727	Matching event in parent	Validated
11526.p1	chr3	47,539,775	47,619,418	Matching event in parent	Did Not Validate
13741.p1	chr3	100,277,248	100,455,560	Matching event in parent	No Data
11722.p1	chr3	100,287,665	100,451,516	Matching event in parent	Validated
11303.p1	chr3	100,295,768	100,447,702	Matching event in parent	Validated
13335.p1	chr3	141,712,379	142,090,170	Matching event in parent	Validated
11262.p1	chr3	151,461,880	152,018,156	Matching event in parent	Validated
12565.p1	chr3	151,461,880	152,018,156	Matching event in parent	Validated
11193.p1	chr3	196,454,793	196,626,933	Matching event in parent	Did Not Validate
11224.p1	chr4	5,699,319	5,795,444	Matching event in parent	Validated
11190.p1	chr4	107,845,110	108,935,744	Matching event in parent	Validated
11056.p1	chr5	32,093,012	32,235,235	Matching event in parent	Validated
11788.p1	chr5	32,093,012	32,235,235	Matching event in parent	Validated
11480.p1	chr5	32,097,384	32,242,233	Matching event in parent	Validated
13733.p1	chr5	80,659,606	80,742,741	Manual inspection of parental SVD-ZRPKM values	No Data
11469.p1	chr5	112,899,555	113,740,553	Matching event in parent	Validated
12130.p1	chr5	112,902,788	113,740,553	Matching event in parent	Validated
13102.p1	chr5	138,456,723	138,699,611	Manual inspection of parental SVD-ZRPKM values	No Data
11303.p1	chr5	138,643,104	138,700,432	Matching event in parent	Did Not Validate
11198.p1	chr5	156,378,522	156,482,544	Manual inspection of parental SVD-ZRPKM values	Did Not Validate
11480.p1	chr6	25,923,922	26,368,495	Matching event in parent	Validated
13207.p1	chr6	49,421,297	49,459,988	Manual inspection of parental SVD-ZRPKM values	Did Not Validate
13822.p1	chr6	55,142,177	55,300,561	Manual inspection of parental SVD-ZRPKM values	No Data
11425.p1	chr6	56,882,004	56,993,638	Manual inspection of parental SVD-ZRPKM values	Validated
11459.p1	chr6	88,311,501	88,374,577	Matching event in parent	Validated
12212.p1	chr6	107,420,452	107,824,999	Matching event in parent	Validated
11518.p1	chr6	168,317,768	168,442,831	Matching event in parent	Validated
13726.p1	chr6	168,319,414	168,443,396	Matching event in parent	No Data
12933.p1	chr6	168,319,414	168,711,126	Matching event in parent	Validated

Sample	hg19 Coordinates (chr - start - stop)			Inheritance	SNP Microarray
11472.p1	chr6	168,319,414	168,711,964	Matching event in parent	Validated
11722.p1	chr6	168,323,535	168,439,409	Matching event in parent	Validated
12667.p1	chr6	168,323,535	168,442,831	Matching event in parent	Validated
11863.p1	chr6	168,323,535	168,458,019	Matching event in parent	Validated
13557.p1	chr6	168,325,684	168,711,126	Matching event in parent	Validated
11398.p1	chr7	11,101,590	12,620,846	Matching event in parent	Validated
11696.p1	chr7	16,834,559	17,838,777	Matching event in parent	Validated
12667.p1	chr7	33,066,428	33,297,022	Manual inspection of parental SVD-ZRPKM values	Validated
11722.p1	chr7	48,285,108	48,431,736	Manual inspection of parental SVD-ZRPKM values	Validated
13822.p1	chr7	142,659,290	142,960,678	Matching event in parent	No Data
11526.p1	chr7	142,659,290	142,961,260	Matching event in parent	Validated
11218.p1	chr7	142,723,286	142,960,678	Matching event in parent	Validated
11843.p1	chr7	152,740,571	154,664,403	Matching event in parent	Did Not Validate
11141.p1	chr8	13,071,835	15,480,758	Matching event in parent	Validated
12130.p1	chr8	15,601,046	16,032,809	Manual inspection of parental SVD-ZRPKM values	Validated
11556.p1	chr8	15,601,046	16,035,497	Matching event in parent	Validated
11753.p1	chr8	29,197,614	29,953,044	Matching event in parent	Did Not Validate
12933.p1	chr8	29,197,614	29,959,489	Matching event in parent	Did Not Validate
13222.p1	chr8	29,197,614	29,959,489	Matching event in parent	Did Not Validate
11660.p1	chr8	29,202,886	29,959,489	Matching event in parent	Did Not Validate
11722.p1	chr8	29,940,362	30,335,353	Matching event in parent	Validated
11303.p1	chr8	98,725,889	98,973,758	Matching event in parent	Did Not Validate
11638.p1	chr8	98,731,276	98,863,702	Matching event in parent	Did Not Validate
11479.p1	chr8	98,731,276	98,943,750	Matching event in parent	Did Not Validate
11023.p1	chr8	98,735,106	98,900,470	Matching event in parent	Did Not Validate
11414.p1	chr8	98,735,106	98,954,127	Matching event in parent	Did Not Validate
11827.p1	chr8	98,735,106	98,954,127	Matching event in parent	Did Not Validate
13741.p1	chr8	102,678,816	103,662,619	Matching event in parent	No Data
12378.p1	chr9	134,360,072	134,458,089	Matching event in parent	Validated
11218.p1	chr9	139,327,606	139,354,326	Matching event in parent	Did Not Validate
13844.p1	chr10	1,598,506	3,824,406	Matching event in parent	No Data
13031.p1	chr10	75,005,679	75,034,352	Matching event in parent	Did Not Validate
12378.p1	chr10	82,040,435	82,122,829	Matching event in parent	Validated
12130.p1	chr10	132,965,059	133,761,295	Matching event in parent	Did Not Validate
12118.p1	chr10	133,106,473	134,523,960	Matching event in parent	Validated
11498.p1	chr10	135,233,529	135,368,588	Matching event in parent	Validated
11148.p1	chr10	135,233,529	135,372,455	Matching event in parent	Validated
11707.p1	chr10	135,340,899	135,372,455	Matching event in parent	Validated
11498.p1	chr10	135,370,262	135,372,455	Matching event in parent	Validated
11964.p1	chr11	14,856,527	14,989,400	Matching event in parent	Validated
12430.p1	chr11	31,128,044	31,451,948	Matching event in parent	Validated
11141.p1	chr11	84,822,704	85,366,752	Matching event in parent	Did Not Validate
13532.p1	chr11	95,555,662	95,724,887	Matching event in parent	Did Not Validate
11452.p1	chr11	95,560,949	95,724,887	Matching event in parent	Did Not Validate
13008.p1	chr12	306,542	922,980	Matching event in parent	Validated
11526.p1	chr12	15,035,072	15,090,986	Manual inspection of parental SVD-ZRPKM values	Validated
13793.p1	chr12	19,440,406	19,626,289	Matching event in parent	No Data
13741.p1	chr12	50,536,856	50,642,534	Matching event in parent	No Data
12581.p1	chr12	112,167,609	112,323,840	Matching event in parent	Validated
11193.p1	chr13	21,720,943	21,950,794	Matching event in parent	Did Not Validate

Sample	hg19 Coordinates (chr - start - stop)			Inheritance	SNP Microarray
11083.p1	chr13	50,118,872	50,237,331	Matching event in parent	Validated
11257.p1	chr13	115,004,824	115,048,418	Matching event in parent	Validated
13530.p1	chr14	67,940,136	68,276,006	Matching event in parent	Validated
13533.p1	chr14	74,512,762	74,551,696	Matching event in parent	Validated
11479.p1	chr15	43,692,241	43,708,007	Matching event in parent	Did Not Validate
13812.p1	chr15	57,555,309	57,815,839	Matching event in parent	No Data
13415.p1	chr15	57,555,309	57,816,949	Matching event in parent	Validated
11556.p1	chr15	89,760,350	89,817,535	Matching event in parent	Validated
12430.p1	chr15	100,269,327	100,537,794	Matching event in parent	Did Not Validate
13812.p1	chr16	3,656,474	3,725,398	Matching event in parent	No Data
11834.p1	chr16	21,763,689	22,538,986	Matching event in parent	Validated
11526.p1	chr16	75,481,455	75,600,805	Matching event in parent	Validated
11184.p1	chr16	81,171,041	81,194,510	Matching event in parent	Validated
11964.p1	chr16	84,402,221	84,474,564	Matching event in parent	Validated
13335.p1	chr17	644,540	708,487	Matching event in parent	Validated
11707.p1	chr17	3,981,176	4,434,078	Matching event in parent	Validated
13844.p1	chr17	4,192,524	4,391,189	Matching event in parent	No Data
13733.p1	chr17	10,322,203	10,358,127	Matching event in parent	No Data
13409.p1	chr17	72,322,488	72,733,256	Matching event in parent	Validated
12667.p1	chr18	39,613,789	40,503,728	Matching event in parent	Validated
13668.p1	chr18	47,813,109	48,252,545	Matching event in parent	No Data
13494.p1	chr18	76,873,240	77,132,882	Matching event in parent	Validated
13812.p1	chr19	4,212,596	4,233,304	Matching event in parent	No Data
11141.p1	chr19	4,212,596	4,249,325	Manual inspection of parental SVD-ZRPKM values	Did Not Validate
11205.p1	chr19	4,215,974	4,233,304	Matching event in parent	Did Not Validate
13530.p1	chr19	4,215,974	4,234,821	Matching event in parent	Did Not Validate
12118.p1	chr19	11,319,586	11,363,226	Matching event in parent	Did Not Validate
11193.p1	chr19	17,440,933	17,452,512	Matching event in parent	Did Not Validate
13116.p1	chr19	45,822,778	45,909,976	Matching event in parent	Validated
13815.p1	chr19	57,792,174	57,967,854	Matching event in parent	No Data
11013.p1	chr20	6,100,050	8,352,097	Matching event in parent	Validated
13844.p1	chr21	35,446,000	36,079,687	Matching event in parent	No Data
13593.p1	chr21	37,635,843	37,710,244	Matching event in parent	Did Not Validate
12810.p1	chr22	32,495,169	32,788,346	Matching event in parent	Validated
11947.p1	chr22	40,711,286	41,077,932	Matching event in parent	Validated
11653.p1	chr22	41,568,502	41,634,889	Matching event in parent	Validated
11043.p1	chr23	6,451,785	8,434,424	Matching event in parent	No Data
12810.p1	chr23	135,487,850	135,594,172	Matching event in parent	No Data

Table S3(b): CNPs in 122 ASD probands

Sample	hg19 Coordinates (chr - start - stop)			Gene	Processed Pseudogene	Type of Event	Validation
13812.p1	chr6	32,609,086	32,632,844			HLA	NA
11523.p1	chr14	106,329,088	106,376,628			IgHeavy	NA
12114.p1	chr2	179,255,799	179,315,757	PRKA	Y	CNP	NA
11141.p1	chr2	179,296,823	179,315,757	PRKA	Y	CNP	NA
11190.p1	chr2	179,296,823	179,315,757	PRKA	Y	CNP	NA
12744.p1	chr2	179,296,823	179,315,757	PRKA	Y	CNP	NA
11788.p1	chr2	179,296,823	179,318,347	PRKA	Y	CNP	NA
11707.p1	chr2	179,300,871	179,312,313	PRKA	Y	CNP	NA
11346.p1	chr2	179,300,871	179,315,170	PRKA	Y	CNP	NA
11414.p1	chr2	179,300,871	179,315,170	PRKA	Y	CNP	NA
11452.p1	chr2	179,300,871	179,315,170	PRKA	Y	CNP	NA
11009.p1	chr2	179,300,871	179,315,757	PRKA	Y	CNP	NA
11504.p1	chr2	179,300,871	179,315,757	PRKA	Y	CNP	NA
11571.p1	chr2	179,300,871	179,315,757	PRKA	Y	CNP	NA
11834.p1	chr2	179,300,871	179,315,757	PRKA	Y	CNP	NA
12810.p1	chr2	179,300,871	179,315,757	PRKA	Y	CNP	NA
13726.p1	chr2	179,300,871	179,318,347	PRKA	Y	CNP	NA
11013.p1	chr2	179,300,871	179,320,878	PRKA	Y	CNP	NA
11843.p1	chr2	179,306,336	179,312,313	PRKA	Y	CNP	NA
11587.p1	chr2	179,306,336	179,315,170	PRKA	Y	CNP	NA
11722.p1	chr3	16,635,161	16,640,105	DAZL		CNP	NA
13741.p1	chr3	16,636,039	16,640,105	DAZL		CNP	NA
11471.p1	chr3	16,636,820	16,639,048	DAZL		CNP	NA
13517.p1	chr3	16,636,820	16,640,105	DAZL		CNP	NA
11193.p1	chr5	158,523,981	158,634,904	RNF145	Y	CNP	NA
13812.p1	chr5	158,595,880	158,609,059	RNF145	Y	CNP	NA
11459.p1	chr5	158,600,990	158,697,453	RNF145	Y	CNP	NA
13409.p1	chr5	180,218,633	180,430,876	BTNL8		CNP	NA
11964.p1	chr5	180,375,919	180,420,160	BTNL8		CNP	NA
11224.p1	chr5	180,375,919	180,430,876	BTNL8		CNP	NA
11753.p1	chr5	180,375,919	180,430,876	BTNL8		CNP	NA
12212.p1	chr5	180,375,919	180,430,876	BTNL8		CNP	NA
11064.p1	chr5	180,375,919	180,431,443	BTNL8		CNP	NA
11518.p1	chr5	180,375,919	180,431,443	BTNL8		CNP	NA
11599.p1	chr5	180,375,919	180,431,443	BTNL8		CNP	NA
12249.p1	chr5	180,375,919	180,431,443	BTNL8		CNP	NA
13793.p1	chr5	180,375,919	180,431,443	BTNL8		CNP	NA
12667.p1	chr5	180,376,238	180,430,876	BTNL8		CNP	NA
13335.p1	chr6	35,745,235	35,787,224	CLPS		CNP	NA
11498.p1	chr11	55,339,603	55,419,315	OR4		CNP	NA
13741.p1	chr11	55,339,603	55,419,315	OR4		CNP	NA
11471.p1	chr11	55,339,603	55,433,572	OR4		CNP	NA
11587.p1	chr11	55,339,603	55,433,572	OR4		CNP	NA
11013.p1	chr11	55,370,916	55,419,315	OR4		CNP	NA
11205.p1	chr11	55,370,916	55,419,315	OR4		CNP	NA
11291.p1	chr11	55,370,916	55,419,315	OR4		CNP	NA
11753.p1	chr11	55,370,916	55,419,315	OR4		CNP	NA
13409.p1	chr12	54,639,898	54,718,965	HNRNPA1	Y	CNP	NA

Sample	hg19 Coordinates (chr - start - stop)			Gene	Processed Pseudogene	Type of Event	Validation
12249.p1	chr12	104,376,576	104,387,282	TDG	Y	CNP	NA
13409.p1	chr13	27,679,867	27,847,631	RPL21	Y	CNP	NA
11638.p1	chr17	45,201,251	45,297,419	CDC27	Y	CNP	NA
13409.p1	chr17	45,201,251	45,297,419	CDC27	Y	CNP	NA
11653.p1	chr19	52,132,290	52,149,893	SIGLEC14		CNP	NA
12641.p1	chr19	52,133,551	52,149,313	SIGLEC14		CNP	NA
13409.p1	chr12	53,291,211	53,410,394	KRT8/KRT18		CNP	NA
11257.p1	chr14	20,248,481	20,529,142	OR4		CNP	NA
11947.p1	chr17	39,502,370	39,553,791	KRT33/KRT34		CNP	NA

Table S4: Genotyping Correlation with Whole-Genome Absolute Copy Number

Location	Genes	mrsFAST r ²	BWA r ²	Median Copy Number
chr1:104230039-104238912	<i>AMY1A</i>	0.98		8.11
chr1:110222301-110242933	<i>GSTM2,GSTM1</i>	0.99	0.86	3.12
chr1:144951760-145076079	<i>PDE4DIP</i>	0.99	0.61	6.66
chr1:145209110-145285912	<i>NOTCH2NL</i>	0.92	0.17	8.73
chr1:145293370-145368682	<i>NBPF10</i>	0.03	0.14	258.86
chr1:196788860-196801319	<i>CFHR1</i>	0.88	0.64	2.66
chr1:196825137-196896065	<i>CFHR4</i>	0.54	0.50	2.53
chr1:202415009-202496465	<i>PPP1R12B</i>	0.99	0.05	2.03
chr1:21766630-21811393	<i>NBPF3</i>	0.40	0.01	13.85
chr1:25598980-25656936	<i>RHD</i>	0.98	0.87	4.01
chr11:55403116-55451172	<i>OR4P4,OR4S2,OR4C6</i>	0.94	0.88	1.04
chr11:61008668-61018915	<i>PGA5</i>	0.98	0.21	5.98
chr12:11505418-11542473	<i>PRB1</i>	0.46	0.83	4.39
chr14:20202606-20420924	<i>OR4Q3,OR4M1,OR4N2,OR4K2,OR4K5</i>	0.96	0.93	3.75
chr14:74035771-74042359	<i>ACOT2</i>	0.98	0.08	3.00
chr15:22304656-22588026	<i>OR4N4</i>	0.97	0.61	4.23
chr15:30605924-30675622	<i>CHRFAM7A</i>	0.84	0.07	3.92
chr16:14766404-14788526	<i>PLA2G10</i>	0.22	0.11	8.57
chr16:15068832-15131552	<i>PDXDC1</i>	0.93	0.36	4.73
chr16:22524883-22547861	<i>LOC100132247</i>	0.08		48.81
chr16:32684848-32688053	<i>TP53TG3B,TP53TG3</i>	0.89		8.28
chr16:70148739-70196427	<i>PDPR</i>	0.96	0.91	4.73
chr17:18362101-18425291	<i>LGALS9C</i>	0.96	0.80	6.74
chr17:20353175-20370848	<i>LGALS9B</i>	0.93	0.04	6.57
chr17:34431219-34433014	<i>CCL4</i>	0.98	0.01	5.45
chr17:34522268-34524156	<i>CCL3L1</i>	0.95		6.88
chr17:34746118-34808091	<i>TBC1D3H,TBC1D3G,TBC1D3C</i>	0.75		47.70
chr17:36337711-36348666	<i>TBC1D3</i>	0.92		49.01
chr17:39506594-39525574	<i>KRT33A,KRT33B</i>	0.67	0.61	2.19
chr17:39531902-39536694	<i>KRT34</i>	0.60	0.58	2.46
chr17:39738532-39743147	<i>KRT14</i>	0.92	0.05	4.35
chr17:44165239-44800231	<i>KIAA1267,LRRRC37A,ARL17A,LRRRC37A2,NSF</i>	0.97	0.35	3.95
chr17:45608443-45700642	<i>NPEPPS</i>	0.79	0.07	8.42
chr17:62850487-62914903	<i>LRRRC37A3</i>	0.83	0.24	10.79
chr19:49535129-49536495	<i>CGB2</i>	0.43	0.01	19.19
chr19:54799854-54804238	<i>LILRA3</i>	0.85	0.76	7.05
chr2:97779232-97915915	<i>ANKRD36</i>	0.06	0.11	18.23
chr22:16256331-16287937	<i>POTEH</i>	0.58		18.93
chr22:23043312-23249272	<i>IGLL5</i>	0.90	0.10	2.13
chr22:24376138-24384284	<i>GSTT1</i>	0.98	0.44	1.09
chr22:25677318-25911586	<i>LRP5L</i>	0.99	0.93	2.96
chr3:197879236-197907728	<i>FAM157A</i>	0.03	0.36	8.71
chr3:75786028-75834255	<i>ZNF717</i>	0.10	0.03	39.36
chr4:69366860-69554789	<i>UGT2B17,UGT2B15</i>	0.91	0.00	2.72
chr4:70127619-70235027	<i>UGT2B28</i>	0.91	0.94	5.50
chr5:180377202-180416706	<i>BTNL3</i>	0.99	0.97	2.36
chr5:68821588-68854548	<i>OCLN</i>	0.92		3.19

Location	Genes	mrsFAST r^2	BWA r^2	Median Copy Number
chr5:69316084-69343660	<i>SERF1A</i>	0.77		4.01
chr5:69345316-69374572	<i>SMN1</i>	0.95		3.84
chr5:795743-825341	<i>ZDHC11</i>	0.92	0.78	3.98
chr6:257332-380527	<i>DUSP22</i>	0.99	0.99	4.01
chr6:32455238-32493130	<i>HLA-DRB5</i>	0.92	0.09	1.41
chr7:101986192-101996889	<i>SPDYE6</i>	0.02		38.21
chr7:102114556-102332921	<i>POLR2J,SPDYE2,POLR2J3,RASA4,UPK3BL,POLR2J2</i>	0.95	0.12	10.58
chr7:143223558-143541003	<i>CTAGE15P,FAM115C</i>	0.98		3.66
chr7:144052488-144077725	<i>ARHGEF5</i>	0.95	0.45	5.29
chr7:43980493-44058748	<i>UBE2D4,SPDYE1</i>	0.05	0.06	9.20
chr8:11946846-11973025	<i>ZNF705D</i>	0.18		9.32
chr9:141106636-141134172	<i>FAM157B</i>	0.00	0.61	9.16
chr9:14510-29739	<i>WASH1</i>	0.26		21.12
chr9:33795558-33799229	<i>PRSS3</i>	0.96	0.33	5.75
chr9:67926760-67969840	<i>ANKRD20A1</i>	0.66		28.15

Table S5: Accuracy of absolute copy number prediction

location	genes	# correct HapMap genotypes (of 7; from Campbell et al 2011)
chr1:25592663-25663607	<i>RHD</i>	7
chr1:110222301-110242933	<i>GSTM1,GSTM2</i>	7
chr1:144959523-145081011	<i>PDE4DIP</i>	5
chr1:196738897-196801697	<i>CFHR3,CFHR1</i>	4
chr1:196825137-196896065	<i>CFHR4</i>	5
chr1:202389905-202402001	<i>PPP1R12B</i>	7
chr1:202415009-202496465	<i>PPP1R12B</i>	2
chr2:89160037-89262733	<i>Ig Light chain locus</i>	4
chr3:100547342-100670846	<i>ABI3BP</i>	6
chr3:151511518-151550270	<i>AADAC</i>	7
chr3:189364074-189538586	<i>TP63</i>	5
chr4:68793517-68833125	<i>TMPRSS11A</i>	6
chr4:69386965-69483317	<i>UGT2B17,UGT2B15</i>	7
chr4:70127619-70235027	<i>UGT2B28</i>	5
chr4:144921494-145040886	<i>GYP A,GYP B</i>	7
chr5:795743-825341	<i>ZDHHC11</i>	5
chr5:32107113-32169449	<i>PDZD2,GOLPH3</i>	7
chr5:68821588-68854548	<i>OCLN</i>	5
chr5:69316084-69343660	<i>SERF1A</i>	4
chr5:69345316-69374572	<i>SMN1</i>	6
chr5:180377202-180416706	<i>BTNL3</i>	6
chr6:257332-380527	<i>DUSP22</i>	5
chr6:32455238-32493130	<i>HLA-DRB5</i>	6
chr7:143223558-143541003	<i>FAM115C,CTAGE15P</i>	4
chr9:115383227-115585827	<i>KIAA1958,C9orf80,SNX30</i>	7
chr10:51008386-51114434	<i>PARG</i>	7
chr10:135232058-135377386	<i>MTG1,CYP2E1,SYCE1</i>	7
chr11:55403116-55451172	<i>OR4P4,OR4S2,OR4C6</i>	3
chr12:11505418-11542473	<i>PRB1</i>	5
chr14:20202606-20420924	<i>OR4Q3,OR4M1,OR4N2,OR4K2,OR4K</i>	6
chr14:88400031-88414591	<i>GALC</i>	6
chr15:22304656-22588026	<i>OR4N4</i>	4
chr15:30605924-30675622	<i>CHRFAM7A</i>	7
chr16:70148739-70196427	<i>PDP R</i>	5
chr17:18362101-18425291	<i>LGALS9C</i>	5
chr17:34416411-34496071	<i>CCL3,CCL4,TBC1D3B</i>	2
chr17:39506594-39525574	<i>KRT33A,KRT33B</i>	6
chr17:39531902-39536694	<i>KRT34</i>	7
chr19:54724572-54740148	<i>LILRB3</i>	4
chr22:22754320-23038160	<i>PRAME,GGTLC2</i>	6
chr22:23043312-23249272	<i>IGLL5</i>	6
chr22:24347958-24395540	<i>GSTT1,LOC391322</i>	5

Table S6: Signal-to-Noise ratios for mrsFAST and BWA calls

Sample	Chro m	Start	Stop	mrsFAST Signal	mrsFAST StdDev	mrsFAST SNR	BWA Signal	BWA StdDev	BWA SNR	mrsFAST SNR improvement
NA18517	16	21396577	21756357	1.927	0.183	10.516	0.373	0.059	6.314	67%
NA19240	12	133659688	133727740	0.881	0.138	6.383	0.212	0.052	4.046	58%
NA15510	7	99507187	99627998	1.670	0.160	10.463	0.284	0.048	5.916	77%
NA19129	6	29910533	30043566	0.894	0.173	5.183	0.377	0.056	6.748	-23%
NA18517	4	68788472	69057034	1.646	0.152	10.841	0.470	0.055	8.520	27%
NA15510	3	19492646	21465556	1.565	0.169	9.240	0.246	0.042	5.833	58%
NA15510	1	155227075	155264543	1.487	0.157	9.496	0.459	0.051	9.025	5%

Appendix C (Chapter 3)

Exome-based CNV calling

Exome data preparation and alignment

Previously generated FASTQ data from four exome sequencing studies (Iossifov et al., 2012; O’Roak et al., 2012; Sanders et al., 2012) was used in this study. In addition, we generated sequence for unaffected sibling in 20 published trios (O’Roak et al., 2011) for a complete set of 412 quads, comprising (One family was excluded during a QC step, see “Coverage and quality control”, below). Reads were split into consecutive 36mers, up to two per read, and mapped using the single-end mode of mrsFAST (Hach et al., 2010), allowing for up to two mismatches per 36mer. We aligned reads to a concatenated hg19 reference genome, which included exome-capture targets based on the Nimblegen EZ Exome v2.0 platform (194,080 targets) as well as 300bp up- and downstream of each target.

CoNIFER-based CNV discovery from exome read-depth

Using CoNIFER v0.2.2 (<http://conifer.sf.net>, (Krumm et al., 2012)), we processed each of the three datasets separately. RPKM values were calculated for 194,080 probes and exons targeted by the Nimblegen EZ Exome v2.0 exome sequence enrichment platform. We set the --svd option to 12 (for samples in the Iossifov data set) or 15 (for O’Roak and Sanders data sets), and used default CoNIFER settings for all other options. After CoNIFER analysis, the raw SVD-ZRPKM values were exported (using the export command) for downstream analysis.

Coverage, Quality Control and data uniformity

We examined the overall mapped coverage after mrsFAST alignment in all samples, as this can affect the sensitivity of exome-based CNV detection (Figure S2). There was no significant difference in coverage between proband-sibling pairs (paired two-tailed t-test = 1.68, $p = 0.09$). However, we noted that samples

from the lossifov dataset had overall lower coverage than did samples from the other two data sets, likely due to the more aggressive multiplexing strategy used by lossifov and colleagues in sequencing their families.

We used the `--write_sd` option in the CoNIFER program to calculate the per-sample standard deviation of SVD-ZRPKM values after processing (listed in Table S1). Abnormally high values would indicate an excessively noisy sample, or a sample for which exon read-depth values could not be adequately normalized with CoNIFER. Fortunately, no samples were excluded using this metric (all std. dev ≤ 0.61). We did note what appeared to be significant contamination between reads in family 12154, and removed this family from all further analysis, resulting in 411 families.

CNV segmentation and single sample calling

All four members of each quad were analyzed for CNVs, in order to minimize false negatives in probands and siblings. We used DNACopy and CGHCall to segment and assign deletion or duplication probabilities to SVD-ZRPKM values. In order to prevent excessively strong SVD-ZRPKM signals from interfering with the models used by CGHCall to assign copy number, we clipped the signal at ± 3 for each exon. Parameters for DNACopy were as follows: $\alpha = 0.01$, using the `undo.split="sdundo"` option with `undo.SD = 2`. Default options for CGHcall were used, and we allowed only "deletion" and "duplication" as called states. Using these parameters, we obtained 32,672 raw segments as either "deleted" or "duplicated".

After segmentation, we excluded 48 calls which clearly failed to segment properly. These calls were larger than 250 exons and had a absolute median signal less than 0.2. We visually examined these calls and confirmed that they were clear false positives generated by CGHcall. These calls were removed to reduce interference with downstream CNVR clustering and analysis, leaving

32,624 calls.

Unified probe set for three datasets

As we analyzed each of the three datasets (describe above) separately, and the CoNIFER algorithm dynamically excludes poorly performing probes, the final datasets each had a slightly differing total number of probes. We unified these probes into a unified probe set by merging non-overlapping probes (i.e., those that did not overlap in all three datasets) to their closest neighbor by genomic distance. The probes from each dataset had greater than 99% overlap with the other two datasets, as well as with the unified probeset (O’Roak: 99.7%, Sanders: 99.7%, lossifov: 99.1%). However, after unification, 92 calls were eliminated from one or more datasets, on the basis that the unified probe set did not support our minimum size threshold of two exons for these calls.

Clustering CNVs and frequency estimates

Next, we grouped individual CNV calls into similar CNV Regions (CNVRs) using pairwise distances between all CNVs based on a modified reciprocal overlap (RO) heuristic. This function calculates the RO between two CNVs based on the minimum fraction of number of overlapping probes, and weights this percentage based on the total number of non-overlapping probes on each end. In this way, the function takes into account the uncertainty in breakpoints and RO for two small CNVs, while allowing two large overlapping CNVs to be count as distinct entities. The function is given by:

$$RO = \min\left(\frac{C}{l_a}, \frac{C}{l_b}\right) \cdot \frac{(g^{LO} + g^{RO})}{2}$$

Where RO is the modified reciprocal overlap, C is the number of overlapping probes between two CNVs A and B, l_a and l_b are the lengths (in number of probes) for A and B, and LO and RO correspond to the number “left overhanging” and “right overhanging” probes on either side of the common probes. We empirically

picked $g = 0.9$.

CNVs were clustered into CNVRs by a weighted hierarchical clustering (WPGMA, weighted pair group method with averaging) method based on the pairwise weighted RO calculated above for each pair of CNVs. Each resulting tree connects all CNVs based on their similarity under the modified RO function (above), and the distances between leaves or clusters measures how likely the CNVs are part of the same CNVR.

To generate CNVRs and frequency counts, we flattened the trees based on the cophenetic distance (i.e., the “height” of each branch of the dendrogram) at an empirical value of 0.85. Using this clustering method, we initially generated 3,473 CNVRs. Of these, 332 had more than 15 calls, and we removed these CNVRs from subsequent analysis of rare CNVs on the basis that they correspond to CNPs, which we analyzed separately. The 3,141 CNVRs (containing 7,424 calls) were carried forward to the next step.

Quad-based genotyping method

We reduced the rate of false-positive de novo calls in probands (and conversely, reduced the false-negative rate inherited CNVs) by applying a novel algorithm designed specifically to find inherited CNVs within a family, under the assumption that an false positive for an inherited call would be much more likely than a true positive de novo event. Using the Continuous Mutual Information (CMI) metric, which assess how “informative” a set of observations X is for set Y , we made pair-wise comparisons of SVD-ZRPKM values at each CNV locus within each family. Specifically, for each rare call (termed the “index call” here), the CMI between all pairs of samples in the quad was calculated:

for each quad member a:

 for each other quad member b:

$$CMI(X;Y) = \int_{\substack{x < 0: 0 \\ x \geq 0: 100}}^{\substack{x < 0: -100 \\ x \geq 0: 0}} \int_{-100}^{100} f(x,y) \cdot \log\left(\frac{f(x,y)}{f(x) \cdot f(y)}\right) dx dy$$

Where X and Y are vectors containing SVD-ZRPKM values for the first sample and the second sample, including the values of the 30-100 probes surrounding the call on either side (Specifically, 30 probes for events ≤ 3 exons; 50 flanking probes for events 4-150 exons; 100 probes flanking for events ≥ 150 exons). To incorporate our high prior probability on the location of the index call, we weighted the values X and Y by a vector W, which was 10 for the probes in the index call and 1 elsewhere. Then, $f(x,y)$ provides the joint probability density of the two vectors, calculated by estimating the gaussian kernel density for both vectors; similarly, $f(x)$ and $f(y)$ are the marginal densities as estimated by the gaussian kernel density for X and Y separately. In order to calculate the CMI value, we used the `scipy.quadpack` library and integrated the CMI function across the interval $(-100,100)$ for all values of x; the limits of y were determined by a step function (if $x < 0$: y: $(-100, 0]$; if $x \geq 0$: y: $(0, 100)$), which prevented reciprocal calls or reciprocal signals from contributing positively to the CMI. We calculate the pairwise CMI for all pairs (excluding self-self) of members in each quad, and averaged the values for each pair of pairs (e.g., the CMI of mother-proband was averaged with proband-mother). In some cases, the numerical integration was not able to converge on a value for a particular pair X;Y, and in this case, we used the successfully calculated CMI. In no cases did the integration for both $I(X;Y)$ and $I(Y;X)$ fail. As our focus was to genotype and impute rare CNVs only, and the computational run time for this step grows at a rate of $O(N^2)$, we limited the calculation of the CMI values to CNVR clusters with 15 or fewer calls.

Depending on the individual carrying the index call, we used two different CMI cutoff values to determine if the pairwise CMI value indicated inheritance: if the parent carried the index call, we required a CMI threshold of 1.3; conversely, if the offspring carried the index call, we required only 1.0 (this discrepancy takes

into account the high likelihood that a call is inherited if seen in an offspring, and the lower likelihood that a parental call is passed on). Additionally, we required that the CMI score was outside of the 0.005 percentile of a fitted gamma distribution of CMI values between 25 randomly unrelated samples at that locus. If these conditions were met, we imputed the CNV in the members of the quad for which no call had been made, based on the boundaries of the index call (i.e., the prior).

We calculated pairwise CMI values for the calls in 3,141 CNVR clusters with 15 or fewer calls. This step inferred a total of 903 calls passing the CMI cutoff(s), of which 204 calls (23%) were part of rare CNVR clusters (<15 calls) and 699 which joined the existing CNP CNVR clusters.

After adding the newly found calls, we re-clustered all calls into CNVRs and recomputed the final frequencies for each, as describe above (“Clustering CNVs and frequency estimates”). This resulted in 3,102 CNVRs containing 7,628 calls carried into our final filtering and classification step.

Filtering segmentally duplicated regions and processed pseudogenes:

Our primary interest in rare and private CNVs prompted us to exclude CNVs which primarily lie in segmental duplications or otherwise multi-copy regions of the genome. These regions are polymorphic and the independent assortment of parental alleles prevents accurate assessment of inheritance patterns in these regions. We excluded CNVs which were found to have more than 50% of their probes within segmental duplications or duplicated regions of the genome (defined using previous methods from 1000 Genomes whole-genome depth-of-coverage analysis, where >80% of 34 unrelated genomes had a copy number of three or greater in 500bp repeat-masked windows across the genome).

Excluding calls which overlapped at least 50% with these regions resulted in the exclusion of 2,353 calls (45% of all calls), corresponding to 609 CNVRs.

Next, we excluded calls which were likely to due solely to the insertion of processed pseudogenes. CoNIFER and most exome-based read-depth methods are sensitive to copy number changes specifically of exons, which can be the result of retro-insertion of processed mRNA transcripts (see Krumm et al., 2012 for more details). Processed pseudogenes can be polymorphic or fixed among individuals; furthermore, inserted copies may be present in variable copy number or absent altogether within the reference genome. To eliminate CNV calls in our dataset due to polymorphic or novel insertions of processed pseudogenes, we referenced three sources to define likely processed pseudogenes: (1) a list of commonly polymorphic processed pseudogenes generated using SPLIT-READ (Karakoc et al., 2011) from 20 control exomes and (2) 225 autism trios (data not reported here). We excluded calls from our call list for which $\geq 90\%$ of the probes corresponded to a gene which had been observed at least once in 225 trios. This excluded a total of 1,098/7,628 calls (17%, in line with previous estimates from (Krumm et al., 2012)).

Final filtering and call set generation

Our final set of calls was created by requiring an absolute median SVD-ZRPKM score (i.e., signal strength) of ≥ 0.5 for calls with 5 or more probes, ≥ 1.0 for calls 3-5 probes in length, and ≥ 1.0 for calls 2 probes in length. We excluded any calls on the X or Y chromosomes for all analyses in this work.

Validation of CNV calls

Targeted array CGH microarray design

To validate the small rare CNVs discovered using the discovery pipeline, we designed a custom Agilent SurePrint G3 4x180k CGH microarray to confirm CNVs. As the variable spacing of the exome probes prevents precise knowledge of CNV breakpoints, we used a variable density array design with ± 3 exon

overhang based on the exome-based breakpoints. (with min/max limits of 5kbp and 50kbp) where possible. Probe density within the CNV call ranged from 125bp^{-1} for calls smaller than 10kbp to 5kbp^{-1} for large calls up to 500kbp, in order to insure at least 10 probes per call. Due to the high density of probes required for validation of small CNVs, some of the probes were of lower quality (as based on the manufacturer's quality score), and their performance was accordingly lower.

Array CGH methods

Test and reference DNA (we used DNA from HapMap sample NA18507) from each sample was labeled with Cy3 and Cy5 dye using a NimbleGen array labeling kit according to manufacturer's instructions. Five micrograms of labelled test and reference DNA was hybridized for 24 hours using Agilent reagents to the microarray slide and washed according to manufacturer's directions. Slides were scanned using an Agilent Microarray Scanner and analyzed using Agilent Feature Extract v10.5.1.1.

Data processing and array quality control

Array intensity ratios were log-transformed and assessed for quality control. Arrays with a per-sample standard deviation of LogR values > 0.5 were repeated. In order to reduce systematic and batch noise between probes and samples, we employed a similar normalization strategy to the CoNIFER pipeline and used SVD to remove the three strongest components of variance.

Receiver Operating Curve determination of array CGH thresholds

We determined minimum logR thresholds for the validation arrays by leveraging the logR values across the 60 previously identified CNVs (from Sanders et al., 2011), each found in at least one of our validation samples. We calculated Receiver Operating Curves (ROC, Figure S3) for duplications (39 calls) and deletions (21 calls), using the samples *without* the previously identified CNVs as

the “true negatives”. Next, we individually picked the optimal operating point (OOP) for deletions (median LogR OOP ≤ -0.178) and duplications (median LogR OOP ≥ 0.24), such that we maximally discerned our known true positives from true negatives. Both OOPs had a FPR of $\sim 1\%$, and a recall rate $>90\%$, indicating our array was highly specific and sensitive to true events. These logR cutoff values were used in assessing if novel CNVs were true positives or not: if the mean LogR across all probes in the call interval was greater than the duplication threshold (or lower than the deletion threshold), we considered the call validated.

Estimating false positive rate

We started with 161 exome-based calls (Table S5) among 80 randomly selected probands and siblings. Of these 161 calls, 69 could be confirmed by a CNV already reported by Sanders and colleagues. We used array CGH data from our customized microarray and the OOPs determined above and the mean of all the array CGH probes across the exome-based CNV start and stop in order to validate remaining calls. The OOP thresholds were exceeded in 61 calls, and based on inheritance across multiple and a combination of available raw Illumina 1M data, we scored four additional calls as validated (Table S5). Six calls did not have sufficient array probe coverage (our upper estimate of the reported FPR includes these calls as false positives, the lower estimate excludes these six from all calculations).

We found that 14 of the unconfirmed calls were rare processed pseudogenes specific to the family or samples tested. To find these, we mapped exome sequencing reads from each sample to a customized reference sequence composed of mRNA sequences extracted from RefSeq. If more reads mapped across the exon-exon junctions within the CNV call in the sample tested than across the same junction in other samples, we considered the elevation of exome read-depth signal to be due to a processed pseudogene insertion, rather than a

true genomic CNV.

Additional analyses

Bootstrap permutations of burden analysis

We tested the robustness of the overall effect of burden by a bootstrap method, in which we calculated the CNV burden ratio (for CNVs and genes) of 10,000 randomly sampled (with replacement) sets of families from the overall set of 411. The resulting distributions of total CNV counts and total gene count and the distribution of burden for both are shown in Figure S5 (compare to actual results for all 411 families in Figure 2a and 2b). The empirical 95% confidence intervals for both the burden of CNVs (CI: 1.09-1.29) and genic burden (1.10-1.52) reject the null hypothesis (at $\alpha = 0.05$) of no differential CNV burden between probands and siblings.

Phenotypes and regression models

Full-scale IQ, SRS t-scores, and individual components of the SRS were downloaded from the SSC database and release 14 of the SSC. For all analyses we excluded the entire family if any values were missing for either the proband or sibling.

To clarify how we classified SRS discordant and concordant families, we provide the following table:

		Proband SRS		
		< 60 ("Unaffected")	60-75 ("Moderate")	76+ ("Severe")
Sibling SRS	< 60	16 Concordant Quads	84 Concordant Quads	276 Discordant Quads
	60-75	—	1 Concordant Quad	13 Concordant Quads
	76+	—	—	1 Concordant Quads

Gene expression data and enrichment analysis

We used publicly available gene expression data from the Human U133A/GNF1H Gene Atlas (GEO: GSE1133), comprising 79 human tissues, including 18 nervous system tissues (Su, 2004). We associated the microarray probe IDs with HUGO gene names and average expression across multiple probes in the same gene. For each tissue, genes were sorted by expression and we considered the top 5% of each category to be “highly expressed”. To calculate enrichment, we took the unique sets of genes disrupted in probands and those disrupted in siblings and intersected each with the set of highly expressed genes in each category. The ratio of counts between these two intersections constituted the fold enrichment for each category.

In order to correct for the 79 multiple comparisons, we employed a permutation and false discovery rate (FDR) strategy. First, we derived a null distribution of enrichment between probands and siblings by shuffling the proband-only and sibling-only sets of genes and recomputing the enrichment. Next, an empirical p-value was derived by scoring the actual enrichment value against the null distributions for each tissue. Using the FDR method described in (Storey & Tibshirani, 2003) and the R package qvalue, we calculated q values for each tissue and assessed statistical significance at $q < 0.05$. In order to calculate the brain and non-brain averages, we averaged gene expression across all 18 brain- and nervous system tissues and 61 non-brain tissues. These two categories were corrected for two comparisons each.

Previously associated genes

To establish the list of genes previously associated with autism/ASD/intellectual disability/schizophrenia, we attempted to identify all genes that were “causal” or associated with developmental delay, intellectual disabilities and schizophrenia. We conducted searches of OMIM with the following terms: “mental retardation” “intellectual disabilities”, “autism” schizophrenia. We also included genes from the

Simons SFARI autism candidate genes with” association scores” ranging from 1 to 4 (n=155 genes)

(https://gene.sfari.org/autdb/submitsearch?selfld_0=GENES_GENE_SYMBOL&selfldv_0=&numOfFields=1&userAction=viewall&tableName=AUT_HG&submit2=View+All#GS).

Additional control exomes

The set of 2,972 exomes used to assess population frequency of ultra rare CNVs were taken from the National Heart Lung and Blood Institute’s Exome Sequencing Project (ESP). These exomes were processed in bulk using CoNIFER (with 21 components removed) and locus-specific population frequencies were determined by manual inspection of outlier samples for each locus.

Combined mutation model

We used published lists of *de novo* SNV and indel mutations from published lists in the three studies ((lossifov et al., 2012; O’Roak et al., 2012; Sanders et al., 2012)). In our combined model (see discussion), we only counted disruptive SNVs and indels (i.e., nonsense, splice, and frameshifting), as these have been shown to be most enriched in probands. Inherited and *de novo* CNV counts were derived from this work (Tables S7). We used a logistic regression model, which transforms a binary outcome (i.e., affected vs. unaffected) such that linear predictors can be used. The model as shown in Figure 5 is:

$$\text{logit}[P(\text{affected} = 1)] \sim \text{intercept} + \text{Burden}_{\text{de novo CNVs}} * \text{Burden}_{\text{inherited CNVs}} * \text{Burden}_{\text{de novo SNVs}}$$

Code availability

CoNIFER and CNV calling

CoNIFER can be downloaded from <http://conifer.sf.net>. Version 0.2.2 was used in this work. The custom pipeline for CNV calling as described in this work is

available at <http://conifer.sf.net>, although the authors cannot guarantee or provide any technical support for this.

Supplement References

- Hach, F., Hormozdiari, F., Alkan, C., Hormozdiari, F., Birol, I., Eichler, E. E., & Sahinalp, S. C. (2010). mrsFAST: a cache-oblivious algorithm for short-read mapping. *Nature Methods*, 7(8), 576–577. doi:10.1038/nmeth0810-576
- Iossifov, I., Ronemus, M., Levy, D., Wang, Z., Hakker, I., Rosenbaum, J., et al. (2012). De novo gene disruptions in children on the autistic spectrum. *Neuron*, 74(2), 285–299. doi:10.1016/j.neuron.2012.04.009
- Krumm, N., Sudmant, P. H., Ko, A., O'Roak, B. J., Malig, M., Coe, B. P., et al. (2012). Copy number variation detection and genotyping from exome sequence data. *Genome Research*. doi:10.1101/gr.138115.112
- O'Roak, B. J., Deriziotis, P., Lee, C., Vives, L., Schwartz, J. J., Girirajan, S., et al. (2011). Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nature Genetics*, 43(6), 585–589. doi:10.1038/ng.835
- O'Roak, B. J., Vives, L., Girirajan, S., Karakoc, E., Krumm, N., Coe, B. P., et al. (2012). Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature*. doi:10.1038/nature10989
- Sanders, S. J., Murtha, M. T., Gupta, A. R., Murdoch, J. D., Raubeson, M. J., Willsey, A. J., et al. (2012). De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature*, 485(7397), 237–241. doi:10.1038/nature10945
- Storey, J. D., & Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences*, 100(16), 9440–9445. doi:10.1073/pnas.1530509100
- Su, A. I. (2004). A gene atlas of the mouse and human protein-encoding transcriptomes. *Proceedings of the National Academy of Sciences*, 101(16), 6062–6067. doi:10.1073/pnas.0400782101

Figure S1: CNV Calling Flowchart

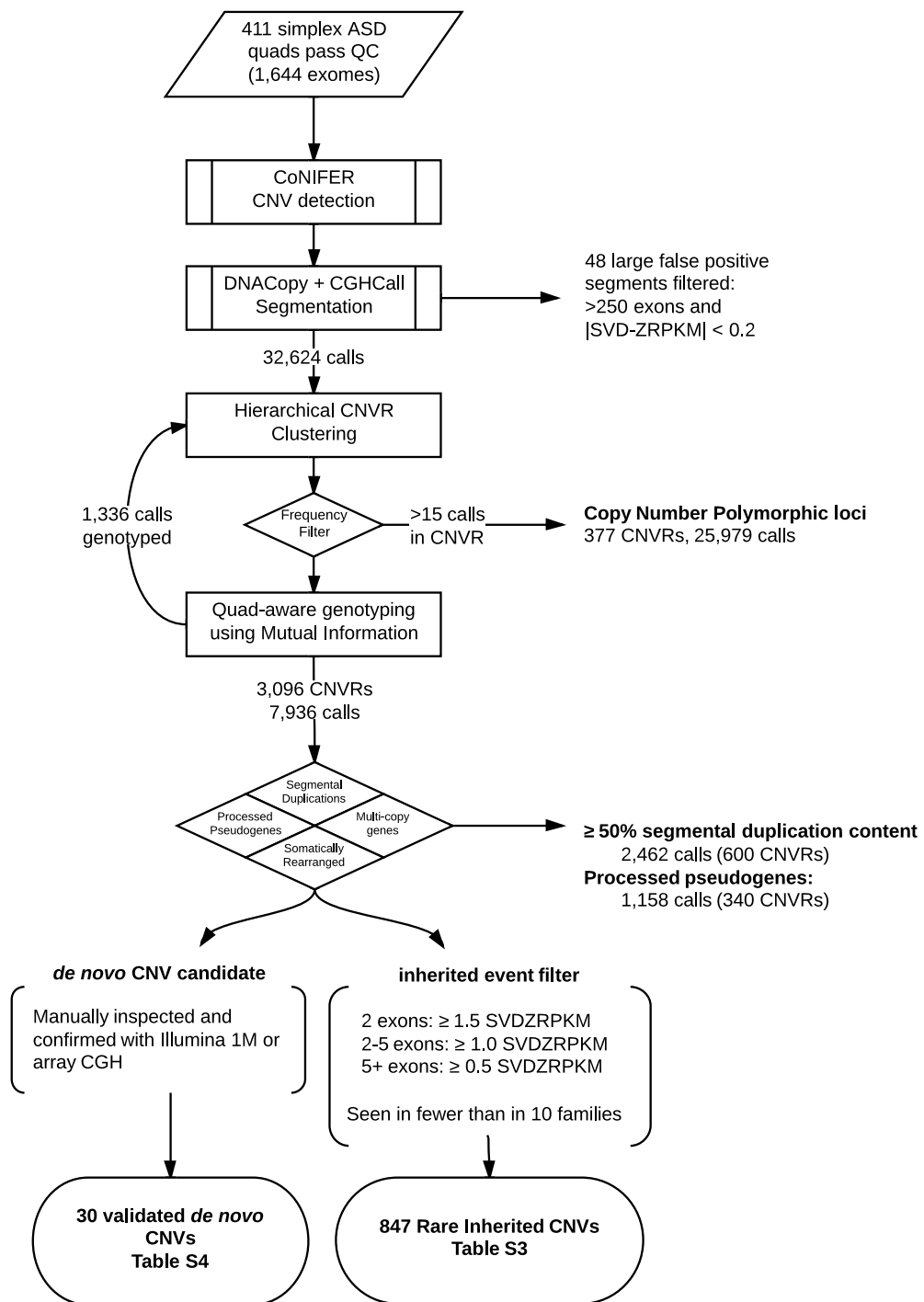


Figure S1: Flow chart for inherited CNV detection. See Methods and Supplemental Methods for details.

Figure S2: Mapped coverage between probands/siblings and by data source

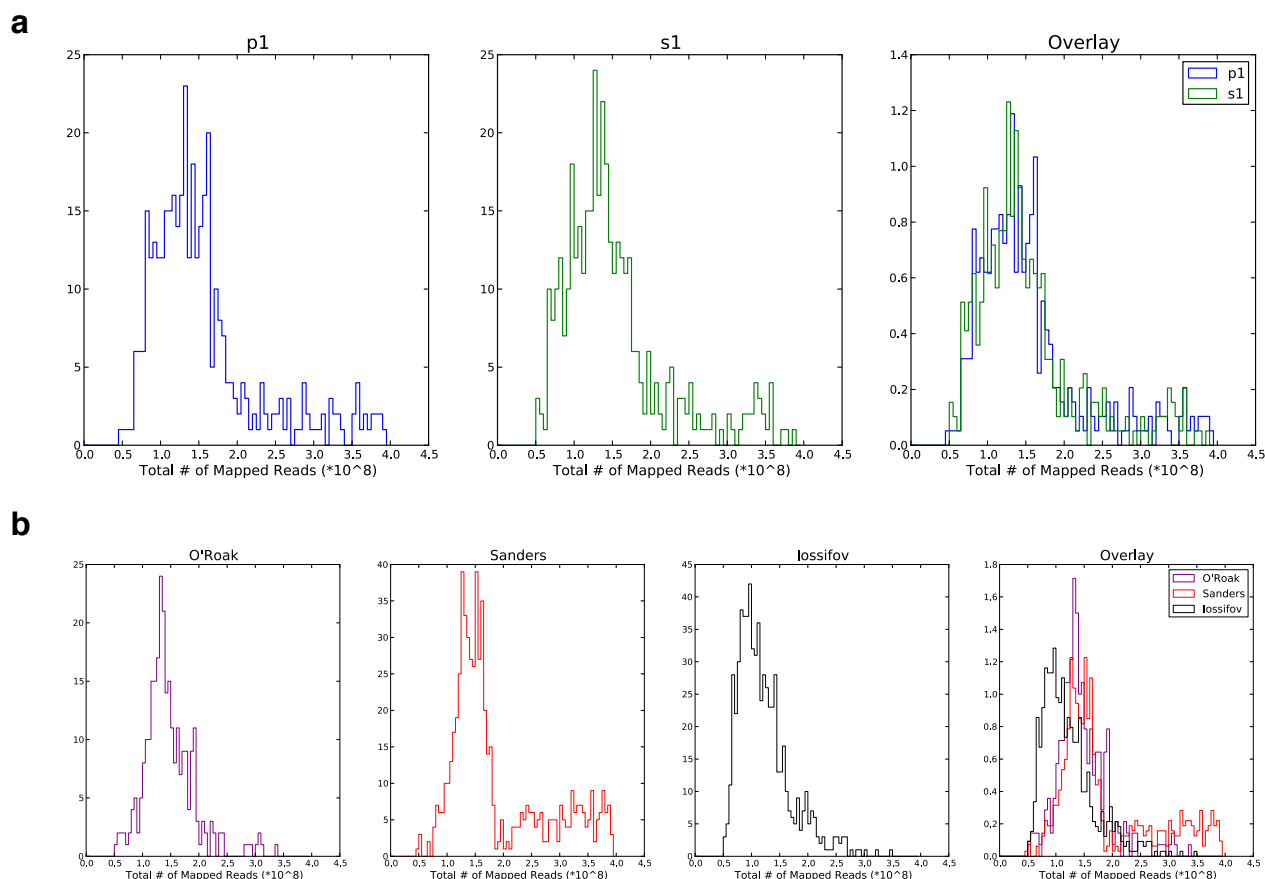


Figure S2: Mapped Coverage between probands/siblings and by data source. X-axis: total mapped 36mer reads ($\times 10^8$) by the mrsFAST alignment program to the human exome. **(a)** Histograms of Probands (left) and Siblings (center) and overlap (right) shows no significant difference in coverage levels (Paired t-test $p = 0.09$). **(b)**. Same as in **(a)**, but by dataset, revealing that the lossifov dataset had lower coverage than the O'Roak or Sanders datasets.

Figure S3: Array-CGH validation ROC curves

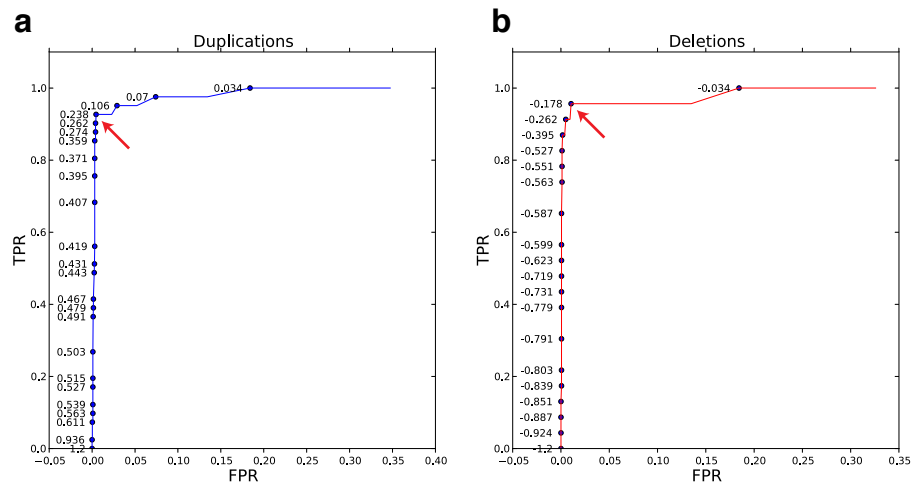


Figure S3: Receiver-Operator Curve determining deletion and duplication thresholds in array-CGH validation. ROC curves based on 60 true-positive deletions **(a)** and duplications **(b)** from Sanders et al., 2011 in these samples. Arrows indicate chosen optimal operating point (OOP), which was used as the threshold for validation of unknown calls.

Figure S4: CNV Size, inheritance, and copy number

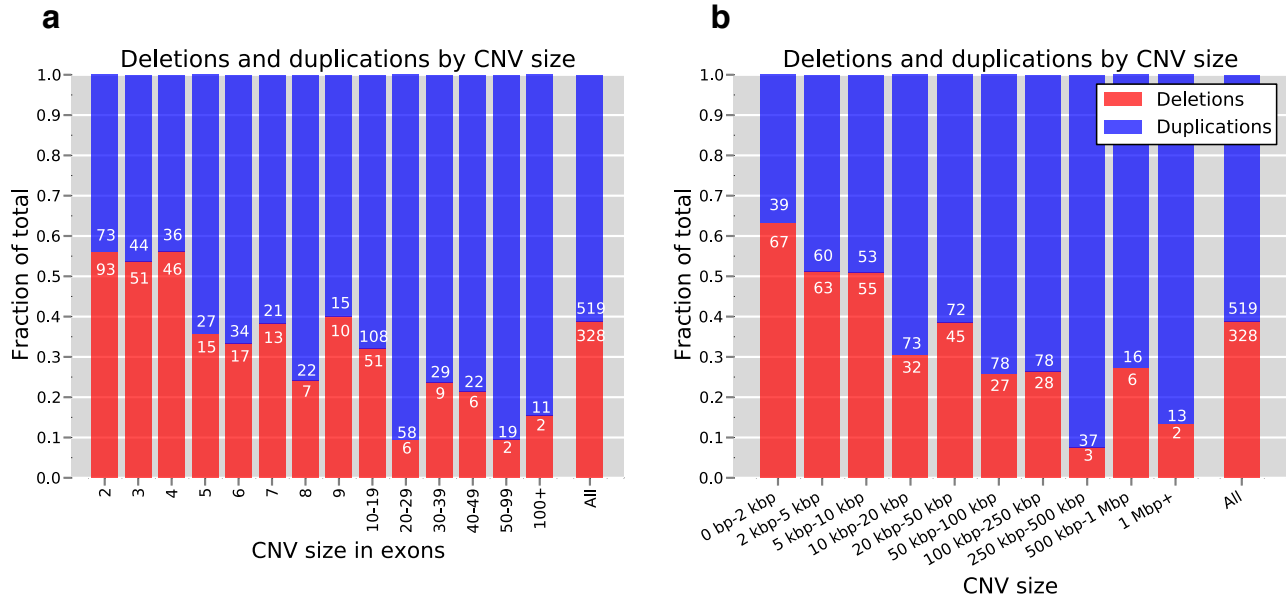


Figure S4: CNV size and copy number. Inherited CNVs in probands and siblings, binned by size in exons **(a)** or estimated genomic size **(b)**. As expected, larger CNVs are more likely to be duplications, an effect we found true for both probands and siblings.

Figure S5: Bootstrap results

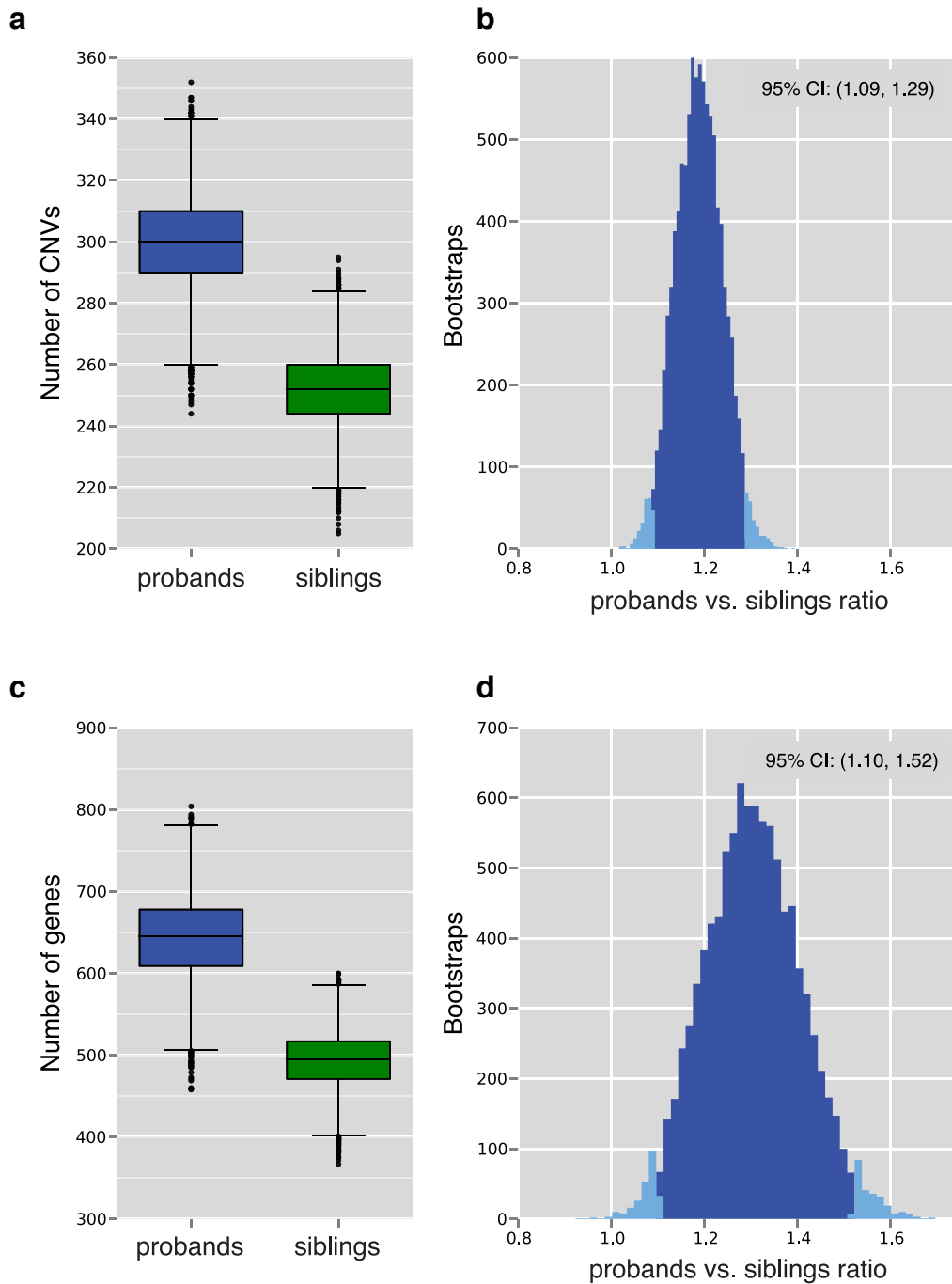


Figure S5: Results of bootstrap permutation test. We bootstrapped our set of inherited CNVs (sampling CNVs by family, with replacement), and calculated the total CNV counts **(a)** for probands (green) and siblings (blue) and CNV burden **(b)** between probands and siblings (dark blue: inner 95% of empirical distribution). In **(c)** and **(d)**, the results when counting total number of genes and genic burden.

Figure S6: Rare vs. Private burden in 411 quads

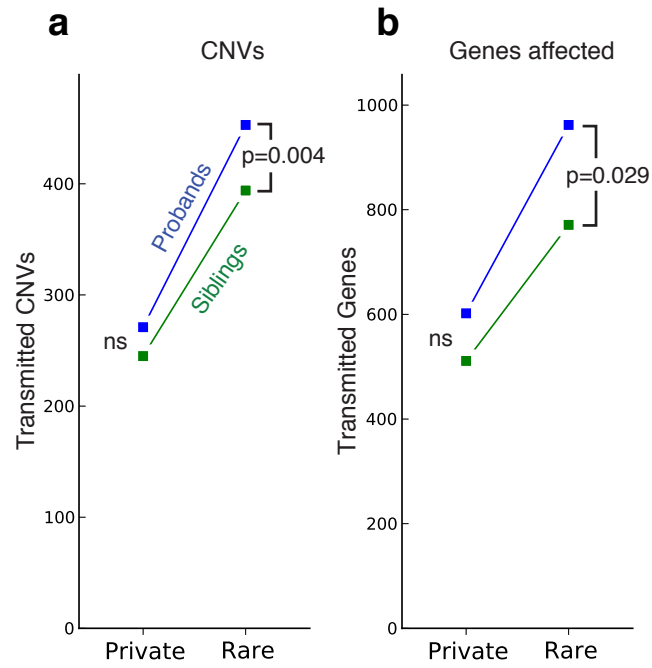


Figure S6: Rare vs. Private burden in 411 quads. There was no increased burden for CNVs **(a)** observed only once in 411 families, or for genes in those CNVs **(b)**.

Figure S7: Phenotypes in 411 probands and siblings

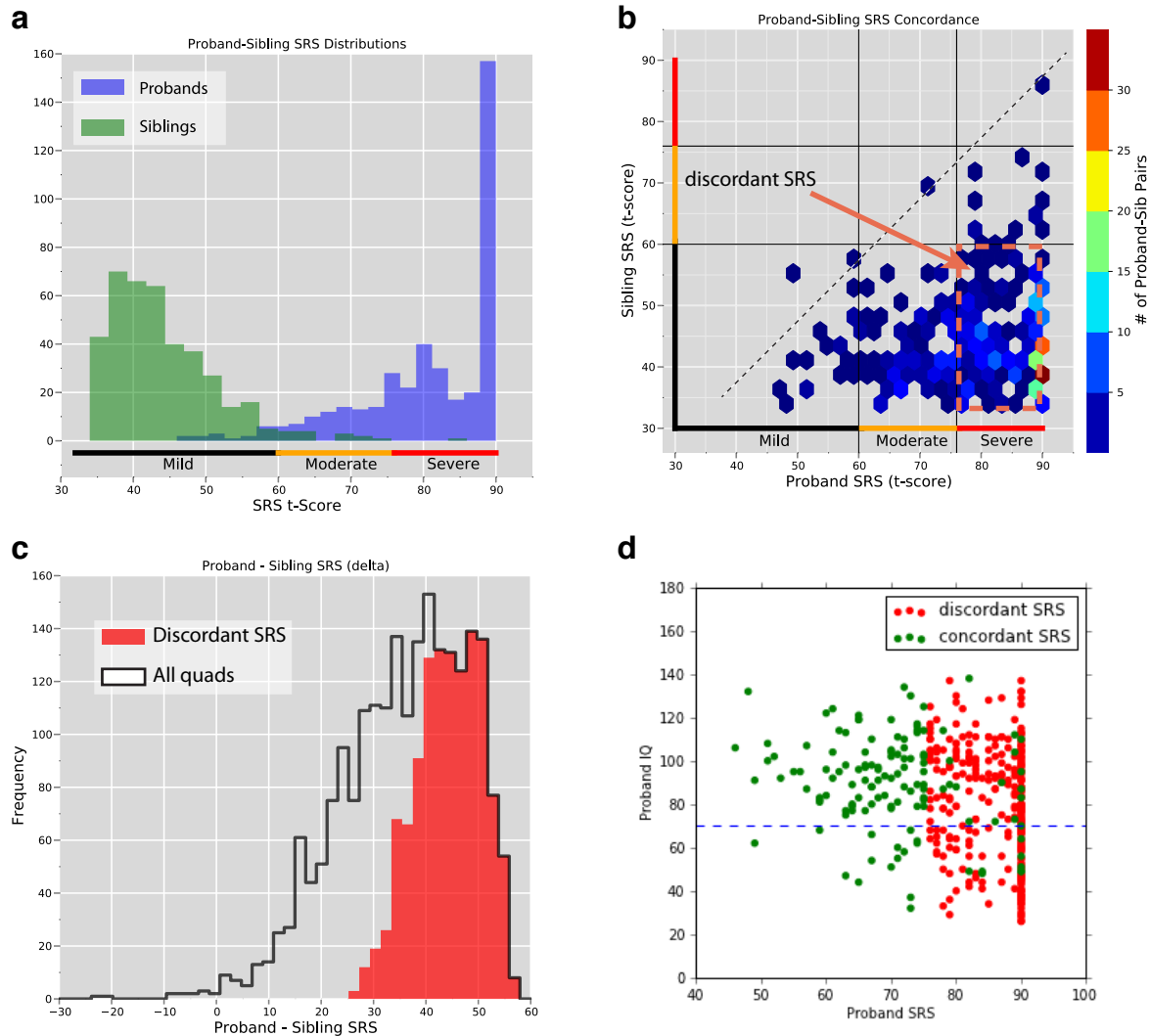


Figure S7: Phenotypes (SRS and IQ) in probands and siblings. (a) Distribution of SRS t-scores in probands (blue) and siblings (green). Higher scores are more affected, and SRS t-scores greater than 75 are considered “severely affected”. **(b)** Heatmap plot of SRS values for probands (x-axis) and their siblings (y-axis). In almost all cases, the probands have higher SRS scores, but the difference in SRS score between probands and siblings varies widely among all pairs. We designated the pairs with the most extreme differences of SRS score between them as “Discordant SRS” pairs (indicated by arrow and dashed orange box, lower right). **(c)** All of the SRS discordant pairs had SRS differences > 25 (by definition, as we required these pairs to have a proband SRS > 75 and a sibling SRS ≤ 50) **(d)** Scatter plot showing both proband SRS score and proband IQ score. Dashed blue line indicates cutoff for High and Low IQ in our comparisons (IQ = 70).

Figure S8: Burden contrasts including SRS and IQ

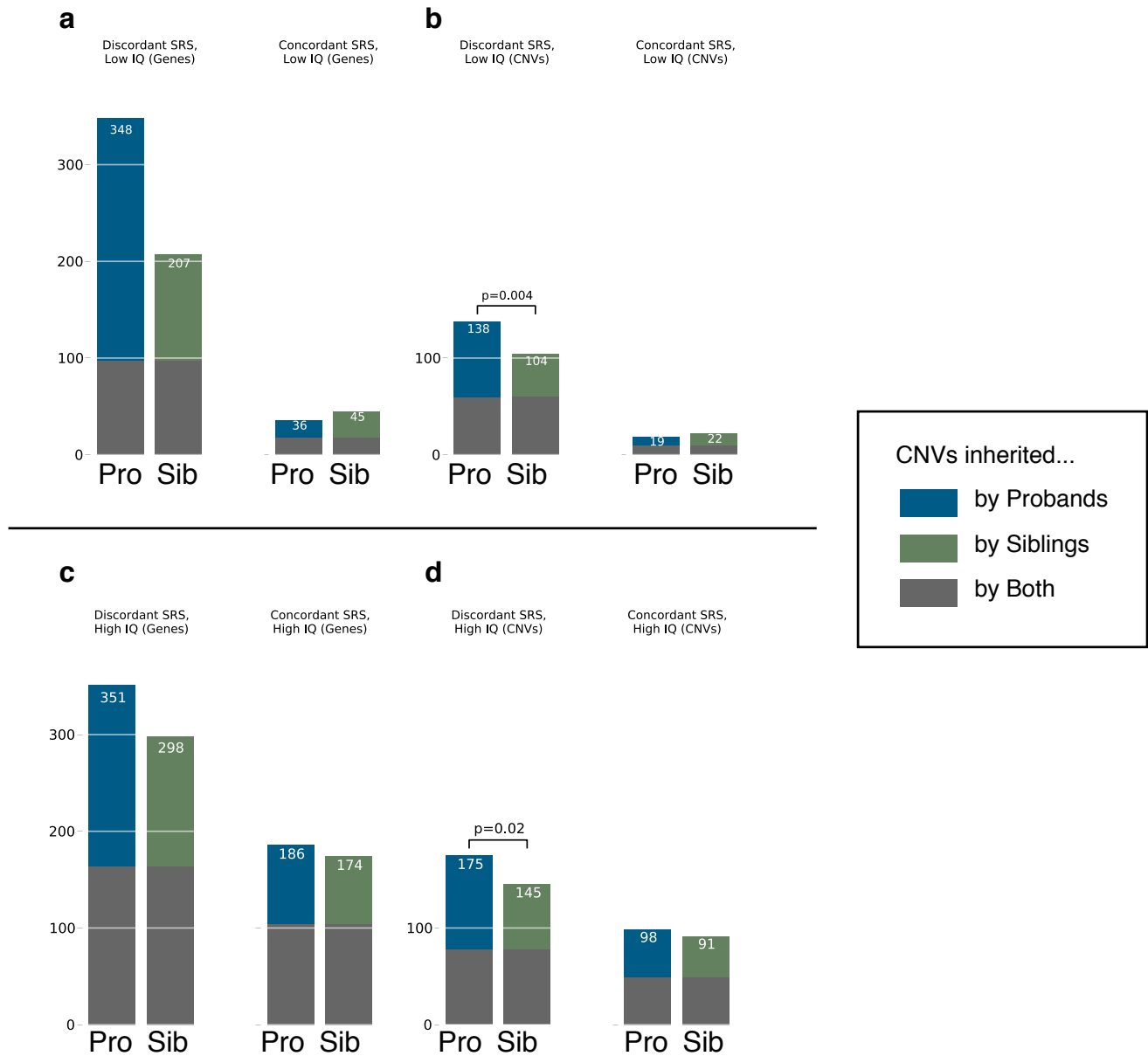


Figure S8. Burden between SRS and IQ in probands and siblings. **(a)** Genic burden and **(b)** CNV burden for proband-sibling pairs where the proband has low IQ (< 70) for both discordant and concordant SRS quads. Burden for probands with high IQ (≥70) shown in **(c)** and **(d)**. P-value bars drawn if two-tailed paired t-test p value is less than 0.05.

Figure S9: Enrichment of brain expressed genes in SRS discordant quads (A) and all quads (B)

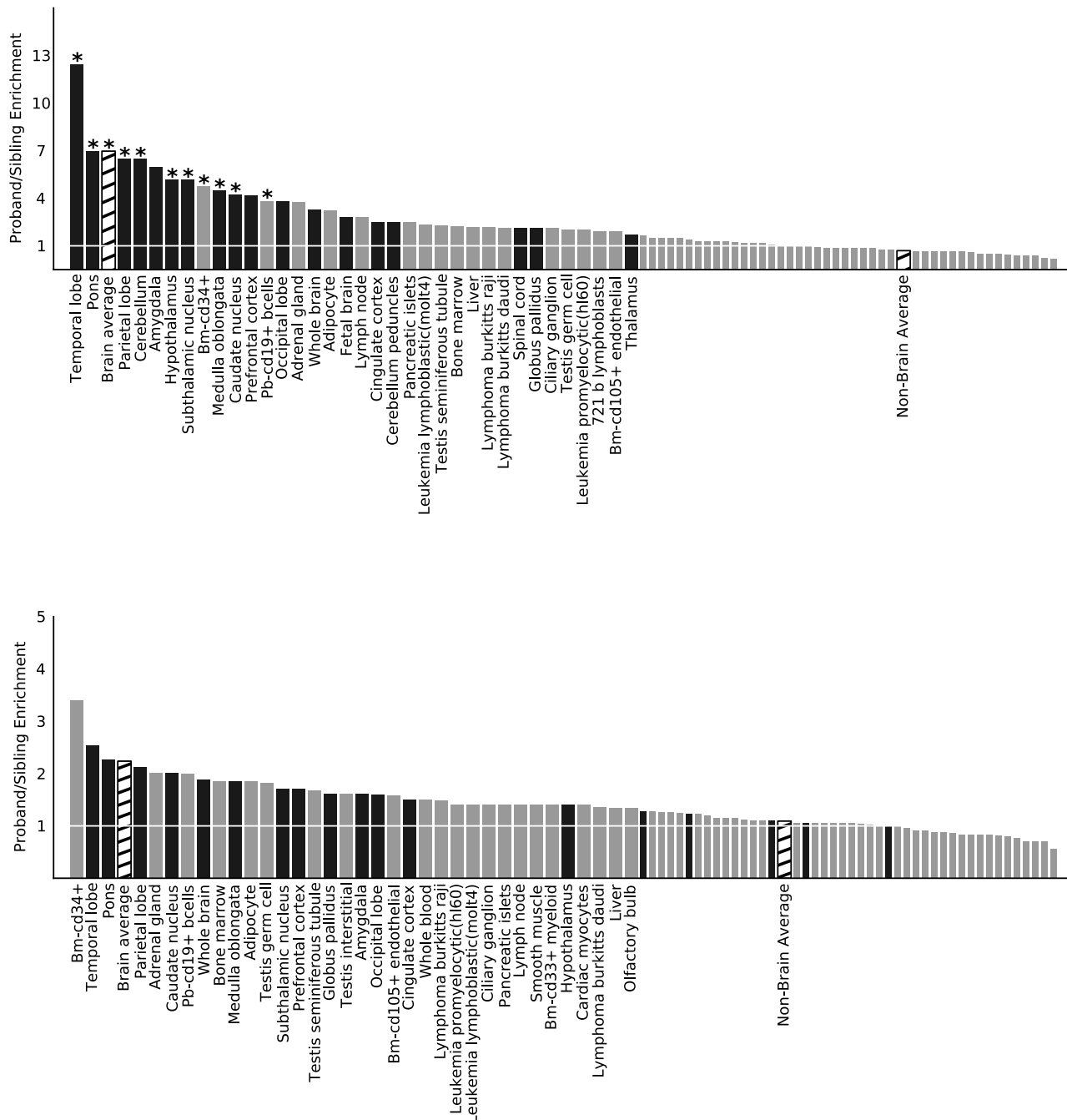


Figure S9. Enrichment of brain expressed genes in probands vs. siblings Bars (y-axis) represent ratio of enrichment between proband and siblings for genes highly expressed in each tissue (defined as top 5%, see Methods). Black bars: tissue is part of brain or nervous system; white bars: non-brain or nervous system tissues; hatched bars: are computed averages. Asterix indicates significance using a FDR-based multiple testing correction q -value < 0.05 . **(a)** probands from SRS discordant quads only show greater enrichment for brain-expressed genes than do all quads, **(b)**.

Figure S10: Intersection between brain-expressed genes and previously associated genes in proband CNVs, but not sibling CNVs

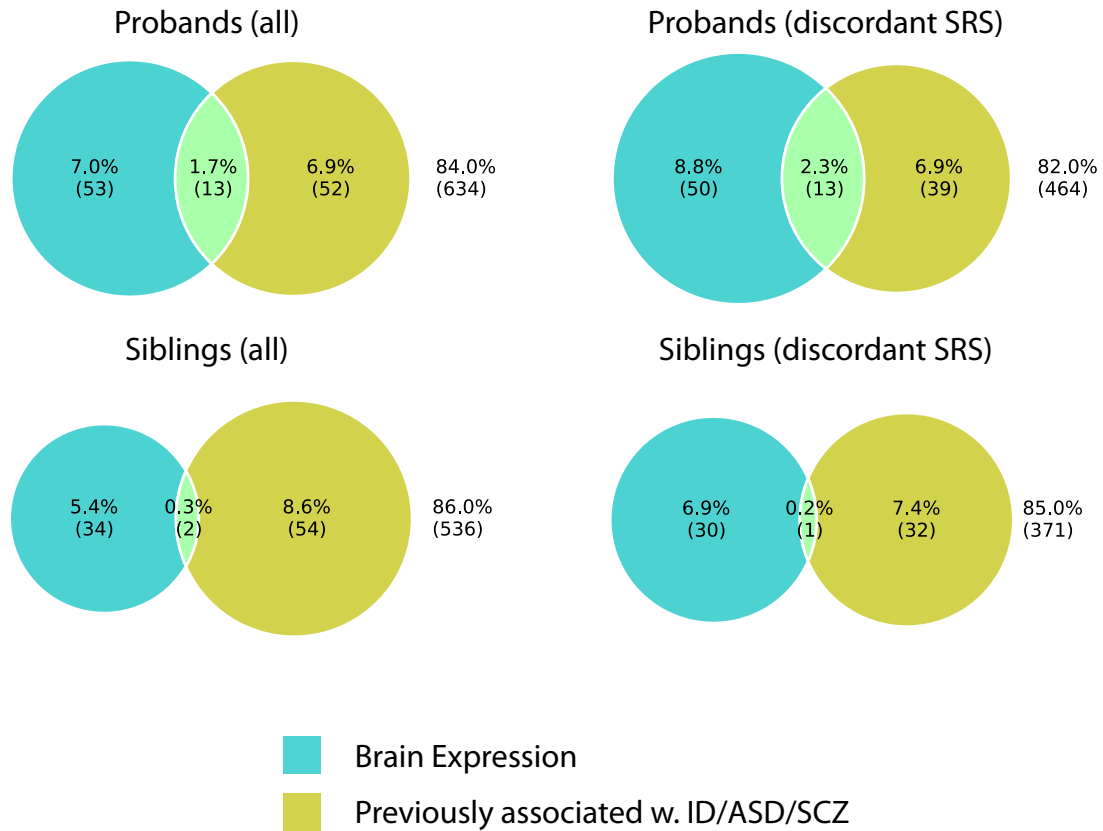


Figure S10. Intersection of brain-expressed and disease genes in probands
 We intersected the sets of genes found in probands (top row) and siblings (bottom row) that were either brain expressed (teal circles) or had previously been observed in ASD/Schizophrenia/ID (yellow circles). Probands—especially those in SRS discordant pairs— had a higher fraction of intersecting genes (13 genes, Table S11) than other groups or their siblings, suggesting that these genes may be top candidates for follow-up study in the pathogenesis of ASD.

Table S1

	Sample	Source	SNV data available	Part of Sanders 2011 (Illumina SNP)	Part of QC analysis	CoNIFER Std Dev
	11000.fa	Sanders et al	TRUE	TRUE	FALSE	0.30
	11000.mo	Sanders et al	TRUE	TRUE	FALSE	0.30
	11000.p1	Sanders et al	TRUE	TRUE	FALSE	0.28
	11000.s1	Sanders et al	TRUE	TRUE	FALSE	0.37
	11008.fa	Sanders et al	TRUE	TRUE	FALSE	0.27
	11008.mo	Sanders et al	TRUE	TRUE	FALSE	0.27
	11008.p1	Sanders et al	TRUE	TRUE	FALSE	0.27
	11008.s1	Sanders et al	TRUE	TRUE	FALSE	0.28
	11010.fa	Sanders et al	TRUE	TRUE	FALSE	0.28
	11010.mo	Sanders et al	TRUE	TRUE	FALSE	0.29
	11010.p1	Sanders et al	TRUE	TRUE	FALSE	0.28
	11010.s1	Sanders et al	TRUE	TRUE	FALSE	0.30
	11013.fa	O'Roak et al	TRUE	TRUE	FALSE	0.38
	11013.mo	O'Roak et al	TRUE	TRUE	FALSE	0.42
	11013.p1	O'Roak et al	TRUE	TRUE	FALSE	0.39
	11013.s1	O'Roak et al	TRUE	TRUE	TRUE	0.36
	11014.fa	Sanders et al	TRUE	TRUE	FALSE	0.24
	11014.mo	Sanders et al	TRUE	TRUE	FALSE	0.26
	11014.p1	Sanders et al	TRUE	TRUE	FALSE	0.25
	11014.s1	Sanders et al	TRUE	TRUE	FALSE	0.26
	11029.fa	O'Roak et al	TRUE	TRUE	FALSE	0.39
	11029.mo	O'Roak et al	TRUE	TRUE	FALSE	0.40
	11029.p1	O'Roak et al	TRUE	TRUE	TRUE	0.38
	11029.s1	O'Roak et al	TRUE	TRUE	TRUE	0.39
	11045.fa	Sanders et al	TRUE	TRUE	FALSE	0.32
	11045.mo	Sanders et al	TRUE	TRUE	FALSE	0.29
	11045.p1	Sanders et al	TRUE	TRUE	TRUE	0.34
	11045.s1	Sanders et al	TRUE	TRUE	FALSE	0.28
	11057.fa	Sanders et al	TRUE	TRUE	FALSE	0.24
	11057.mo	Sanders et al	TRUE	TRUE	FALSE	0.25
	11057.p1	Sanders et al	TRUE	TRUE	FALSE	0.24
	11057.s1	Sanders et al	TRUE	TRUE	FALSE	0.23
	11060.fa	Sanders et al	TRUE	TRUE	FALSE	0.31
	11060.mo	Sanders et al	TRUE	TRUE	FALSE	0.32
	11060.p1	Sanders et al	TRUE	TRUE	FALSE	0.38
	11060.s2	Sanders et al	TRUE	FALSE	FALSE	0.35
	11066.fa	Sanders et al	TRUE	TRUE	FALSE	0.37
	11066.mo	Sanders et al	TRUE	TRUE	FALSE	0.30
	11066.p1	Sanders et al	TRUE	TRUE	FALSE	0.23
	11066.s2	Sanders et al	TRUE	FALSE	FALSE	0.34
	11067.fa	Sanders et al	TRUE	TRUE	FALSE	0.26
	11067.mo	Sanders et al	TRUE	TRUE	FALSE	0.21
	11067.p1	Sanders et al	TRUE	TRUE	FALSE	0.23
	11067.s1	Sanders et al	TRUE	TRUE	FALSE	0.24
	11074.fa	Sanders et al	TRUE	TRUE	FALSE	0.24
	11074.mo	Sanders et al	TRUE	TRUE	FALSE	0.24
	11074.p1	Sanders et al	TRUE	TRUE	FALSE	0.26
	11074.s1	Sanders et al	TRUE	TRUE	FALSE	0.27
	11075.fa	Sanders et al	TRUE	TRUE	FALSE	0.31
	11075.mo	Sanders et al	TRUE	TRUE	FALSE	0.20
	11075.p1	Sanders et al	TRUE	TRUE	FALSE	0.24
	11075.s1	Sanders et al	TRUE	TRUE	FALSE	0.23
	11077.fa	Sanders et al	TRUE	TRUE	FALSE	0.26
	11077.mo	Sanders et al	TRUE	TRUE	FALSE	0.21
	11077.p1	Sanders et al	TRUE	TRUE	FALSE	0.22
	11077.s1	Sanders et al	TRUE	TRUE	FALSE	0.27
	11079.fa	Sanders et al	TRUE	TRUE	FALSE	0.29
	11079.mo	Sanders et al	TRUE	TRUE	FALSE	0.30
	11079.p1	Sanders et al	TRUE	TRUE	FALSE	0.22
	11079.s1	Sanders et al	TRUE	TRUE	FALSE	0.43
	11085.fa	Sanders et al	TRUE	TRUE	FALSE	0.30
	11085.mo	Sanders et al	TRUE	TRUE	FALSE	0.30
	11085.p1	Sanders et al	TRUE	TRUE	FALSE	0.27
	11085.s1	Sanders et al	TRUE	TRUE	FALSE	0.30
	11089.fa	Sanders et al	TRUE	TRUE	FALSE	0.24
	11089.mo	Sanders et al	TRUE	TRUE	FALSE	0.23
	11089.p1	Sanders et al	TRUE	TRUE	FALSE	0.23
	11089.s1	Sanders et al	TRUE	TRUE	FALSE	0.25
	11090.fa	Sanders et al	TRUE	TRUE	FALSE	0.26
	11090.mo	Sanders et al	TRUE	TRUE	FALSE	0.31
	11090.p1	Sanders et al	TRUE	TRUE	TRUE	0.29
	11090.s1	Sanders et al	TRUE	TRUE	FALSE	0.39
	11092.fa	Sanders et al	TRUE	TRUE	FALSE	0.27
	11092.mo	Sanders et al	TRUE	TRUE	FALSE	0.30
	11092.p1	Sanders et al	TRUE	TRUE	FALSE	0.29
	11092.s1	Sanders et al	TRUE	TRUE	FALSE	0.29
	11094.fa	Sanders et al	TRUE	TRUE	FALSE	0.24
	11094.mo	Sanders et al	TRUE	TRUE	FALSE	0.22
	11094.p1	Sanders et al	TRUE	TRUE	FALSE	0.39
	11094.s1	Sanders et al	TRUE	TRUE	FALSE	0.24
	11107.fa	Sanders et al	TRUE	TRUE	FALSE	0.24
	11107.mo	Sanders et al	TRUE	TRUE	FALSE	0.24
	11107.p1	Sanders et al	TRUE	TRUE	FALSE	0.27
	11107.s1	Sanders et al	TRUE	TRUE	FALSE	0.25
	11108.fa	Sanders et al	TRUE	TRUE	FALSE	0.32

11108.mo	Sanders et al	TRUE	TRUE	FALSE	0.32
11108.p1	Sanders et al	TRUE	TRUE	FALSE	0.30
11108.s1	Sanders et al	TRUE	TRUE	FALSE	0.38
11114.fa	Sanders et al	TRUE	TRUE	FALSE	0.28
11114.mo	Sanders et al	TRUE	TRUE	FALSE	0.30
11114.p1	Sanders et al	TRUE	TRUE	FALSE	0.29
11114.s1	Sanders et al	TRUE	TRUE	FALSE	0.28
11115.fa	Sanders et al	TRUE	TRUE	FALSE	0.29
11115.mo	Sanders et al	TRUE	TRUE	FALSE	0.40
11115.p1	Sanders et al	TRUE	TRUE	FALSE	0.31
11115.s1	Sanders et al	TRUE	TRUE	TRUE	0.32
11117.fa	Sanders et al	TRUE	TRUE	FALSE	0.24
11117.mo	Sanders et al	TRUE	TRUE	FALSE	0.31
11117.p1	Sanders et al	TRUE	TRUE	FALSE	0.24
11117.s1	Sanders et al	TRUE	TRUE	FALSE	0.35
11118.fa	Sanders et al	TRUE	TRUE	FALSE	0.46
11118.mo	Sanders et al	TRUE	TRUE	FALSE	0.31
11118.p1	Sanders et al	TRUE	TRUE	TRUE	0.30
11118.s1	Sanders et al	TRUE	TRUE	FALSE	0.35
11132.fa	Sanders et al	TRUE	TRUE	FALSE	0.25
11132.mo	Sanders et al	TRUE	TRUE	FALSE	0.23
11132.p1	Sanders et al	TRUE	TRUE	FALSE	0.21
11132.s1	Sanders et al	TRUE	TRUE	FALSE	0.31
11146.fa	Sanders et al	TRUE	TRUE	FALSE	0.25
11146.mo	Sanders et al	TRUE	TRUE	FALSE	0.27
11146.p1	Sanders et al	TRUE	TRUE	FALSE	0.35
11146.s1	Sanders et al	TRUE	TRUE	FALSE	0.36
11154.fa	Sanders et al	TRUE	TRUE	FALSE	0.29
11154.mo	Sanders et al	TRUE	TRUE	FALSE	0.29
11154.p1	Sanders et al	TRUE	TRUE	FALSE	0.29
11154.s1	Sanders et al	TRUE	TRUE	FALSE	0.42
11172.fa	O'Roak et al	TRUE	TRUE	FALSE	0.37
11172.mo	O'Roak et al	TRUE	TRUE	FALSE	0.44
11172.p1	O'Roak et al	TRUE	TRUE	FALSE	0.39
11172.s1	O'Roak et al	TRUE	TRUE	FALSE	0.36
11180.fa	Sanders et al	TRUE	FALSE	FALSE	0.31
11180.mo	Sanders et al	TRUE	FALSE	FALSE	0.33
11180.p1	Sanders et al	TRUE	FALSE	FALSE	0.28
11180.s1	Sanders et al	TRUE	FALSE	FALSE	0.33
11190.fa	O'Roak et al	TRUE	TRUE	FALSE	0.55
11190.mo	O'Roak et al	TRUE	TRUE	FALSE	0.46
11190.p1	O'Roak et al	TRUE	TRUE	FALSE	0.36
11190.s1	O'Roak et al	TRUE	TRUE	FALSE	0.41
11196.fa	Sanders et al	TRUE	TRUE	FALSE	0.22
11196.mo	Sanders et al	TRUE	TRUE	FALSE	0.24
11196.p1	Sanders et al	TRUE	TRUE	FALSE	0.24
11196.s1	Sanders et al	TRUE	TRUE	TRUE	0.23
11203.fa	Sanders et al	TRUE	TRUE	FALSE	0.23
11203.mo	Sanders et al	TRUE	TRUE	FALSE	0.22
11203.p1	Sanders et al	TRUE	TRUE	FALSE	0.21
11203.s1	Sanders et al	TRUE	TRUE	FALSE	0.25
11219.fa	Sanders et al	TRUE	TRUE	FALSE	0.27
11219.mo	Sanders et al	TRUE	TRUE	FALSE	0.26
11219.p1	Sanders et al	TRUE	TRUE	FALSE	0.29
11219.s1	Sanders et al	TRUE	TRUE	FALSE	0.23
11220.fa	Sanders et al	TRUE	TRUE	FALSE	0.27
11220.mo	Sanders et al	TRUE	TRUE	FALSE	0.27
11220.p1	Sanders et al	TRUE	TRUE	FALSE	0.26
11220.s1	Sanders et al	TRUE	TRUE	FALSE	0.28
11229.fa	O'Roak et al	TRUE	TRUE	FALSE	0.45
11229.mo	O'Roak et al	TRUE	TRUE	FALSE	0.44
11229.p1	O'Roak et al	TRUE	TRUE	TRUE	0.42
11229.s1	O'Roak et al	TRUE	TRUE	TRUE	0.39
11241.fa	Sanders et al	TRUE	TRUE	FALSE	0.28
11241.mo	Sanders et al	TRUE	TRUE	FALSE	0.29
11241.p1	Sanders et al	TRUE	TRUE	TRUE	0.28
11241.s1	Sanders et al	TRUE	TRUE	FALSE	0.28
11242.fa	Sanders et al	TRUE	TRUE	FALSE	0.27
11242.mo	Sanders et al	TRUE	TRUE	FALSE	0.24
11242.p1	Sanders et al	TRUE	TRUE	FALSE	0.26
11242.s1	Sanders et al	TRUE	TRUE	FALSE	0.26
11247.fa	Sanders et al	TRUE	TRUE	FALSE	0.26
11247.mo	Sanders et al	TRUE	TRUE	FALSE	0.23
11247.p1	Sanders et al	TRUE	TRUE	FALSE	0.29
11247.s1	Sanders et al	TRUE	TRUE	FALSE	0.21
11252.fa	Sanders et al	TRUE	TRUE	FALSE	0.31
11252.mo	Sanders et al	TRUE	TRUE	FALSE	0.32
11252.p1	Sanders et al	TRUE	TRUE	TRUE	0.26
11252.s1	Sanders et al	TRUE	TRUE	FALSE	0.31
11265.fa	Sanders et al	TRUE	TRUE	FALSE	0.31
11265.mo	Sanders et al	TRUE	TRUE	FALSE	0.31
11265.p1	Sanders et al	TRUE	TRUE	FALSE	0.30
11265.s1	Sanders et al	TRUE	TRUE	FALSE	0.38
11267.fa	Sanders et al	TRUE	TRUE	FALSE	0.29
11267.mo	Sanders et al	TRUE	TRUE	FALSE	0.31
11267.p1	Sanders et al	TRUE	TRUE	FALSE	0.35
11267.s1	Sanders et al	TRUE	TRUE	TRUE	0.32
11282.fa	Sanders et al	TRUE	TRUE	FALSE	0.31
11282.mo	Sanders et al	TRUE	TRUE	FALSE	0.31

11282.p1	Sanders et al	TRUE	TRUE	TRUE	0.32
11282.s1	Sanders et al	TRUE	TRUE	TRUE	0.35
11285.fa	Sanders et al	TRUE	TRUE	FALSE	0.27
11285.mo	Sanders et al	TRUE	TRUE	FALSE	0.29
11285.p1	Sanders et al	TRUE	TRUE	FALSE	0.29
11285.s1	Sanders et al	TRUE	TRUE	FALSE	0.28
11290.fa	Sanders et al	TRUE	TRUE	FALSE	0.29
11290.mo	Sanders et al	TRUE	TRUE	FALSE	0.30
11290.p1	Sanders et al	TRUE	TRUE	FALSE	0.24
11290.s1	Sanders et al	TRUE	TRUE	FALSE	0.36
11291.fa	O'Roak et al	TRUE	TRUE	FALSE	0.39
11291.mo	O'Roak et al	TRUE	TRUE	FALSE	0.41
11291.p1	O'Roak et al	TRUE	TRUE	FALSE	0.41
11291.s1	O'Roak et al	TRUE	TRUE	FALSE	0.37
11298.fa	Sanders et al	TRUE	TRUE	FALSE	0.35
11298.mo	Sanders et al	TRUE	TRUE	FALSE	0.32
11298.p1	Sanders et al	TRUE	TRUE	FALSE	0.29
11298.s1	Sanders et al	TRUE	TRUE	FALSE	0.31
11301.fa	Sanders et al	TRUE	TRUE	FALSE	0.27
11301.mo	Sanders et al	TRUE	TRUE	FALSE	0.23
11301.p1	Sanders et al	TRUE	TRUE	FALSE	0.26
11301.s1	Sanders et al	TRUE	TRUE	FALSE	0.26
11304.fa	Sanders et al	TRUE	TRUE	FALSE	0.30
11304.mo	Sanders et al	TRUE	TRUE	FALSE	0.32
11304.p1	Sanders et al	TRUE	TRUE	FALSE	0.31
11304.s1	Sanders et al	TRUE	TRUE	TRUE	0.26
11316.fa	Sanders et al	TRUE	TRUE	FALSE	0.28
11316.mo	Sanders et al	TRUE	TRUE	FALSE	0.29
11316.p1	Sanders et al	TRUE	TRUE	FALSE	0.30
11316.s1	Sanders et al	TRUE	TRUE	FALSE	0.28
11336.fa	Sanders et al	TRUE	TRUE	FALSE	0.27
11336.mo	Sanders et al	TRUE	TRUE	FALSE	0.22
11336.p1	Sanders et al	TRUE	TRUE	FALSE	0.23
11336.s1	Sanders et al	TRUE	TRUE	FALSE	0.23
11353.fa	Sanders et al	TRUE	TRUE	FALSE	0.31
11353.mo	Sanders et al	TRUE	TRUE	FALSE	0.40
11353.p1	Sanders et al	TRUE	TRUE	FALSE	0.30
11353.s1	Sanders et al	TRUE	TRUE	FALSE	0.36
11356.fa	Sanders et al	TRUE	TRUE	FALSE	0.21
11356.mo	Sanders et al	TRUE	TRUE	FALSE	0.21
11356.p1	Sanders et al	TRUE	TRUE	FALSE	0.22
11356.s1	Sanders et al	TRUE	TRUE	FALSE	0.25
11364.fa	O'Roak et al	TRUE	TRUE	FALSE	0.38
11364.mo	O'Roak et al	TRUE	TRUE	FALSE	0.41
11364.p1	O'Roak et al	TRUE	TRUE	FALSE	0.39
11364.s1	O'Roak et al	TRUE	TRUE	FALSE	0.39
11382.fa	Sanders et al	TRUE	TRUE	FALSE	0.23
11382.mo	Sanders et al	TRUE	TRUE	FALSE	0.27
11382.p1	Sanders et al	TRUE	TRUE	FALSE	0.24
11382.s1	Sanders et al	TRUE	TRUE	FALSE	0.26
11390.fa	O'Roak et al	TRUE	TRUE	FALSE	0.32
11390.mo	O'Roak et al	TRUE	TRUE	FALSE	0.36
11390.p1	O'Roak et al	TRUE	TRUE	FALSE	0.35
11390.s1	O'Roak et al	TRUE	TRUE	FALSE	0.38
11411.fa	Sanders et al	TRUE	FALSE	FALSE	0.22
11411.mo	Sanders et al	TRUE	FALSE	FALSE	0.25
11411.p1	Sanders et al	TRUE	FALSE	FALSE	0.21
11411.s1	Sanders et al	TRUE	FALSE	FALSE	0.26
11412.fa	Sanders et al	TRUE	TRUE	FALSE	0.26
11412.mo	Sanders et al	TRUE	TRUE	FALSE	0.28
11412.p1	Sanders et al	TRUE	TRUE	TRUE	0.27
11412.s1	Sanders et al	TRUE	TRUE	TRUE	0.30
11429.fa	Sanders et al	TRUE	TRUE	FALSE	0.27
11429.mo	Sanders et al	TRUE	TRUE	FALSE	0.30
11429.p1	Sanders et al	TRUE	TRUE	FALSE	0.27
11429.s1	Sanders et al	TRUE	TRUE	FALSE	0.26
11433.fa	Sanders et al	TRUE	TRUE	FALSE	0.30
11433.mo	Sanders et al	TRUE	TRUE	FALSE	0.30
11433.p1	Sanders et al	TRUE	TRUE	FALSE	0.35
11433.s1	Sanders et al	TRUE	TRUE	TRUE	0.41
11437.fa	Sanders et al	TRUE	TRUE	FALSE	0.32
11437.mo	Sanders et al	TRUE	TRUE	FALSE	0.35
11437.p1	Sanders et al	TRUE	TRUE	FALSE	0.35
11437.s1	Sanders et al	TRUE	TRUE	FALSE	0.46
11452.fa	O'Roak et al	TRUE	TRUE	FALSE	0.39
11452.mo	O'Roak et al	TRUE	TRUE	FALSE	0.44
11452.p1	O'Roak et al	TRUE	TRUE	FALSE	0.47
11452.s1	O'Roak et al	TRUE	TRUE	FALSE	0.37
11456.fa	Sanders et al	TRUE	TRUE	FALSE	0.29
11456.mo	Sanders et al	TRUE	TRUE	FALSE	0.31
11456.p1	Sanders et al	TRUE	TRUE	FALSE	0.29
11456.s1	Sanders et al	TRUE	TRUE	TRUE	0.29
11459.fa	O'Roak et al	TRUE	TRUE	FALSE	0.38
11459.mo	O'Roak et al	TRUE	TRUE	FALSE	0.41
11459.p1	O'Roak et al	TRUE	TRUE	FALSE	0.44
11459.s1	O'Roak et al	TRUE	TRUE	FALSE	0.36
11462.fa	Sanders et al	TRUE	TRUE	FALSE	0.24
11462.mo	Sanders et al	TRUE	TRUE	FALSE	0.26
11462.p1	Sanders et al	TRUE	TRUE	FALSE	0.22

11462.s1	Sanders et al	TRUE	TRUE	FALSE	0.26
11469.fa	O'Roak et al	TRUE	TRUE	FALSE	0.42
11469.mo	O'Roak et al	TRUE	TRUE	FALSE	0.46
11469.p1	O'Roak et al	TRUE	TRUE	FALSE	0.44
11469.s1	O'Roak et al	TRUE	TRUE	FALSE	0.43
11472.fa	O'Roak et al	TRUE	TRUE	FALSE	0.48
11472.mo	O'Roak et al	TRUE	TRUE	FALSE	0.42
11472.p1	O'Roak et al	TRUE	TRUE	FALSE	0.44
11472.s1	O'Roak et al	TRUE	TRUE	FALSE	0.44
11474.fa	Sanders et al	TRUE	TRUE	FALSE	0.34
11474.mo	Sanders et al	TRUE	TRUE	FALSE	0.31
11474.p1	Sanders et al	TRUE	TRUE	FALSE	0.26
11474.s1	Sanders et al	TRUE	TRUE	FALSE	0.24
11479.fa	O'Roak et al	TRUE	TRUE	FALSE	0.37
11479.mo	O'Roak et al	TRUE	TRUE	FALSE	0.48
11479.p1	O'Roak et al	TRUE	TRUE	FALSE	0.39
11479.s1	O'Roak et al	TRUE	TRUE	FALSE	0.38
11484.fa	Sanders et al	TRUE	TRUE	FALSE	0.25
11484.mo	Sanders et al	TRUE	TRUE	FALSE	0.25
11484.p1	Sanders et al	TRUE	TRUE	FALSE	0.26
11484.s1	Sanders et al	TRUE	TRUE	TRUE	0.24
11490.fa	Sanders et al	TRUE	TRUE	FALSE	0.29
11490.mo	Sanders et al	TRUE	TRUE	FALSE	0.30
11490.p1	Sanders et al	TRUE	TRUE	FALSE	0.25
11490.s1	Sanders et al	TRUE	TRUE	FALSE	0.43
11491.fa	O'Roak et al	TRUE	TRUE	FALSE	0.34
11491.mo	O'Roak et al	TRUE	TRUE	FALSE	0.37
11491.p1	O'Roak et al	TRUE	TRUE	FALSE	0.38
11491.s1	O'Roak et al	TRUE	TRUE	FALSE	0.35
11501.fa	Sanders et al	TRUE	TRUE	FALSE	0.28
11501.mo	Sanders et al	TRUE	TRUE	FALSE	0.27
11501.p1	Sanders et al	TRUE	TRUE	FALSE	0.27
11501.s1	Sanders et al	TRUE	TRUE	FALSE	0.26
11509.fa	Sanders et al	TRUE	TRUE	FALSE	0.26
11509.mo	Sanders et al	TRUE	TRUE	FALSE	0.25
11509.p1	Sanders et al	TRUE	TRUE	FALSE	0.25
11509.s1	Sanders et al	TRUE	TRUE	FALSE	0.26
11519.fa	Sanders et al	TRUE	TRUE	FALSE	0.27
11519.mo	Sanders et al	TRUE	TRUE	FALSE	0.26
11519.p1	Sanders et al	TRUE	TRUE	TRUE	0.29
11519.s1	Sanders et al	TRUE	TRUE	FALSE	0.34
11524.fa	Sanders et al	TRUE	TRUE	FALSE	0.29
11524.mo	Sanders et al	TRUE	TRUE	FALSE	0.29
11524.p1	Sanders et al	TRUE	TRUE	FALSE	0.26
11524.s1	Sanders et al	TRUE	TRUE	FALSE	0.27
11532.fa	Sanders et al	TRUE	TRUE	FALSE	0.30
11532.mo	Sanders et al	TRUE	TRUE	FALSE	0.30
11532.p1	Sanders et al	TRUE	TRUE	TRUE	0.30
11532.s1	Sanders et al	TRUE	TRUE	FALSE	0.26
11551.fa	Sanders et al	TRUE	TRUE	FALSE	0.24
11551.mo	Sanders et al	TRUE	TRUE	FALSE	0.28
11551.p1	Sanders et al	TRUE	TRUE	TRUE	0.25
11551.s1	Sanders et al	TRUE	TRUE	FALSE	0.29
11561.fa	Sanders et al	TRUE	TRUE	FALSE	0.26
11561.mo	Sanders et al	TRUE	TRUE	FALSE	0.29
11561.p1	Sanders et al	TRUE	TRUE	FALSE	0.23
11561.s1	Sanders et al	TRUE	TRUE	FALSE	0.35
11569.fa	O'Roak et al	TRUE	TRUE	FALSE	0.30
11569.mo	O'Roak et al	TRUE	TRUE	FALSE	0.38
11569.p1	O'Roak et al	TRUE	TRUE	TRUE	0.39
11569.s1	O'Roak et al	TRUE	TRUE	FALSE	0.35
11571.fa	O'Roak et al	TRUE	TRUE	FALSE	0.41
11571.mo	O'Roak et al	TRUE	TRUE	FALSE	0.42
11571.p1	O'Roak et al	TRUE	TRUE	FALSE	0.41
11571.s1	O'Roak et al	TRUE	TRUE	FALSE	0.38
11581.fa	Sanders et al	TRUE	TRUE	FALSE	0.25
11581.mo	Sanders et al	TRUE	TRUE	FALSE	0.28
11581.p1	Sanders et al	TRUE	TRUE	FALSE	0.27
11581.s1	Sanders et al	TRUE	TRUE	FALSE	0.26
11610.fa	O'Roak et al	TRUE	TRUE	FALSE	0.38
11610.mo	O'Roak et al	TRUE	TRUE	FALSE	0.41
11610.p1	O'Roak et al	TRUE	TRUE	FALSE	0.35
11610.s1	O'Roak et al	TRUE	TRUE	FALSE	0.38
11611.fa	Sanders et al	TRUE	TRUE	FALSE	0.29
11611.mo	Sanders et al	TRUE	TRUE	FALSE	0.29
11611.p1	Sanders et al	TRUE	TRUE	FALSE	0.27
11611.s1	Sanders et al	TRUE	FALSE	FALSE	0.32
11622.fa	Sanders et al	TRUE	TRUE	FALSE	0.27
11622.mo	Sanders et al	TRUE	TRUE	FALSE	0.25
11622.p1	Sanders et al	TRUE	TRUE	FALSE	0.30
11622.s1	Sanders et al	TRUE	TRUE	FALSE	0.24
11629.fa	O'Roak et al	TRUE	TRUE	FALSE	0.30
11629.mo	O'Roak et al	TRUE	TRUE	FALSE	0.40
11629.p1	O'Roak et al	TRUE	TRUE	FALSE	0.35
11629.s1	O'Roak et al	TRUE	TRUE	FALSE	0.38
11638.fa	O'Roak et al	TRUE	TRUE	FALSE	0.48
11638.mo	O'Roak et al	TRUE	TRUE	FALSE	0.37
11638.p1	O'Roak et al	TRUE	TRUE	FALSE	0.37
11638.s1	O'Roak et al	TRUE	TRUE	FALSE	0.39

11641.fa	Sanders et al	TRUE	TRUE	FALSE	0.27
11641.mo	Sanders et al	TRUE	TRUE	FALSE	0.24
11641.p1	Sanders et al	TRUE	TRUE	FALSE	0.33
11641.s1	Sanders et al	TRUE	TRUE	FALSE	0.24
11654.fa	Sanders et al	TRUE	TRUE	FALSE	0.35
11654.mo	Sanders et al	TRUE	TRUE	FALSE	0.30
11654.p1	Sanders et al	TRUE	TRUE	FALSE	0.32
11654.s1	Sanders et al	TRUE	TRUE	FALSE	0.27
11659.fa	O'Roak et al	TRUE	TRUE	FALSE	0.37
11659.mo	O'Roak et al	TRUE	TRUE	FALSE	0.38
11659.p1	O'Roak et al	TRUE	TRUE	FALSE	0.40
11659.s1	O'Roak et al	TRUE	TRUE	TRUE	0.32
11667.fa	Sanders et al	TRUE	TRUE	FALSE	0.34
11667.mo	Sanders et al	TRUE	TRUE	FALSE	0.33
11667.p1	Sanders et al	TRUE	TRUE	TRUE	0.34
11667.s1	Sanders et al	TRUE	TRUE	TRUE	0.27
11676.fa	Sanders et al	TRUE	TRUE	FALSE	0.22
11676.mo	Sanders et al	TRUE	TRUE	FALSE	0.22
11676.p1	Sanders et al	TRUE	TRUE	FALSE	0.22
11676.s1	Sanders et al	TRUE	TRUE	FALSE	0.26
11691.fa	O'Roak et al	TRUE	TRUE	FALSE	0.36
11691.mo	O'Roak et al	TRUE	TRUE	FALSE	0.40
11691.p1	O'Roak et al	TRUE	TRUE	FALSE	0.37
11691.s1	O'Roak et al	TRUE	TRUE	FALSE	0.37
11696.fa	O'Roak et al	TRUE	TRUE	FALSE	0.47
11696.mo	O'Roak et al	TRUE	TRUE	FALSE	0.35
11696.p1	O'Roak et al	TRUE	TRUE	FALSE	0.44
11696.s1	O'Roak et al	TRUE	TRUE	TRUE	0.39
11700.fa	Sanders et al	TRUE	TRUE	FALSE	0.26
11700.mo	Sanders et al	TRUE	TRUE	FALSE	0.27
11700.p1	Sanders et al	TRUE	TRUE	FALSE	0.26
11700.s1	Sanders et al	TRUE	TRUE	FALSE	0.27
11711.fa	O'Roak et al	TRUE	TRUE	FALSE	0.40
11711.mo	O'Roak et al	TRUE	TRUE	FALSE	0.39
11711.p1	O'Roak et al	TRUE	TRUE	FALSE	0.40
11711.s1	O'Roak et al	TRUE	TRUE	FALSE	0.38
11715.fa	O'Roak et al	TRUE	TRUE	FALSE	0.40
11715.mo	O'Roak et al	TRUE	TRUE	FALSE	0.36
11715.p1	O'Roak et al	TRUE	TRUE	FALSE	0.42
11715.s1	O'Roak et al	TRUE	TRUE	TRUE	0.38
11716.fa	Sanders et al	TRUE	TRUE	FALSE	0.27
11716.mo	Sanders et al	TRUE	TRUE	FALSE	0.26
11716.p1	Sanders et al	TRUE	TRUE	FALSE	0.26
11716.s1	Sanders et al	TRUE	TRUE	FALSE	0.28
11720.fa	Sanders et al	TRUE	TRUE	FALSE	0.27
11720.mo	Sanders et al	TRUE	TRUE	FALSE	0.42
11720.p1	Sanders et al	TRUE	TRUE	FALSE	0.33
11720.s1	Sanders et al	TRUE	TRUE	FALSE	0.36
11722.fa	O'Roak et al	TRUE	TRUE	FALSE	0.39
11722.mo	O'Roak et al	TRUE	TRUE	FALSE	0.36
11722.p1	O'Roak et al	TRUE	TRUE	FALSE	0.40
11722.s1	O'Roak et al	TRUE	TRUE	FALSE	0.37
11724.fa	Sanders et al	TRUE	TRUE	FALSE	0.47
11724.mo	Sanders et al	TRUE	TRUE	FALSE	0.38
11724.p1	Sanders et al	TRUE	TRUE	FALSE	0.36
11724.s1	Sanders et al	TRUE	TRUE	FALSE	0.29
11740.fa	Sanders et al	TRUE	TRUE	FALSE	0.25
11740.mo	Sanders et al	TRUE	TRUE	FALSE	0.26
11740.p1	Sanders et al	TRUE	TRUE	FALSE	0.25
11740.s1	Sanders et al	TRUE	TRUE	FALSE	0.27
11766.fa	Sanders et al	TRUE	TRUE	FALSE	0.25
11766.mo	Sanders et al	TRUE	TRUE	FALSE	0.27
11766.p1	Sanders et al	TRUE	TRUE	FALSE	0.29
11766.s1	Sanders et al	TRUE	TRUE	FALSE	0.28
11773.fa	O'Roak et al	TRUE	TRUE	FALSE	0.44
11773.mo	O'Roak et al	TRUE	TRUE	FALSE	0.39
11773.p1	O'Roak et al	TRUE	TRUE	FALSE	0.33
11773.s1	O'Roak et al	TRUE	TRUE	FALSE	0.34
11788.fa	O'Roak et al	TRUE	TRUE	FALSE	0.38
11788.mo	O'Roak et al	TRUE	TRUE	FALSE	0.37
11788.p1	O'Roak et al	TRUE	TRUE	FALSE	0.42
11788.s1	O'Roak et al	TRUE	TRUE	TRUE	0.37
11797.fa	Sanders et al	TRUE	TRUE	FALSE	0.28
11797.mo	Sanders et al	TRUE	TRUE	FALSE	0.27
11797.p1	Sanders et al	TRUE	TRUE	FALSE	0.26
11797.s1	Sanders et al	TRUE	TRUE	FALSE	0.26
11808.fa	Sanders et al	TRUE	TRUE	FALSE	0.35
11808.mo	Sanders et al	TRUE	TRUE	FALSE	0.37
11808.p1	Sanders et al	TRUE	TRUE	FALSE	0.34
11808.s1	Sanders et al	TRUE	TRUE	FALSE	0.41
11809.fa	Sanders et al	TRUE	TRUE	FALSE	0.37
11809.mo	Sanders et al	TRUE	TRUE	FALSE	0.31
11809.p1	Sanders et al	TRUE	TRUE	FALSE	0.29
11809.s1	Sanders et al	TRUE	TRUE	FALSE	0.34
11810.fa	Sanders et al	TRUE	TRUE	FALSE	0.31
11810.mo	Sanders et al	TRUE	TRUE	FALSE	0.30
11810.p1	Sanders et al	TRUE	TRUE	FALSE	0.24
11810.s1	Sanders et al	TRUE	TRUE	FALSE	0.44
11824.fa	Sanders et al	TRUE	TRUE	FALSE	0.30

11824.mo	Sanders et al	TRUE	TRUE	FALSE	0.31
11824.p1	Sanders et al	TRUE	TRUE	FALSE	0.30
11824.s1	Sanders et al	TRUE	FALSE	FALSE	0.39
11828.fa	Sanders et al	TRUE	TRUE	FALSE	0.27
11828.mo	Sanders et al	TRUE	TRUE	FALSE	0.27
11828.p1	Sanders et al	TRUE	TRUE	TRUE	0.26
11828.s1	Sanders et al	TRUE	TRUE	FALSE	0.27
11872.fa	O'Roak et al	TRUE	TRUE	FALSE	0.36
11872.mo	O'Roak et al	TRUE	TRUE	FALSE	0.35
11872.p1	O'Roak et al	TRUE	TRUE	TRUE	0.35
11872.s1	O'Roak et al	TRUE	TRUE	FALSE	0.33
11892.fa	Sanders et al	TRUE	TRUE	FALSE	0.30
11892.mo	Sanders et al	TRUE	TRUE	FALSE	0.31
11892.p1	Sanders et al	TRUE	TRUE	FALSE	0.34
11892.s1	Sanders et al	TRUE	TRUE	FALSE	0.27
11894.fa	lossifov et al	TRUE	TRUE	FALSE	0.44
11894.mo	lossifov et al	TRUE	TRUE	FALSE	0.54
11894.p1	lossifov et al	TRUE	TRUE	FALSE	0.54
11894.s1	lossifov et al	TRUE	TRUE	FALSE	0.55
11895.fa	O'Roak et al	TRUE	TRUE	FALSE	0.40
11895.mo	O'Roak et al	TRUE	TRUE	FALSE	0.41
11895.p1	O'Roak et al	TRUE	TRUE	FALSE	0.48
11895.s1	O'Roak et al	TRUE	TRUE	TRUE	0.34
11905.fa	Sanders et al	TRUE	TRUE	FALSE	0.32
11905.mo	Sanders et al	TRUE	TRUE	FALSE	0.33
11905.p1	Sanders et al	TRUE	TRUE	FALSE	0.38
11905.s1	Sanders et al	TRUE	TRUE	FALSE	0.39
11942.fa	O'Roak et al	TRUE	TRUE	FALSE	0.34
11942.mo	O'Roak et al	TRUE	TRUE	FALSE	0.40
11942.p1	O'Roak et al	TRUE	TRUE	TRUE	0.34
11942.s1	O'Roak et al	TRUE	TRUE	FALSE	0.37
11959.fa	O'Roak et al	TRUE	TRUE	FALSE	0.38
11959.mo	O'Roak et al	TRUE	TRUE	FALSE	0.39
11959.p1	O'Roak et al	TRUE	TRUE	FALSE	0.34
11959.s1	O'Roak et al	TRUE	TRUE	FALSE	0.37
11962.fa	Sanders et al	TRUE	TRUE	FALSE	0.28
11962.mo	Sanders et al	TRUE	TRUE	FALSE	0.27
11962.p1	Sanders et al	TRUE	TRUE	FALSE	0.26
11962.s1	Sanders et al	TRUE	TRUE	FALSE	0.27
11964.fa	O'Roak et al	TRUE	TRUE	FALSE	0.46
11964.mo	O'Roak et al	TRUE	TRUE	FALSE	0.41
11964.p1	O'Roak et al	TRUE	TRUE	FALSE	0.43
11964.s1	O'Roak et al	TRUE	TRUE	FALSE	0.37
12011.fa	O'Roak et al	TRUE	TRUE	FALSE	0.33
12011.mo	O'Roak et al	TRUE	TRUE	FALSE	0.40
12011.p1	O'Roak et al	TRUE	TRUE	FALSE	0.39
12011.s1	O'Roak et al	TRUE	TRUE	TRUE	0.38
12051.fa	lossifov et al	TRUE	TRUE	FALSE	0.50
12051.mo	lossifov et al	TRUE	TRUE	FALSE	0.56
12051.p1	lossifov et al	TRUE	TRUE	FALSE	0.55
12051.s1	lossifov et al	TRUE	TRUE	FALSE	0.58
12100.fa	Sanders et al	TRUE	TRUE	FALSE	0.32
12100.mo	Sanders et al	TRUE	TRUE	FALSE	0.30
12100.p1	Sanders et al	TRUE	TRUE	FALSE	0.32
12100.s1	Sanders et al	TRUE	TRUE	FALSE	0.40
12106.fa	O'Roak et al	TRUE	TRUE	FALSE	0.39
12106.mo	O'Roak et al	TRUE	TRUE	FALSE	0.33
12106.p1	O'Roak et al	TRUE	TRUE	FALSE	0.33
12106.s1	O'Roak et al	TRUE	TRUE	TRUE	0.38
12152.fa	O'Roak et al	TRUE	TRUE	FALSE	0.37
12152.mo	O'Roak et al	TRUE	TRUE	FALSE	0.37
12152.p1	O'Roak et al	TRUE	TRUE	FALSE	0.33
12152.s1	O'Roak et al	TRUE	TRUE	FALSE	0.37
12153.fa	O'Roak et al	TRUE	TRUE	FALSE	0.38
12153.mo	O'Roak et al	TRUE	TRUE	FALSE	0.42
12153.p1	O'Roak et al	TRUE	TRUE	FALSE	0.37
12153.s1	O'Roak et al	TRUE	TRUE	FALSE	0.31
12161.fa	O'Roak et al	TRUE	TRUE	FALSE	0.31
12161.mo	O'Roak et al	TRUE	TRUE	FALSE	0.35
12161.p1	O'Roak et al	TRUE	TRUE	FALSE	0.44
12161.s1	O'Roak et al	TRUE	TRUE	TRUE	0.37
12162.fa	Sanders et al	TRUE	TRUE	FALSE	0.29
12162.mo	Sanders et al	TRUE	TRUE	FALSE	0.33
12162.p1	Sanders et al	TRUE	TRUE	FALSE	0.30
12162.s1	Sanders et al	TRUE	TRUE	FALSE	0.26
12175.fa	Sanders et al	TRUE	TRUE	FALSE	0.31
12175.mo	Sanders et al	TRUE	TRUE	FALSE	0.28
12175.p1	Sanders et al	TRUE	TRUE	FALSE	0.30
12175.s1	Sanders et al	TRUE	TRUE	FALSE	0.27
12187.fa	Sanders et al	TRUE	TRUE	FALSE	0.32
12187.mo	Sanders et al	TRUE	TRUE	FALSE	0.34
12187.p1	Sanders et al	TRUE	TRUE	FALSE	0.32
12187.s1	Sanders et al	TRUE	TRUE	FALSE	0.28
12224.fa	Sanders et al	TRUE	TRUE	FALSE	0.29
12224.mo	Sanders et al	TRUE	TRUE	FALSE	0.29
12224.p1	Sanders et al	TRUE	TRUE	FALSE	0.28
12224.s1	Sanders et al	TRUE	TRUE	FALSE	0.27
12228.fa	Sanders et al	TRUE	TRUE	FALSE	0.28
12228.mo	Sanders et al	TRUE	TRUE	FALSE	0.30

12228.p1	Sanders et al	TRUE	TRUE	TRUE	0.25
12228.s1	Sanders et al	TRUE	TRUE	FALSE	0.29
12233.fa	O'Roak et al	TRUE	TRUE	FALSE	0.40
12233.mo	O'Roak et al	TRUE	TRUE	FALSE	0.39
12233.p1	O'Roak et al	TRUE	TRUE	FALSE	0.36
12233.s1	O'Roak et al	TRUE	TRUE	FALSE	0.34
12235.fa	Sanders et al	TRUE	TRUE	FALSE	0.24
12235.mo	Sanders et al	TRUE	TRUE	FALSE	0.28
12235.p1	Sanders et al	TRUE	TRUE	FALSE	0.26
12235.s1	Sanders et al	TRUE	TRUE	FALSE	0.22
12241.fa	Sanders et al	TRUE	TRUE	FALSE	0.32
12241.mo	Sanders et al	TRUE	TRUE	FALSE	0.31
12241.p1	Sanders et al	TRUE	TRUE	FALSE	0.30
12241.s1	Sanders et al	TRUE	TRUE	FALSE	0.29
12243.fa	lossifov et al	TRUE	TRUE	FALSE	0.55
12243.mo	lossifov et al	TRUE	TRUE	FALSE	0.60
12243.p1	lossifov et al	TRUE	TRUE	FALSE	0.57
12243.s1	lossifov et al	TRUE	TRUE	FALSE	0.59
12252.fa	lossifov et al	TRUE	TRUE	FALSE	0.50
12252.mo	lossifov et al	TRUE	TRUE	FALSE	0.50
12252.p1	lossifov et al	TRUE	TRUE	FALSE	0.58
12252.s1	lossifov et al	TRUE	TRUE	TRUE	0.52
12285.fa	O'Roak et al	TRUE	FALSE	FALSE	0.40
12285.mo	O'Roak et al	TRUE	FALSE	FALSE	0.36
12285.p1	O'Roak et al	TRUE	FALSE	FALSE	0.49
12285.s1	O'Roak et al	TRUE	FALSE	FALSE	0.30
12295.fa	Sanders et al	TRUE	TRUE	FALSE	0.37
12295.mo	Sanders et al	TRUE	TRUE	FALSE	0.35
12295.p1	Sanders et al	TRUE	TRUE	FALSE	0.31
12295.s1	Sanders et al	TRUE	TRUE	FALSE	0.35
12297.fa	Sanders et al	TRUE	TRUE	FALSE	0.25
12297.mo	Sanders et al	TRUE	TRUE	FALSE	0.25
12297.p1	Sanders et al	TRUE	TRUE	FALSE	0.21
12297.s1	Sanders et al	TRUE	TRUE	FALSE	0.23
12301.fa	lossifov et al	TRUE	TRUE	FALSE	0.55
12301.mo	lossifov et al	TRUE	TRUE	FALSE	0.55
12301.p1	lossifov et al	TRUE	TRUE	FALSE	0.48
12301.s1	lossifov et al	TRUE	TRUE	FALSE	0.48
12303.fa	Sanders et al	TRUE	TRUE	FALSE	0.35
12303.mo	Sanders et al	TRUE	TRUE	FALSE	0.23
12303.p1	Sanders et al	TRUE	TRUE	FALSE	0.24
12303.s1	Sanders et al	TRUE	TRUE	FALSE	0.26
12304.fa	O'Roak et al	TRUE	TRUE	FALSE	0.37
12304.mo	O'Roak et al	TRUE	TRUE	FALSE	0.42
12304.p1	O'Roak et al	TRUE	TRUE	TRUE	0.40
12304.s1	O'Roak et al	TRUE	TRUE	FALSE	0.35
12308.fa	Sanders et al	TRUE	TRUE	FALSE	0.29
12308.mo	Sanders et al	TRUE	TRUE	FALSE	0.30
12308.p1	Sanders et al	TRUE	TRUE	FALSE	0.31
12308.s1	Sanders et al	TRUE	TRUE	FALSE	0.40
12313.fa	lossifov et al	TRUE	TRUE	FALSE	0.48
12313.mo	lossifov et al	TRUE	TRUE	FALSE	0.50
12313.p1	lossifov et al	TRUE	TRUE	FALSE	0.53
12313.s1	lossifov et al	TRUE	TRUE	FALSE	0.58
12317.fa	Sanders et al	TRUE	TRUE	FALSE	0.23
12317.mo	Sanders et al	TRUE	TRUE	FALSE	0.27
12317.p1	Sanders et al	TRUE	TRUE	FALSE	0.23
12317.s1	Sanders et al	TRUE	TRUE	TRUE	0.25
12321.fa	lossifov et al	TRUE	TRUE	FALSE	0.49
12321.mo	lossifov et al	TRUE	TRUE	FALSE	0.50
12321.p1	lossifov et al	TRUE	TRUE	FALSE	0.51
12321.s1	lossifov et al	TRUE	TRUE	FALSE	0.54
12327.fa	Sanders et al	TRUE	TRUE	FALSE	0.36
12327.mo	Sanders et al	TRUE	TRUE	FALSE	0.33
12327.p1	Sanders et al	TRUE	TRUE	FALSE	0.31
12327.s1	Sanders et al	TRUE	TRUE	FALSE	0.50
12334.fa	lossifov et al	TRUE	TRUE	FALSE	0.41
12334.mo	lossifov et al	TRUE	TRUE	FALSE	0.48
12334.p1	lossifov et al	TRUE	TRUE	FALSE	0.40
12334.s1	lossifov et al	TRUE	TRUE	FALSE	0.41
12340.fa	Sanders et al	TRUE	TRUE	FALSE	0.38
12340.mo	Sanders et al	TRUE	TRUE	FALSE	0.38
12340.p1	Sanders et al	TRUE	TRUE	FALSE	0.43
12340.s1	Sanders et al	TRUE	TRUE	FALSE	0.34
12343.fa	Sanders et al	TRUE	TRUE	FALSE	0.35
12343.mo	Sanders et al	TRUE	TRUE	FALSE	0.39
12343.p1	Sanders et al	TRUE	TRUE	FALSE	0.40
12343.s1	Sanders et al	TRUE	TRUE	FALSE	0.37
12345.fa	Sanders et al	TRUE	TRUE	FALSE	0.24
12345.mo	Sanders et al	TRUE	TRUE	FALSE	0.23
12345.p1	Sanders et al	TRUE	TRUE	FALSE	0.26
12345.s1	Sanders et al	TRUE	TRUE	FALSE	0.24
12358.fa	O'Roak et al	TRUE	TRUE	FALSE	0.46
12358.mo	O'Roak et al	TRUE	TRUE	FALSE	0.39
12358.p1	O'Roak et al	TRUE	TRUE	FALSE	0.43
12358.s1	O'Roak et al	TRUE	TRUE	FALSE	0.38
12360.fa	lossifov et al	TRUE	TRUE	FALSE	0.60
12360.mo	lossifov et al	TRUE	TRUE	FALSE	0.54
12360.p1	lossifov et al	TRUE	TRUE	FALSE	0.57

12360.s1	lossifov et al	TRUE	TRUE	FALSE	0.61
12368.fa	Sanders et al	TRUE	TRUE	FALSE	0.31
12368.mo	Sanders et al	TRUE	TRUE	FALSE	0.29
12368.p1	Sanders et al	TRUE	TRUE	FALSE	0.33
12368.s1	Sanders et al	TRUE	TRUE	FALSE	0.28
12370.fa	Sanders et al	TRUE	TRUE	FALSE	0.27
12370.mo	Sanders et al	TRUE	TRUE	FALSE	0.29
12370.p1	Sanders et al	TRUE	TRUE	TRUE	0.30
12370.s1	Sanders et al	TRUE	TRUE	TRUE	0.27
12373.fa	O'Roak et al	TRUE	TRUE	FALSE	0.35
12373.mo	O'Roak et al	TRUE	TRUE	FALSE	0.33
12373.p1	O'Roak et al	TRUE	TRUE	FALSE	0.44
12373.s1	O'Roak et al	TRUE	TRUE	FALSE	0.36
12375.fa	Sanders et al	TRUE	TRUE	FALSE	0.28
12375.mo	Sanders et al	TRUE	TRUE	FALSE	0.31
12375.p1	Sanders et al	TRUE	TRUE	FALSE	0.30
12375.s1	Sanders et al	TRUE	TRUE	FALSE	0.28
12383.fa	Sanders et al	TRUE	TRUE	FALSE	0.30
12383.mo	Sanders et al	TRUE	TRUE	FALSE	0.30
12383.p1	Sanders et al	TRUE	TRUE	TRUE	0.32
12383.s1	Sanders et al	TRUE	TRUE	TRUE	0.29
12390.fa	O'Roak et al	TRUE	TRUE	FALSE	0.39
12390.mo	O'Roak et al	TRUE	TRUE	FALSE	0.30
12390.p1	O'Roak et al	TRUE	TRUE	FALSE	0.37
12390.s1	O'Roak et al	TRUE	TRUE	TRUE	0.39
12394.fa	lossifov et al	TRUE	FALSE	FALSE	0.50
12394.mo	lossifov et al	TRUE	FALSE	FALSE	0.50
12394.p1	lossifov et al	TRUE	FALSE	FALSE	0.45
12394.s1	lossifov et al	TRUE	FALSE	FALSE	0.47
12396.fa	lossifov et al	TRUE	TRUE	FALSE	0.60
12396.mo	lossifov et al	TRUE	TRUE	FALSE	0.50
12396.p1	lossifov et al	TRUE	TRUE	FALSE	0.45
12396.s1	lossifov et al	TRUE	TRUE	FALSE	0.47
12403.fa	Sanders et al	TRUE	TRUE	FALSE	0.40
12403.mo	Sanders et al	TRUE	TRUE	FALSE	0.36
12403.p1	Sanders et al	TRUE	TRUE	FALSE	0.37
12403.s1	Sanders et al	TRUE	TRUE	FALSE	0.35
12409.fa	lossifov et al	TRUE	TRUE	FALSE	0.44
12409.mo	lossifov et al	TRUE	TRUE	FALSE	0.45
12409.p1	lossifov et al	TRUE	TRUE	FALSE	0.42
12409.s1	lossifov et al	TRUE	TRUE	TRUE	0.48
12412.fa	lossifov et al	TRUE	FALSE	FALSE	0.49
12412.mo	lossifov et al	TRUE	FALSE	FALSE	0.50
12412.p1	lossifov et al	TRUE	FALSE	FALSE	0.46
12412.s1	lossifov et al	TRUE	FALSE	FALSE	0.46
12420.fa	lossifov et al	TRUE	TRUE	FALSE	0.50
12420.mo	lossifov et al	TRUE	TRUE	FALSE	0.56
12420.p1	lossifov et al	TRUE	TRUE	FALSE	0.44
12420.s1	lossifov et al	TRUE	TRUE	FALSE	0.47
12424.fa	lossifov et al	TRUE	TRUE	FALSE	0.46
12424.mo	lossifov et al	TRUE	TRUE	FALSE	0.38
12424.p1	lossifov et al	TRUE	TRUE	TRUE	0.46
12424.s1	lossifov et al	TRUE	TRUE	FALSE	0.48
12438.fa	lossifov et al	TRUE	TRUE	FALSE	0.49
12438.mo	lossifov et al	TRUE	TRUE	FALSE	0.50
12438.p1	lossifov et al	TRUE	TRUE	FALSE	0.51
12438.s1	lossifov et al	TRUE	TRUE	FALSE	0.53
12441.fa	lossifov et al	TRUE	TRUE	FALSE	0.38
12441.mo	lossifov et al	TRUE	TRUE	FALSE	0.41
12441.p1	lossifov et al	TRUE	TRUE	TRUE	0.39
12441.s1	lossifov et al	TRUE	TRUE	FALSE	0.44
12445.fa	lossifov et al	TRUE	TRUE	FALSE	0.45
12445.mo	lossifov et al	TRUE	TRUE	FALSE	0.53
12445.p1	lossifov et al	TRUE	TRUE	FALSE	0.41
12445.s1	lossifov et al	TRUE	TRUE	FALSE	0.42
12460.fa	lossifov et al	TRUE	TRUE	FALSE	0.53
12460.mo	lossifov et al	TRUE	TRUE	FALSE	0.42
12460.p1	lossifov et al	TRUE	TRUE	FALSE	0.46
12460.s1	lossifov et al	TRUE	TRUE	FALSE	0.50
12462.fa	lossifov et al	TRUE	TRUE	FALSE	0.61
12462.mo	lossifov et al	TRUE	TRUE	FALSE	0.55
12462.p1	lossifov et al	TRUE	TRUE	FALSE	0.50
12462.s1	lossifov et al	TRUE	TRUE	FALSE	0.59
12463.fa	lossifov et al	TRUE	TRUE	FALSE	0.51
12463.mo	lossifov et al	TRUE	TRUE	FALSE	0.46
12463.p1	lossifov et al	TRUE	TRUE	FALSE	0.59
12463.s1	lossifov et al	TRUE	TRUE	FALSE	0.58
12467.fa	lossifov et al	TRUE	FALSE	FALSE	0.48
12467.mo	lossifov et al	TRUE	FALSE	FALSE	0.50
12467.p1	lossifov et al	TRUE	FALSE	FALSE	0.47
12467.s1	lossifov et al	TRUE	FALSE	FALSE	0.42
12473.fa	lossifov et al	TRUE	FALSE	FALSE	0.33
12473.mo	lossifov et al	TRUE	FALSE	FALSE	0.45
12473.p1	lossifov et al	TRUE	FALSE	FALSE	0.30
12473.s1	lossifov et al	TRUE	FALSE	FALSE	0.30
12480.fa	lossifov et al	FALSE	TRUE	FALSE	0.46
12480.mo	lossifov et al	FALSE	TRUE	FALSE	0.53
12480.p1	lossifov et al	FALSE	TRUE	FALSE	0.43
12480.s1	lossifov et al	FALSE	FALSE	FALSE	0.41

12481.fa	lossifov et al	TRUE	TRUE	FALSE	0.43
12481.mo	lossifov et al	TRUE	TRUE	FALSE	0.49
12481.p1	lossifov et al	TRUE	TRUE	TRUE	0.48
12481.s1	lossifov et al	TRUE	TRUE	FALSE	0.45
12498.fa	lossifov et al	TRUE	TRUE	FALSE	0.38
12498.mo	lossifov et al	TRUE	TRUE	FALSE	0.40
12498.p1	lossifov et al	TRUE	TRUE	FALSE	0.44
12498.s1	lossifov et al	TRUE	TRUE	FALSE	0.40
12507.fa	Sanders et al	TRUE	TRUE	FALSE	0.28
12507.mo	Sanders et al	TRUE	TRUE	FALSE	0.27
12507.p1	Sanders et al	TRUE	TRUE	FALSE	0.29
12507.s1	Sanders et al	TRUE	TRUE	FALSE	0.34
12510.fa	lossifov et al	TRUE	TRUE	FALSE	0.42
12510.mo	lossifov et al	TRUE	TRUE	FALSE	0.49
12510.p1	lossifov et al	TRUE	TRUE	TRUE	0.49
12510.s1	lossifov et al	TRUE	TRUE	TRUE	0.53
12512.fa	Sanders et al	TRUE	TRUE	FALSE	0.34
12512.mo	Sanders et al	TRUE	TRUE	FALSE	0.32
12512.p1	Sanders et al	TRUE	TRUE	FALSE	0.40
12512.s1	Sanders et al	TRUE	TRUE	FALSE	0.35
12515.fa	lossifov et al	TRUE	TRUE	FALSE	0.50
12515.mo	lossifov et al	TRUE	TRUE	FALSE	0.54
12515.p1	lossifov et al	TRUE	TRUE	FALSE	0.49
12515.s1	lossifov et al	TRUE	TRUE	FALSE	0.51
12518.fa	lossifov et al	TRUE	TRUE	FALSE	0.47
12518.mo	lossifov et al	TRUE	TRUE	FALSE	0.55
12518.p1	lossifov et al	TRUE	TRUE	FALSE	0.46
12518.s1	lossifov et al	TRUE	TRUE	FALSE	0.48
12522.fa	Sanders et al	TRUE	FALSE	FALSE	0.19
12522.mo	Sanders et al	TRUE	FALSE	FALSE	0.19
12522.p1	Sanders et al	TRUE	FALSE	FALSE	0.24
12522.s1	Sanders et al	TRUE	FALSE	FALSE	0.24
12523.fa	lossifov et al	TRUE	TRUE	FALSE	0.54
12523.mo	lossifov et al	TRUE	TRUE	FALSE	0.56
12523.p1	lossifov et al	TRUE	TRUE	FALSE	0.44
12523.s1	lossifov et al	TRUE	TRUE	FALSE	0.58
12524.fa	Sanders et al	TRUE	TRUE	FALSE	0.29
12524.mo	Sanders et al	TRUE	TRUE	FALSE	0.29
12524.p1	Sanders et al	TRUE	TRUE	FALSE	0.30
12524.s1	Sanders et al	TRUE	TRUE	FALSE	0.28
12526.fa	lossifov et al	TRUE	FALSE	FALSE	0.59
12526.mo	lossifov et al	TRUE	FALSE	FALSE	0.60
12526.p1	lossifov et al	TRUE	FALSE	FALSE	0.58
12526.s1	lossifov et al	TRUE	FALSE	FALSE	0.59
12534.fa	Sanders et al	TRUE	TRUE	FALSE	0.36
12534.mo	Sanders et al	TRUE	TRUE	FALSE	0.38
12534.p1	Sanders et al	TRUE	TRUE	FALSE	0.37
12534.s1	Sanders et al	TRUE	TRUE	FALSE	0.34
12536.fa	Sanders et al	TRUE	TRUE	FALSE	0.36
12536.mo	Sanders et al	TRUE	TRUE	FALSE	0.37
12536.p1	Sanders et al	TRUE	TRUE	FALSE	0.35
12536.s1	Sanders et al	TRUE	TRUE	FALSE	0.45
12552.fa	Sanders et al	TRUE	TRUE	FALSE	0.36
12552.mo	Sanders et al	TRUE	TRUE	FALSE	0.38
12552.p1	Sanders et al	TRUE	TRUE	FALSE	0.35
12552.s1	Sanders et al	TRUE	TRUE	TRUE	0.40
12561.fa	Sanders et al	TRUE	TRUE	FALSE	0.27
12561.mo	Sanders et al	TRUE	TRUE	FALSE	0.21
12561.p1	Sanders et al	TRUE	TRUE	FALSE	0.21
12561.s1	Sanders et al	TRUE	TRUE	TRUE	0.22
12578.fa	O'Roak et al	TRUE	TRUE	FALSE	0.37
12578.mo	O'Roak et al	TRUE	TRUE	FALSE	0.38
12578.p1	O'Roak et al	TRUE	TRUE	FALSE	0.35
12578.s1	O'Roak et al	TRUE	TRUE	FALSE	0.36
12579.fa	lossifov et al	TRUE	TRUE	FALSE	0.58
12579.mo	lossifov et al	TRUE	TRUE	FALSE	0.52
12579.p1	lossifov et al	TRUE	TRUE	FALSE	0.55
12579.s1	lossifov et al	TRUE	TRUE	FALSE	0.60
12581.fa	O'Roak et al	TRUE	TRUE	FALSE	0.45
12581.mo	O'Roak et al	TRUE	TRUE	FALSE	0.46
12581.p1	O'Roak et al	TRUE	TRUE	FALSE	0.44
12581.s1	O'Roak et al	TRUE	TRUE	FALSE	0.38
12582.fa	lossifov et al	TRUE	TRUE	FALSE	0.60
12582.mo	lossifov et al	TRUE	TRUE	FALSE	0.59
12582.p1	lossifov et al	TRUE	TRUE	TRUE	0.55
12582.s1	lossifov et al	TRUE	TRUE	FALSE	0.53
12588.fa	lossifov et al	TRUE	TRUE	FALSE	0.54
12588.mo	lossifov et al	TRUE	TRUE	FALSE	0.53
12588.p1	lossifov et al	TRUE	TRUE	FALSE	0.57
12588.s1	lossifov et al	TRUE	TRUE	TRUE	0.57
12616.fa	Sanders et al	TRUE	TRUE	FALSE	0.23
12616.mo	Sanders et al	TRUE	TRUE	FALSE	0.21
12616.p1	Sanders et al	TRUE	TRUE	FALSE	0.22
12616.s1	Sanders et al	TRUE	TRUE	FALSE	0.22
12618.fa	lossifov et al	TRUE	TRUE	FALSE	0.48
12618.mo	lossifov et al	TRUE	TRUE	FALSE	0.54
12618.p1	lossifov et al	TRUE	TRUE	TRUE	0.55
12618.s1	lossifov et al	TRUE	TRUE	FALSE	0.55
12620.fa	lossifov et al	TRUE	FALSE	FALSE	0.54

12620.mo	lossifov et al	TRUE	FALSE	FALSE	0.54
12620.p1	lossifov et al	TRUE	FALSE	FALSE	0.51
12620.s1	lossifov et al	TRUE	FALSE	FALSE	0.49
12626.fa	lossifov et al	TRUE	TRUE	FALSE	0.50
12626.mo	lossifov et al	TRUE	TRUE	FALSE	0.52
12626.p1	lossifov et al	TRUE	TRUE	FALSE	0.52
12626.s1	lossifov et al	TRUE	TRUE	FALSE	0.52
12628.fa	lossifov et al	TRUE	TRUE	FALSE	0.52
12628.mo	lossifov et al	TRUE	TRUE	FALSE	0.49
12628.p1	lossifov et al	TRUE	TRUE	FALSE	0.51
12628.s1	lossifov et al	TRUE	TRUE	FALSE	0.55
12630.fa	O'Roak et al	TRUE	TRUE	FALSE	0.39
12630.mo	O'Roak et al	TRUE	TRUE	FALSE	0.39
12630.p1	O'Roak et al	TRUE	TRUE	FALSE	0.40
12630.s1	O'Roak et al	TRUE	TRUE	FALSE	0.37
12631.fa	lossifov et al	TRUE	TRUE	FALSE	0.50
12631.mo	lossifov et al	TRUE	TRUE	FALSE	0.52
12631.p1	lossifov et al	TRUE	TRUE	FALSE	0.53
12631.s1	lossifov et al	TRUE	TRUE	TRUE	0.52
12633.fa	lossifov et al	TRUE	FALSE	FALSE	0.47
12633.mo	lossifov et al	TRUE	FALSE	FALSE	0.48
12633.p1	lossifov et al	TRUE	FALSE	FALSE	0.45
12633.s1	lossifov et al	TRUE	FALSE	FALSE	0.47
12637.fa	lossifov et al	TRUE	TRUE	FALSE	0.40
12637.mo	lossifov et al	TRUE	TRUE	FALSE	0.41
12637.p1	lossifov et al	TRUE	TRUE	TRUE	0.40
12637.s1	lossifov et al	TRUE	TRUE	FALSE	0.43
12638.fa	lossifov et al	TRUE	TRUE	FALSE	0.45
12638.mo	lossifov et al	TRUE	TRUE	FALSE	0.45
12638.p1	lossifov et al	TRUE	TRUE	FALSE	0.43
12638.s1	lossifov et al	TRUE	TRUE	FALSE	0.47
12642.fa	lossifov et al	TRUE	TRUE	FALSE	0.56
12642.mo	lossifov et al	TRUE	TRUE	FALSE	0.54
12642.p1	lossifov et al	TRUE	TRUE	FALSE	0.57
12642.s1	lossifov et al	TRUE	TRUE	FALSE	0.53
12644.fa	lossifov et al	TRUE	TRUE	FALSE	0.54
12644.mo	lossifov et al	TRUE	TRUE	FALSE	0.51
12644.p1	lossifov et al	TRUE	TRUE	FALSE	0.50
12644.s1	lossifov et al	TRUE	TRUE	FALSE	0.49
12645.fa	lossifov et al	TRUE	TRUE	FALSE	0.41
12645.mo	lossifov et al	TRUE	TRUE	FALSE	0.43
12645.p1	lossifov et al	TRUE	TRUE	FALSE	0.41
12645.s1	lossifov et al	TRUE	TRUE	FALSE	0.42
12647.fa	Sanders et al	TRUE	TRUE	FALSE	0.31
12647.mo	Sanders et al	TRUE	TRUE	FALSE	0.26
12647.p1	Sanders et al	TRUE	TRUE	FALSE	0.40
12647.s1	Sanders et al	TRUE	TRUE	FALSE	0.39
12650.fa	Sanders et al	TRUE	TRUE	FALSE	0.24
12650.mo	Sanders et al	TRUE	TRUE	FALSE	0.21
12650.p1	Sanders et al	TRUE	TRUE	FALSE	0.20
12650.s1	Sanders et al	TRUE	TRUE	TRUE	0.26
12651.fa	Sanders et al	TRUE	TRUE	FALSE	0.31
12651.mo	Sanders et al	TRUE	TRUE	FALSE	0.30
12651.p1	Sanders et al	TRUE	TRUE	FALSE	0.37
12651.s1	Sanders et al	TRUE	TRUE	FALSE	0.28
12652.fa	lossifov et al	TRUE	TRUE	FALSE	0.50
12652.mo	lossifov et al	TRUE	TRUE	FALSE	0.55
12652.p1	lossifov et al	TRUE	TRUE	FALSE	0.54
12652.s1	lossifov et al	TRUE	TRUE	FALSE	0.54
12653.fa	lossifov et al	TRUE	TRUE	FALSE	0.53
12653.mo	lossifov et al	TRUE	TRUE	FALSE	0.49
12653.p1	lossifov et al	TRUE	TRUE	FALSE	0.49
12653.s1	lossifov et al	TRUE	TRUE	FALSE	0.53
12655.fa	lossifov et al	TRUE	TRUE	FALSE	0.50
12655.mo	lossifov et al	TRUE	TRUE	FALSE	0.58
12655.p1	lossifov et al	TRUE	TRUE	TRUE	0.50
12655.s1	lossifov et al	TRUE	TRUE	TRUE	0.46
12656.fa	Sanders et al	TRUE	TRUE	FALSE	0.37
12656.mo	Sanders et al	TRUE	TRUE	FALSE	0.35
12656.p1	Sanders et al	TRUE	TRUE	FALSE	0.34
12656.s1	Sanders et al	TRUE	TRUE	TRUE	0.33
12657.fa	Sanders et al	TRUE	TRUE	FALSE	0.31
12657.mo	Sanders et al	TRUE	TRUE	FALSE	0.35
12657.p1	Sanders et al	TRUE	TRUE	FALSE	0.34
12657.s1	Sanders et al	TRUE	TRUE	FALSE	0.40
12664.fa	lossifov et al	TRUE	TRUE	FALSE	0.50
12664.mo	lossifov et al	TRUE	TRUE	FALSE	0.52
12664.p1	lossifov et al	TRUE	TRUE	FALSE	0.46
12664.s1	lossifov et al	TRUE	TRUE	FALSE	0.56
12680.fa	Sanders et al	TRUE	FALSE	FALSE	0.33
12680.mo	Sanders et al	TRUE	FALSE	FALSE	0.32
12680.p1	Sanders et al	TRUE	FALSE	FALSE	0.31
12680.s1	Sanders et al	TRUE	FALSE	FALSE	0.44
12683.fa	lossifov et al	TRUE	FALSE	FALSE	0.52
12683.mo	lossifov et al	TRUE	FALSE	FALSE	0.54
12683.p1	lossifov et al	TRUE	FALSE	FALSE	0.58
12683.s1	lossifov et al	TRUE	FALSE	FALSE	0.53
12685.fa	Sanders et al	TRUE	TRUE	FALSE	0.32
12685.mo	Sanders et al	TRUE	TRUE	FALSE	0.29

12685.p1	Sanders et al	TRUE	TRUE	FALSE	0.29
12685.s1	Sanders et al	TRUE	TRUE	FALSE	0.32
12688.fa	lossifov et al	TRUE	FALSE	FALSE	0.60
12688.mo	lossifov et al	TRUE	FALSE	FALSE	0.55
12688.p1	lossifov et al	TRUE	FALSE	FALSE	0.57
12688.s1	lossifov et al	TRUE	FALSE	FALSE	0.56
12690.fa	Sanders et al	TRUE	TRUE	FALSE	0.21
12690.mo	Sanders et al	TRUE	TRUE	FALSE	0.23
12690.p1	Sanders et al	TRUE	TRUE	FALSE	0.21
12690.s1	Sanders et al	TRUE	TRUE	FALSE	0.22
12691.fa	lossifov et al	TRUE	TRUE	FALSE	0.46
12691.mo	lossifov et al	TRUE	TRUE	FALSE	0.48
12691.p1	lossifov et al	TRUE	TRUE	FALSE	0.48
12691.s1	lossifov et al	TRUE	TRUE	FALSE	0.49
12697.fa	lossifov et al	TRUE	FALSE	FALSE	0.50
12697.mo	lossifov et al	TRUE	FALSE	FALSE	0.50
12697.p1	lossifov et al	TRUE	FALSE	FALSE	0.46
12697.s1	lossifov et al	TRUE	FALSE	FALSE	0.52
12703.fa	O'Roak et al	TRUE	TRUE	FALSE	0.50
12703.mo	O'Roak et al	TRUE	TRUE	FALSE	0.40
12703.p1	O'Roak et al	TRUE	TRUE	FALSE	0.39
12703.s1	O'Roak et al	TRUE	TRUE	FALSE	0.31
12705.fa	lossifov et al	TRUE	FALSE	FALSE	0.57
12705.mo	lossifov et al	TRUE	FALSE	FALSE	0.55
12705.p1	lossifov et al	TRUE	FALSE	FALSE	0.57
12705.s1	lossifov et al	TRUE	FALSE	FALSE	0.55
12708.fa	lossifov et al	TRUE	TRUE	FALSE	0.48
12708.mo	lossifov et al	TRUE	TRUE	FALSE	0.50
12708.p1	lossifov et al	TRUE	TRUE	FALSE	0.47
12708.s1	lossifov et al	TRUE	TRUE	FALSE	0.53
12716.fa	lossifov et al	TRUE	FALSE	FALSE	0.48
12716.mo	lossifov et al	TRUE	FALSE	FALSE	0.43
12716.p1	lossifov et al	TRUE	FALSE	FALSE	0.43
12716.s1	lossifov et al	TRUE	FALSE	FALSE	0.43
12719.fa	lossifov et al	TRUE	FALSE	FALSE	0.45
12719.mo	lossifov et al	TRUE	FALSE	FALSE	0.46
12719.p1	lossifov et al	TRUE	FALSE	FALSE	0.45
12719.s1	lossifov et al	TRUE	FALSE	FALSE	0.46
12720.fa	lossifov et al	TRUE	FALSE	FALSE	0.45
12720.mo	lossifov et al	TRUE	FALSE	FALSE	0.46
12720.p1	lossifov et al	TRUE	FALSE	FALSE	0.44
12720.s1	lossifov et al	TRUE	FALSE	FALSE	0.49
12723.fa	lossifov et al	TRUE	TRUE	FALSE	0.46
12723.mo	lossifov et al	TRUE	TRUE	FALSE	0.43
12723.p1	lossifov et al	TRUE	TRUE	FALSE	0.43
12723.s1	lossifov et al	TRUE	TRUE	FALSE	0.44
12724.fa	lossifov et al	TRUE	FALSE	FALSE	0.45
12724.mo	lossifov et al	TRUE	FALSE	FALSE	0.42
12724.p1	lossifov et al	TRUE	FALSE	FALSE	0.44
12724.s1	lossifov et al	TRUE	FALSE	FALSE	0.41
12727.fa	lossifov et al	TRUE	FALSE	FALSE	0.42
12727.mo	lossifov et al	TRUE	FALSE	FALSE	0.41
12727.p1	lossifov et al	TRUE	FALSE	FALSE	0.43
12727.s1	lossifov et al	TRUE	FALSE	FALSE	0.43
12729.fa	Sanders et al	TRUE	TRUE	FALSE	0.34
12729.mo	Sanders et al	TRUE	TRUE	FALSE	0.37
12729.p1	Sanders et al	TRUE	TRUE	FALSE	0.34
12729.s1	Sanders et al	TRUE	TRUE	FALSE	0.27
12733.fa	lossifov et al	TRUE	FALSE	FALSE	0.45
12733.mo	lossifov et al	TRUE	FALSE	FALSE	0.50
12733.p1	lossifov et al	TRUE	FALSE	FALSE	0.45
12733.s1	lossifov et al	TRUE	FALSE	FALSE	0.47
12735.fa	lossifov et al	TRUE	TRUE	FALSE	0.41
12735.mo	lossifov et al	TRUE	TRUE	FALSE	0.42
12735.p1	lossifov et al	TRUE	TRUE	FALSE	0.41
12735.s1	lossifov et al	TRUE	TRUE	FALSE	0.41
12736.fa	Sanders et al	TRUE	TRUE	FALSE	0.31
12736.mo	Sanders et al	TRUE	TRUE	FALSE	0.31
12736.p1	Sanders et al	TRUE	TRUE	FALSE	0.24
12736.s1	Sanders et al	TRUE	TRUE	FALSE	0.41
12739.fa	lossifov et al	TRUE	TRUE	FALSE	0.55
12739.mo	lossifov et al	TRUE	TRUE	FALSE	0.58
12739.p1	lossifov et al	TRUE	TRUE	FALSE	0.57
12739.s1	lossifov et al	TRUE	TRUE	FALSE	0.59
12741.fa	O'Roak et al	TRUE	TRUE	FALSE	0.38
12741.mo	O'Roak et al	TRUE	TRUE	FALSE	0.38
12741.p1	O'Roak et al	TRUE	TRUE	FALSE	0.40
12741.s1	O'Roak et al	TRUE	TRUE	TRUE	0.40
12743.fa	lossifov et al	TRUE	TRUE	FALSE	0.47
12743.mo	lossifov et al	TRUE	TRUE	FALSE	0.49
12743.p1	lossifov et al	TRUE	TRUE	FALSE	0.47
12743.s1	lossifov et al	TRUE	TRUE	FALSE	0.49
12748.fa	lossifov et al	TRUE	TRUE	FALSE	0.51
12748.mo	lossifov et al	TRUE	TRUE	FALSE	0.53
12748.p1	lossifov et al	TRUE	TRUE	FALSE	0.55
12748.s1	lossifov et al	TRUE	TRUE	FALSE	0.57
12758.fa	lossifov et al	TRUE	TRUE	FALSE	0.55
12758.mo	lossifov et al	TRUE	TRUE	FALSE	0.54
12758.p1	lossifov et al	TRUE	TRUE	FALSE	0.54

12758.s1	lossifov et al	TRUE	TRUE	FALSE	0.55
12759.fa	lossifov et al	TRUE	TRUE	FALSE	0.51
12759.mo	lossifov et al	TRUE	TRUE	FALSE	0.53
12759.p1	lossifov et al	TRUE	TRUE	FALSE	0.50
12759.s1	lossifov et al	TRUE	TRUE	FALSE	0.54
12763.fa	Sanders et al	TRUE	FALSE	FALSE	0.30
12763.mo	Sanders et al	TRUE	FALSE	FALSE	0.32
12763.p1	Sanders et al	TRUE	FALSE	FALSE	0.35
12763.s1	Sanders et al	TRUE	FALSE	FALSE	0.37
12764.fa	lossifov et al	TRUE	TRUE	FALSE	0.52
12764.mo	lossifov et al	TRUE	TRUE	FALSE	0.52
12764.p1	lossifov et al	TRUE	TRUE	FALSE	0.51
12764.s1	lossifov et al	TRUE	TRUE	FALSE	0.55
12770.fa	lossifov et al	TRUE	FALSE	FALSE	0.52
12770.mo	lossifov et al	TRUE	FALSE	FALSE	0.56
12770.p1	lossifov et al	TRUE	FALSE	FALSE	0.54
12770.s1	lossifov et al	TRUE	FALSE	FALSE	0.57
12780.fa	Sanders et al	TRUE	TRUE	FALSE	0.24
12780.mo	Sanders et al	TRUE	TRUE	FALSE	0.22
12780.p1	Sanders et al	TRUE	TRUE	FALSE	0.23
12780.s1	Sanders et al	TRUE	TRUE	FALSE	0.21
12790.fa	Sanders et al	TRUE	TRUE	FALSE	0.38
12790.mo	Sanders et al	TRUE	TRUE	FALSE	0.32
12790.p1	Sanders et al	TRUE	TRUE	FALSE	0.31
12790.s1	Sanders et al	TRUE	TRUE	FALSE	0.31
12802.fa	Sanders et al	TRUE	FALSE	FALSE	0.37
12802.mo	Sanders et al	TRUE	FALSE	FALSE	0.35
12802.p1	Sanders et al	TRUE	FALSE	FALSE	0.35
12802.s1	Sanders et al	TRUE	FALSE	FALSE	0.33
12810.fa	O'Roak et al	TRUE	TRUE	FALSE	0.49
12810.mo	O'Roak et al	TRUE	TRUE	FALSE	0.39
12810.p1	O'Roak et al	TRUE	TRUE	TRUE	0.42
12810.s1	O'Roak et al	TRUE	TRUE	FALSE	0.33
12826.fa	lossifov et al	TRUE	FALSE	FALSE	0.40
12826.mo	lossifov et al	TRUE	FALSE	FALSE	0.44
12826.p1	lossifov et al	TRUE	FALSE	FALSE	0.41
12826.s1	lossifov et al	TRUE	FALSE	FALSE	0.41
12829.fa	lossifov et al	TRUE	TRUE	FALSE	0.40
12829.mo	lossifov et al	TRUE	TRUE	FALSE	0.42
12829.p1	lossifov et al	TRUE	TRUE	FALSE	0.39
12829.s1	lossifov et al	TRUE	TRUE	TRUE	0.41
12833.fa	lossifov et al	TRUE	FALSE	FALSE	0.46
12833.mo	lossifov et al	TRUE	FALSE	FALSE	0.45
12833.p1	lossifov et al	TRUE	FALSE	FALSE	0.49
12833.s1	lossifov et al	TRUE	FALSE	FALSE	0.45
12836.fa	lossifov et al	TRUE	TRUE	FALSE	0.42
12836.mo	lossifov et al	TRUE	TRUE	FALSE	0.43
12836.p1	lossifov et al	TRUE	TRUE	TRUE	0.43
12836.s1	lossifov et al	TRUE	TRUE	TRUE	0.43
12837.fa	lossifov et al	TRUE	TRUE	FALSE	0.45
12837.mo	lossifov et al	TRUE	TRUE	FALSE	0.45
12837.p1	lossifov et al	TRUE	TRUE	FALSE	0.45
12837.s1	lossifov et al	TRUE	TRUE	TRUE	0.45
12838.fa	lossifov et al	TRUE	TRUE	FALSE	0.44
12838.mo	lossifov et al	TRUE	TRUE	FALSE	0.44
12838.p1	lossifov et al	TRUE	TRUE	FALSE	0.46
12838.s1	lossifov et al	TRUE	TRUE	TRUE	0.44
12840.fa	lossifov et al	TRUE	FALSE	FALSE	0.39
12840.mo	lossifov et al	TRUE	FALSE	FALSE	0.38
12840.p1	lossifov et al	TRUE	FALSE	FALSE	0.39
12840.s1	lossifov et al	TRUE	FALSE	FALSE	0.42
12843.fa	lossifov et al	FALSE	TRUE	FALSE	0.42
12843.mo	lossifov et al	FALSE	TRUE	FALSE	0.40
12843.p1	lossifov et al	FALSE	TRUE	FALSE	0.38
12843.s1	lossifov et al	FALSE	FALSE	FALSE	0.41
12851.fa	lossifov et al	TRUE	TRUE	FALSE	0.41
12851.mo	lossifov et al	TRUE	TRUE	FALSE	0.40
12851.p1	lossifov et al	TRUE	TRUE	FALSE	0.40
12851.s1	lossifov et al	TRUE	TRUE	TRUE	0.41
12852.fa	lossifov et al	TRUE	TRUE	FALSE	0.39
12852.mo	lossifov et al	TRUE	TRUE	FALSE	0.42
12852.p1	lossifov et al	TRUE	TRUE	FALSE	0.40
12852.s1	lossifov et al	TRUE	TRUE	FALSE	0.40
12869.fa	Sanders et al	TRUE	TRUE	FALSE	0.35
12869.mo	Sanders et al	TRUE	TRUE	FALSE	0.43
12869.p1	Sanders et al	TRUE	TRUE	FALSE	0.40
12869.s1	Sanders et al	TRUE	FALSE	FALSE	0.32
12905.fa	O'Roak et al	TRUE	FALSE	FALSE	0.38
12905.mo	O'Roak et al	TRUE	FALSE	FALSE	0.33
12905.p1	O'Roak et al	TRUE	FALSE	FALSE	0.34
12905.s1	O'Roak et al	TRUE	FALSE	FALSE	0.39
12906.fa	Sanders et al	TRUE	FALSE	FALSE	0.35
12906.mo	Sanders et al	TRUE	FALSE	FALSE	0.29
12906.p1	Sanders et al	TRUE	FALSE	FALSE	0.28
12906.s1	Sanders et al	TRUE	FALSE	FALSE	0.29
12937.fa	lossifov et al	TRUE	TRUE	FALSE	0.48
12937.mo	lossifov et al	TRUE	TRUE	FALSE	0.51
12937.p1	lossifov et al	TRUE	TRUE	FALSE	0.50
12937.s1	lossifov et al	TRUE	TRUE	FALSE	0.42

12958.fa	Sanders et al	TRUE	FALSE	FALSE	0.37
12958.mo	Sanders et al	TRUE	FALSE	FALSE	0.36
12958.p1	Sanders et al	TRUE	FALSE	FALSE	0.44
12958.s1	Sanders et al	TRUE	FALSE	FALSE	0.36
12962.fa	lossifov et al	TRUE	TRUE	FALSE	0.43
12962.mo	lossifov et al	TRUE	TRUE	FALSE	0.51
12962.p1	lossifov et al	TRUE	TRUE	FALSE	0.50
12962.s1	lossifov et al	TRUE	TRUE	FALSE	0.47
12975.fa	lossifov et al	TRUE	TRUE	FALSE	0.41
12975.mo	lossifov et al	TRUE	TRUE	FALSE	0.56
12975.p1	lossifov et al	TRUE	TRUE	FALSE	0.50
12975.s1	lossifov et al	TRUE	TRUE	FALSE	0.52
12984.fa	Sanders et al	TRUE	TRUE	FALSE	0.32
12984.mo	Sanders et al	TRUE	TRUE	FALSE	0.34
12984.p1	Sanders et al	TRUE	TRUE	FALSE	0.31
12984.s1	Sanders et al	TRUE	TRUE	FALSE	0.40
12997.fa	lossifov et al	TRUE	TRUE	FALSE	0.43
12997.mo	lossifov et al	TRUE	TRUE	FALSE	0.50
12997.p1	lossifov et al	TRUE	TRUE	TRUE	0.48
12997.s1	lossifov et al	TRUE	TRUE	TRUE	0.57
13000.fa	Sanders et al	TRUE	FALSE	FALSE	0.28
13000.mo	Sanders et al	TRUE	FALSE	FALSE	0.28
13000.p1	Sanders et al	TRUE	FALSE	FALSE	0.29
13000.s1	Sanders et al	TRUE	FALSE	FALSE	0.30
13016.fa	lossifov et al	TRUE	TRUE	FALSE	0.47
13016.mo	lossifov et al	TRUE	TRUE	FALSE	0.54
13016.p1	lossifov et al	TRUE	TRUE	FALSE	0.52
13016.s1	lossifov et al	TRUE	TRUE	FALSE	0.60
13018.fa	lossifov et al	TRUE	TRUE	FALSE	0.58
13018.mo	lossifov et al	TRUE	TRUE	FALSE	0.61
13018.p1	lossifov et al	TRUE	TRUE	TRUE	0.48
13018.s1	lossifov et al	TRUE	TRUE	FALSE	0.44
13048.fa	O'Roak et al	TRUE	TRUE	FALSE	0.46
13048.mo	O'Roak et al	TRUE	TRUE	FALSE	0.40
13048.p1	O'Roak et al	TRUE	TRUE	FALSE	0.40
13048.s1	O'Roak et al	TRUE	TRUE	FALSE	0.33
13063.fa	Sanders et al	TRUE	TRUE	FALSE	0.28
13063.mo	Sanders et al	TRUE	TRUE	FALSE	0.28
13063.p1	Sanders et al	TRUE	TRUE	FALSE	0.29
13063.s1	Sanders et al	TRUE	TRUE	FALSE	0.27
13073.fa	Sanders et al	TRUE	TRUE	FALSE	0.28
13073.mo	Sanders et al	TRUE	TRUE	FALSE	0.30
13073.p1	Sanders et al	TRUE	TRUE	FALSE	0.32
13073.s1	Sanders et al	TRUE	TRUE	FALSE	0.33
13094.fa	lossifov et al	TRUE	FALSE	FALSE	0.46
13094.mo	lossifov et al	TRUE	FALSE	FALSE	0.48
13094.p1	lossifov et al	TRUE	FALSE	FALSE	0.47
13094.s1	lossifov et al	TRUE	FALSE	FALSE	0.45
13096.fa	lossifov et al	TRUE	TRUE	FALSE	0.51
13096.mo	lossifov et al	TRUE	TRUE	FALSE	0.52
13096.p1	lossifov et al	TRUE	TRUE	FALSE	0.44
13096.s1	lossifov et al	TRUE	FALSE	FALSE	0.49
13097.fa	lossifov et al	TRUE	TRUE	FALSE	0.43
13097.mo	lossifov et al	TRUE	TRUE	FALSE	0.41
13097.p1	lossifov et al	TRUE	TRUE	TRUE	0.50
13097.s1	lossifov et al	TRUE	TRUE	FALSE	0.51
13099.fa	lossifov et al	TRUE	FALSE	FALSE	0.54
13099.mo	lossifov et al	TRUE	FALSE	FALSE	0.58
13099.p1	lossifov et al	TRUE	FALSE	FALSE	0.50
13099.s1	lossifov et al	TRUE	FALSE	FALSE	0.44
13101.fa	lossifov et al	FALSE	FALSE	FALSE	0.56
13101.mo	lossifov et al	FALSE	FALSE	FALSE	0.56
13101.p1	lossifov et al	FALSE	FALSE	FALSE	0.36
13101.s1	lossifov et al	FALSE	FALSE	FALSE	0.61
13104.fa	lossifov et al	TRUE	TRUE	FALSE	0.51
13104.mo	lossifov et al	TRUE	TRUE	FALSE	0.48
13104.p1	lossifov et al	TRUE	TRUE	FALSE	0.40
13104.s1	lossifov et al	TRUE	TRUE	FALSE	0.54
13116.fa	O'Roak et al	TRUE	FALSE	FALSE	0.50
13116.mo	O'Roak et al	TRUE	FALSE	FALSE	0.39
13116.p1	O'Roak et al	TRUE	FALSE	FALSE	0.49
13116.s1	O'Roak et al	TRUE	FALSE	FALSE	0.30
13120.fa	lossifov et al	TRUE	TRUE	FALSE	0.49
13120.mo	lossifov et al	TRUE	TRUE	FALSE	0.47
13120.p1	lossifov et al	TRUE	TRUE	FALSE	0.53
13120.s1	lossifov et al	TRUE	TRUE	FALSE	0.56
13125.fa	lossifov et al	TRUE	FALSE	FALSE	0.54
13125.mo	lossifov et al	TRUE	FALSE	FALSE	0.44
13125.p1	lossifov et al	TRUE	FALSE	FALSE	0.52
13125.s1	lossifov et al	TRUE	FALSE	FALSE	0.40
13129.fa	lossifov et al	TRUE	FALSE	FALSE	0.58
13129.mo	lossifov et al	TRUE	FALSE	FALSE	0.52
13129.p1	lossifov et al	TRUE	FALSE	FALSE	0.39
13129.s1	lossifov et al	TRUE	FALSE	FALSE	0.48
13131.fa	lossifov et al	TRUE	FALSE	FALSE	0.48
13131.mo	lossifov et al	TRUE	FALSE	FALSE	0.46
13131.p1	lossifov et al	TRUE	FALSE	FALSE	0.46
13131.s1	lossifov et al	TRUE	FALSE	FALSE	0.47
13139.fa	lossifov et al	TRUE	FALSE	FALSE	0.45

13139.mo	lossifov et al	TRUE	FALSE	FALSE	0.44
13139.p1	lossifov et al	TRUE	FALSE	FALSE	0.43
13139.s1	lossifov et al	TRUE	FALSE	FALSE	0.46
13144.fa	lossifov et al	TRUE	TRUE	FALSE	0.47
13144.mo	lossifov et al	TRUE	TRUE	FALSE	0.46
13144.p1	lossifov et al	TRUE	TRUE	FALSE	0.45
13144.s1	lossifov et al	TRUE	TRUE	FALSE	0.47
13146.fa	lossifov et al	TRUE	TRUE	FALSE	0.46
13146.mo	lossifov et al	TRUE	TRUE	FALSE	0.47
13146.p1	lossifov et al	TRUE	TRUE	FALSE	0.46
13146.s1	lossifov et al	TRUE	FALSE	FALSE	0.47
13148.fa	lossifov et al	TRUE	FALSE	FALSE	0.44
13148.mo	lossifov et al	TRUE	FALSE	FALSE	0.45
13148.p1	lossifov et al	TRUE	FALSE	FALSE	0.46
13148.s1	lossifov et al	TRUE	FALSE	FALSE	0.42
13152.fa	lossifov et al	TRUE	TRUE	FALSE	0.51
13152.mo	lossifov et al	TRUE	TRUE	FALSE	0.50
13152.p1	lossifov et al	TRUE	TRUE	FALSE	0.52
13152.s1	lossifov et al	TRUE	FALSE	FALSE	0.51
13153.fa	lossifov et al	TRUE	TRUE	FALSE	0.44
13153.mo	lossifov et al	TRUE	TRUE	FALSE	0.43
13153.p1	lossifov et al	TRUE	TRUE	FALSE	0.42
13153.s1	lossifov et al	TRUE	TRUE	FALSE	0.42
13154.fa	Sanders et al	TRUE	FALSE	FALSE	0.23
13154.mo	Sanders et al	TRUE	FALSE	FALSE	0.23
13154.p1	Sanders et al	TRUE	FALSE	FALSE	0.22
13154.s1	Sanders et al	TRUE	FALSE	FALSE	0.23
13159.fa	lossifov et al	TRUE	TRUE	FALSE	0.43
13159.mo	lossifov et al	TRUE	TRUE	FALSE	0.44
13159.p1	lossifov et al	TRUE	TRUE	FALSE	0.43
13159.s1	lossifov et al	TRUE	TRUE	FALSE	0.43
13162.fa	lossifov et al	TRUE	TRUE	FALSE	0.44
13162.mo	lossifov et al	TRUE	TRUE	FALSE	0.42
13162.p1	lossifov et al	TRUE	TRUE	FALSE	0.42
13162.s1	lossifov et al	TRUE	TRUE	TRUE	0.43
13165.fa	lossifov et al	TRUE	FALSE	FALSE	0.43
13165.mo	lossifov et al	TRUE	FALSE	FALSE	0.44
13165.p1	lossifov et al	TRUE	FALSE	FALSE	0.43
13165.s1	lossifov et al	TRUE	FALSE	FALSE	0.46
13166.fa	lossifov et al	TRUE	TRUE	FALSE	0.43
13166.mo	lossifov et al	TRUE	TRUE	FALSE	0.42
13166.p1	lossifov et al	TRUE	TRUE	FALSE	0.42
13166.s1	lossifov et al	TRUE	TRUE	FALSE	0.42
13168.fa	lossifov et al	TRUE	TRUE	FALSE	0.48
13168.mo	lossifov et al	TRUE	TRUE	FALSE	0.46
13168.p1	lossifov et al	TRUE	TRUE	FALSE	0.48
13168.s1	lossifov et al	TRUE	TRUE	FALSE	0.49
13169.fa	O'Roak et al	TRUE	TRUE	FALSE	0.37
13169.mo	O'Roak et al	TRUE	TRUE	FALSE	0.40
13169.p1	O'Roak et al	TRUE	TRUE	FALSE	0.41
13169.s1	O'Roak et al	TRUE	TRUE	FALSE	0.35
13171.fa	Sanders et al	TRUE	TRUE	FALSE	0.37
13171.mo	Sanders et al	TRUE	TRUE	FALSE	0.37
13171.p1	Sanders et al	TRUE	TRUE	FALSE	0.45
13171.s1	Sanders et al	TRUE	TRUE	FALSE	0.35
13174.fa	lossifov et al	TRUE	TRUE	FALSE	0.50
13174.mo	lossifov et al	TRUE	TRUE	FALSE	0.50
13174.p1	lossifov et al	TRUE	TRUE	FALSE	0.50
13174.s1	lossifov et al	TRUE	TRUE	FALSE	0.51
13176.fa	lossifov et al	TRUE	FALSE	FALSE	0.45
13176.mo	lossifov et al	TRUE	FALSE	FALSE	0.43
13176.p1	lossifov et al	TRUE	FALSE	FALSE	0.45
13176.s1	lossifov et al	TRUE	FALSE	FALSE	0.45
13183.fa	lossifov et al	TRUE	TRUE	FALSE	0.54
13183.mo	lossifov et al	TRUE	TRUE	FALSE	0.53
13183.p1	lossifov et al	TRUE	TRUE	FALSE	0.51
13183.s1	lossifov et al	TRUE	TRUE	FALSE	0.51
13187.fa	lossifov et al	TRUE	TRUE	FALSE	0.47
13187.mo	lossifov et al	TRUE	TRUE	FALSE	0.47
13187.p1	lossifov et al	TRUE	TRUE	FALSE	0.48
13187.s1	lossifov et al	TRUE	TRUE	FALSE	0.51
13188.fa	O'Roak et al	TRUE	FALSE	FALSE	0.31
13188.mo	O'Roak et al	TRUE	FALSE	FALSE	0.41
13188.p1	O'Roak et al	TRUE	FALSE	FALSE	0.29
13188.s1	O'Roak et al	TRUE	FALSE	FALSE	0.39
13193.fa	lossifov et al	TRUE	TRUE	FALSE	0.42
13193.mo	lossifov et al	TRUE	TRUE	FALSE	0.49
13193.p1	lossifov et al	TRUE	TRUE	FALSE	0.53
13193.s1	lossifov et al	TRUE	TRUE	FALSE	0.47
13195.fa	Sanders et al	TRUE	TRUE	FALSE	0.40
13195.mo	Sanders et al	TRUE	TRUE	FALSE	0.36
13195.p1	Sanders et al	TRUE	TRUE	FALSE	0.33
13195.s1	Sanders et al	TRUE	TRUE	FALSE	0.42
13196.fa	lossifov et al	TRUE	TRUE	FALSE	0.46
13196.mo	lossifov et al	TRUE	TRUE	FALSE	0.47
13196.p1	lossifov et al	TRUE	TRUE	FALSE	0.43
13196.s1	lossifov et al	TRUE	TRUE	FALSE	0.52
13197.fa	lossifov et al	TRUE	FALSE	FALSE	0.47
13197.mo	lossifov et al	TRUE	FALSE	FALSE	0.47

13197.p1	lossifov et al	TRUE	FALSE	FALSE	0.49
13197.s1	lossifov et al	TRUE	FALSE	FALSE	0.49
13215.fa	lossifov et al	TRUE	FALSE	FALSE	0.57
13215.mo	lossifov et al	TRUE	FALSE	FALSE	0.53
13215.p1	lossifov et al	TRUE	FALSE	FALSE	0.46
13215.s1	lossifov et al	TRUE	FALSE	FALSE	0.57
13216.fa	lossifov et al	TRUE	TRUE	FALSE	0.27
13216.mo	lossifov et al	TRUE	TRUE	FALSE	0.34
13216.p1	lossifov et al	TRUE	TRUE	FALSE	0.27
13216.s1	lossifov et al	TRUE	TRUE	FALSE	0.27
13218.fa	lossifov et al	TRUE	TRUE	FALSE	0.49
13218.mo	lossifov et al	TRUE	TRUE	FALSE	0.47
13218.p1	lossifov et al	TRUE	TRUE	FALSE	0.41
13218.s1	lossifov et al	TRUE	FALSE	FALSE	0.52
13227.fa	lossifov et al	TRUE	FALSE	FALSE	0.38
13227.mo	lossifov et al	TRUE	FALSE	FALSE	0.42
13227.p1	lossifov et al	TRUE	FALSE	FALSE	0.37
13227.s1	lossifov et al	TRUE	FALSE	FALSE	0.40
13239.fa	lossifov et al	TRUE	FALSE	FALSE	0.36
13239.mo	lossifov et al	TRUE	FALSE	FALSE	0.40
13239.p1	lossifov et al	TRUE	FALSE	FALSE	0.34
13239.s1	lossifov et al	TRUE	FALSE	FALSE	0.39
13258.fa	lossifov et al	TRUE	FALSE	FALSE	0.42
13258.mo	lossifov et al	TRUE	FALSE	FALSE	0.42
13258.p1	lossifov et al	TRUE	FALSE	FALSE	0.42
13258.s1	lossifov et al	TRUE	FALSE	FALSE	0.40
13263.fa	lossifov et al	TRUE	FALSE	FALSE	0.47
13263.mo	lossifov et al	TRUE	FALSE	FALSE	0.48
13263.p1	lossifov et al	TRUE	FALSE	FALSE	0.46
13263.s1	lossifov et al	TRUE	FALSE	FALSE	0.48
13266.fa	lossifov et al	TRUE	TRUE	FALSE	0.46
13266.mo	lossifov et al	TRUE	TRUE	FALSE	0.52
13266.p1	lossifov et al	TRUE	TRUE	FALSE	0.48
13266.s1	lossifov et al	TRUE	TRUE	FALSE	0.47
13269.fa	lossifov et al	TRUE	FALSE	FALSE	0.50
13269.mo	lossifov et al	TRUE	FALSE	FALSE	0.45
13269.p1	lossifov et al	TRUE	FALSE	FALSE	0.49
13269.s1	lossifov et al	TRUE	FALSE	FALSE	0.50
13271.fa	Sanders et al	TRUE	FALSE	FALSE	0.30
13271.mo	Sanders et al	TRUE	FALSE	FALSE	0.31
13271.p1	Sanders et al	TRUE	FALSE	FALSE	0.31
13271.s1	Sanders et al	TRUE	FALSE	FALSE	0.27
13293.fa	lossifov et al	TRUE	FALSE	FALSE	0.47
13293.mo	lossifov et al	TRUE	FALSE	FALSE	0.51
13293.p1	lossifov et al	TRUE	FALSE	FALSE	0.50
13293.s1	lossifov et al	TRUE	FALSE	FALSE	0.56
13296.fa	lossifov et al	FALSE	TRUE	FALSE	0.47
13296.mo	lossifov et al	FALSE	TRUE	FALSE	0.47
13296.p1	lossifov et al	FALSE	TRUE	FALSE	0.46
13296.s1	lossifov et al	FALSE	TRUE	TRUE	0.47
13307.fa	lossifov et al	TRUE	FALSE	FALSE	0.33
13307.mo	lossifov et al	TRUE	FALSE	FALSE	0.33
13307.p1	lossifov et al	TRUE	FALSE	FALSE	0.34
13307.s1	lossifov et al	TRUE	FALSE	FALSE	0.34
13309.fa	lossifov et al	TRUE	FALSE	FALSE	0.44
13309.mo	lossifov et al	TRUE	FALSE	FALSE	0.44
13309.p1	lossifov et al	TRUE	FALSE	FALSE	0.45
13309.s1	lossifov et al	TRUE	FALSE	FALSE	0.43
13312.fa	lossifov et al	TRUE	FALSE	FALSE	0.53
13312.mo	lossifov et al	TRUE	FALSE	FALSE	0.50
13312.p1	lossifov et al	TRUE	FALSE	FALSE	0.45
13312.s1	lossifov et al	TRUE	FALSE	FALSE	0.50
13315.fa	lossifov et al	TRUE	FALSE	FALSE	0.33
13315.mo	lossifov et al	TRUE	FALSE	FALSE	0.31
13315.p1	lossifov et al	TRUE	FALSE	FALSE	0.35
13315.s1	lossifov et al	TRUE	FALSE	FALSE	0.32
13322.fa	Sanders et al	TRUE	TRUE	FALSE	0.27
13322.mo	Sanders et al	TRUE	TRUE	FALSE	0.30
13322.p1	Sanders et al	TRUE	TRUE	FALSE	0.33
13322.s1	Sanders et al	TRUE	TRUE	FALSE	0.26
13327.fa	lossifov et al	TRUE	TRUE	FALSE	0.30
13327.mo	lossifov et al	TRUE	TRUE	FALSE	0.34
13327.p1	lossifov et al	TRUE	TRUE	FALSE	0.50
13327.s1	lossifov et al	TRUE	TRUE	TRUE	0.45
13328.fa	lossifov et al	TRUE	FALSE	FALSE	0.37
13328.mo	lossifov et al	TRUE	FALSE	FALSE	0.38
13328.p1	lossifov et al	TRUE	FALSE	FALSE	0.40
13328.s1	lossifov et al	TRUE	FALSE	FALSE	0.49
13330.fa	lossifov et al	TRUE	FALSE	FALSE	0.50
13330.mo	lossifov et al	TRUE	FALSE	FALSE	0.39
13330.p1	lossifov et al	TRUE	FALSE	FALSE	0.55
13330.s1	lossifov et al	TRUE	FALSE	FALSE	0.56
13335.fa	O'Roak et al	TRUE	FALSE	FALSE	0.48
13335.mo	O'Roak et al	TRUE	FALSE	FALSE	0.43
13335.p1	O'Roak et al	TRUE	FALSE	FALSE	0.57
13335.s1	O'Roak et al	TRUE	FALSE	FALSE	0.47
13338.fa	lossifov et al	TRUE	FALSE	FALSE	0.41
13338.mo	lossifov et al	TRUE	FALSE	FALSE	0.45
13338.p1	lossifov et al	TRUE	FALSE	FALSE	0.54

13338.s1	lossifov et al	TRUE	FALSE	FALSE	0.51
13346.fa	O'Roak et al	TRUE	FALSE	FALSE	0.38
13346.mo	O'Roak et al	TRUE	FALSE	FALSE	0.37
13346.p1	O'Roak et al	TRUE	FALSE	FALSE	0.38
13346.s1	O'Roak et al	TRUE	FALSE	FALSE	0.34
13349.fa	lossifov et al	TRUE	FALSE	FALSE	0.50
13349.mo	lossifov et al	TRUE	FALSE	FALSE	0.40
13349.p1	lossifov et al	TRUE	FALSE	FALSE	0.38
13349.s1	lossifov et al	TRUE	FALSE	FALSE	0.36
13355.fa	Sanders et al	TRUE	FALSE	FALSE	0.28
13355.mo	Sanders et al	TRUE	FALSE	FALSE	0.28
13355.p1	Sanders et al	TRUE	FALSE	FALSE	0.29
13355.s1	Sanders et al	TRUE	FALSE	FALSE	0.26
13366.fa	lossifov et al	TRUE	FALSE	FALSE	0.48
13366.mo	lossifov et al	TRUE	FALSE	FALSE	0.43
13366.p1	lossifov et al	TRUE	FALSE	FALSE	0.37
13366.s1	lossifov et al	TRUE	FALSE	FALSE	0.52
13374.fa	Sanders et al	TRUE	FALSE	FALSE	0.33
13374.mo	Sanders et al	TRUE	FALSE	FALSE	0.38
13374.p1	Sanders et al	TRUE	FALSE	FALSE	0.38
13374.s1	Sanders et al	TRUE	FALSE	FALSE	0.34
13385.fa	Sanders et al	TRUE	FALSE	FALSE	0.25
13385.mo	Sanders et al	TRUE	FALSE	FALSE	0.24
13385.p1	Sanders et al	TRUE	FALSE	FALSE	0.22
13385.s1	Sanders et al	TRUE	FALSE	FALSE	0.24
13387.fa	lossifov et al	TRUE	FALSE	FALSE	0.45
13387.mo	lossifov et al	TRUE	FALSE	FALSE	0.46
13387.p1	lossifov et al	TRUE	FALSE	FALSE	0.54
13387.s1	lossifov et al	TRUE	FALSE	FALSE	0.51
13393.fa	Sanders et al	TRUE	FALSE	FALSE	0.28
13393.mo	Sanders et al	TRUE	FALSE	FALSE	0.31
13393.p1	Sanders et al	TRUE	FALSE	FALSE	0.30
13393.s1	Sanders et al	TRUE	FALSE	FALSE	0.27
13396.fa	lossifov et al	TRUE	FALSE	FALSE	0.46
13396.mo	lossifov et al	TRUE	FALSE	FALSE	0.45
13396.p1	lossifov et al	TRUE	FALSE	FALSE	0.46
13396.s1	lossifov et al	TRUE	FALSE	FALSE	0.45
13398.fa	lossifov et al	TRUE	FALSE	FALSE	0.42
13398.mo	lossifov et al	TRUE	FALSE	FALSE	0.44
13398.p1	lossifov et al	TRUE	FALSE	FALSE	0.44
13398.s1	lossifov et al	TRUE	FALSE	FALSE	0.43
13412.fa	lossifov et al	TRUE	FALSE	FALSE	0.48
13412.mo	lossifov et al	TRUE	FALSE	FALSE	0.49
13412.p1	lossifov et al	TRUE	FALSE	FALSE	0.47
13412.s1	lossifov et al	TRUE	FALSE	FALSE	0.44
13418.fa	lossifov et al	TRUE	FALSE	FALSE	0.45
13418.mo	lossifov et al	TRUE	FALSE	FALSE	0.43
13418.p1	lossifov et al	TRUE	FALSE	FALSE	0.45
13418.s1	lossifov et al	TRUE	FALSE	FALSE	0.43
13424.fa	lossifov et al	TRUE	FALSE	FALSE	0.44
13424.mo	lossifov et al	TRUE	FALSE	FALSE	0.42
13424.p1	lossifov et al	TRUE	FALSE	FALSE	0.41
13424.s1	lossifov et al	TRUE	FALSE	FALSE	0.44
13439.fa	lossifov et al	TRUE	FALSE	FALSE	0.43
13439.mo	lossifov et al	TRUE	FALSE	FALSE	0.45
13439.p1	lossifov et al	TRUE	FALSE	FALSE	0.45
13439.s1	lossifov et al	TRUE	FALSE	FALSE	0.44
13443.fa	lossifov et al	TRUE	FALSE	FALSE	0.44
13443.mo	lossifov et al	TRUE	FALSE	FALSE	0.44
13443.p1	lossifov et al	TRUE	FALSE	FALSE	0.45
13443.s1	lossifov et al	TRUE	FALSE	FALSE	0.45
13444.fa	lossifov et al	TRUE	FALSE	FALSE	0.34
13444.mo	lossifov et al	TRUE	FALSE	FALSE	0.36
13444.p1	lossifov et al	TRUE	FALSE	FALSE	0.35
13444.s1	lossifov et al	TRUE	FALSE	FALSE	0.36
13447.fa	O'Roak et al	TRUE	FALSE	FALSE	0.38
13447.mo	O'Roak et al	TRUE	FALSE	FALSE	0.34
13447.p1	O'Roak et al	TRUE	FALSE	FALSE	0.38
13447.s1	O'Roak et al	TRUE	FALSE	FALSE	0.32
13462.fa	lossifov et al	TRUE	FALSE	FALSE	0.34
13462.mo	lossifov et al	TRUE	FALSE	FALSE	0.36
13462.p1	lossifov et al	TRUE	FALSE	FALSE	0.35
13462.s1	lossifov et al	TRUE	FALSE	FALSE	0.36
13465.fa	lossifov et al	TRUE	FALSE	FALSE	0.47
13465.mo	lossifov et al	TRUE	FALSE	FALSE	0.44
13465.p1	lossifov et al	TRUE	FALSE	FALSE	0.44
13465.s1	lossifov et al	TRUE	FALSE	FALSE	0.42
13486.fa	lossifov et al	TRUE	FALSE	FALSE	0.47
13486.mo	lossifov et al	TRUE	FALSE	FALSE	0.51
13486.p1	lossifov et al	TRUE	FALSE	FALSE	0.50
13486.s1	lossifov et al	TRUE	FALSE	FALSE	0.53
13487.fa	lossifov et al	TRUE	FALSE	FALSE	0.38
13487.mo	lossifov et al	TRUE	FALSE	FALSE	0.40
13487.p1	lossifov et al	TRUE	FALSE	FALSE	0.39
13487.s1	lossifov et al	TRUE	FALSE	FALSE	0.40
13493.fa	lossifov et al	TRUE	FALSE	FALSE	0.42
13493.mo	lossifov et al	TRUE	FALSE	FALSE	0.39
13493.p1	lossifov et al	TRUE	FALSE	FALSE	0.38
13493.s1	lossifov et al	TRUE	FALSE	FALSE	0.38

13496.fa	lossifov et al	TRUE	FALSE	FALSE	0.41
13496.mo	lossifov et al	TRUE	FALSE	FALSE	0.40
13496.p1	lossifov et al	TRUE	FALSE	FALSE	0.41
13496.s1	lossifov et al	TRUE	FALSE	FALSE	0.43
13502.fa	lossifov et al	TRUE	FALSE	FALSE	0.34
13502.mo	lossifov et al	TRUE	FALSE	FALSE	0.33
13502.p1	lossifov et al	TRUE	FALSE	FALSE	0.34
13502.s1	lossifov et al	TRUE	FALSE	FALSE	0.33
13504.fa	lossifov et al	TRUE	FALSE	FALSE	0.38
13504.mo	lossifov et al	TRUE	FALSE	FALSE	0.37
13504.p1	lossifov et al	TRUE	FALSE	FALSE	0.35
13504.s1	lossifov et al	TRUE	FALSE	FALSE	0.37
13505.fa	lossifov et al	TRUE	FALSE	FALSE	0.35
13505.mo	lossifov et al	TRUE	FALSE	FALSE	0.33
13505.p1	lossifov et al	TRUE	FALSE	FALSE	0.33
13505.s1	lossifov et al	TRUE	FALSE	FALSE	0.34
13507.fa	lossifov et al	TRUE	FALSE	FALSE	0.35
13507.mo	lossifov et al	TRUE	FALSE	FALSE	0.36
13507.p1	lossifov et al	TRUE	FALSE	FALSE	0.35
13507.s1	lossifov et al	TRUE	FALSE	FALSE	0.36
13508.fa	lossifov et al	TRUE	FALSE	FALSE	0.34
13508.mo	lossifov et al	TRUE	FALSE	FALSE	0.39
13508.p1	lossifov et al	TRUE	FALSE	FALSE	0.38
13508.s1	lossifov et al	TRUE	FALSE	FALSE	0.34
13509.fa	Sanders et al	TRUE	FALSE	FALSE	0.34
13509.mo	Sanders et al	TRUE	FALSE	FALSE	0.34
13509.p1	Sanders et al	TRUE	FALSE	FALSE	0.33
13509.s1	Sanders et al	TRUE	FALSE	FALSE	0.35
13512.fa	lossifov et al	TRUE	FALSE	FALSE	0.38
13512.mo	lossifov et al	TRUE	FALSE	FALSE	0.39
13512.p1	lossifov et al	TRUE	FALSE	FALSE	0.39
13512.s1	lossifov et al	TRUE	FALSE	FALSE	0.39
13513.fa	lossifov et al	TRUE	FALSE	FALSE	0.45
13513.mo	lossifov et al	TRUE	FALSE	FALSE	0.44
13513.p1	lossifov et al	TRUE	FALSE	FALSE	0.46
13513.s1	lossifov et al	TRUE	FALSE	FALSE	0.43
13533.fa	O'Roak et al	TRUE	FALSE	FALSE	0.48
13533.mo	O'Roak et al	TRUE	FALSE	FALSE	0.42
13533.p1	O'Roak et al	TRUE	FALSE	FALSE	0.42
13533.s1	O'Roak et al	TRUE	FALSE	FALSE	0.34
13543.fa	Sanders et al	TRUE	FALSE	FALSE	0.31
13543.mo	Sanders et al	TRUE	FALSE	FALSE	0.27
13543.p1	Sanders et al	TRUE	FALSE	FALSE	0.29
13543.s1	Sanders et al	TRUE	FALSE	FALSE	0.30
13589.fa	lossifov et al	TRUE	FALSE	FALSE	0.40
13589.mo	lossifov et al	TRUE	FALSE	FALSE	0.37
13589.p1	lossifov et al	TRUE	FALSE	FALSE	0.42
13589.s1	lossifov et al	TRUE	FALSE	FALSE	0.40
13590.fa	lossifov et al	TRUE	FALSE	FALSE	0.44
13590.mo	lossifov et al	TRUE	FALSE	FALSE	0.43
13590.p1	lossifov et al	TRUE	FALSE	FALSE	0.41
13590.s1	lossifov et al	TRUE	FALSE	FALSE	0.41
13593.fa	O'Roak et al	TRUE	FALSE	FALSE	0.57
13593.mo	O'Roak et al	TRUE	FALSE	FALSE	0.49
13593.p1	O'Roak et al	TRUE	FALSE	FALSE	0.44
13593.s1	O'Roak et al	TRUE	FALSE	FALSE	0.31
13599.fa	lossifov et al	TRUE	FALSE	FALSE	0.47
13599.mo	lossifov et al	TRUE	FALSE	FALSE	0.48
13599.p1	lossifov et al	TRUE	FALSE	FALSE	0.40
13599.s1	lossifov et al	TRUE	FALSE	FALSE	0.46
13601.fa	lossifov et al	TRUE	FALSE	FALSE	0.42
13601.mo	lossifov et al	TRUE	FALSE	FALSE	0.40
13601.p1	lossifov et al	TRUE	FALSE	FALSE	0.44
13601.s1	lossifov et al	TRUE	FALSE	FALSE	0.43
13606.fa	O'Roak et al	TRUE	FALSE	FALSE	0.33
13606.mo	O'Roak et al	TRUE	FALSE	FALSE	0.37
13606.p1	O'Roak et al	TRUE	FALSE	FALSE	0.36
13606.s1	O'Roak et al	TRUE	FALSE	FALSE	0.32
13608.fa	Sanders et al	TRUE	FALSE	FALSE	0.25
13608.mo	Sanders et al	TRUE	FALSE	FALSE	0.24
13608.p1	Sanders et al	TRUE	FALSE	FALSE	0.24
13608.s1	Sanders et al	TRUE	FALSE	FALSE	0.26
13618.fa	Sanders et al	TRUE	FALSE	FALSE	0.27
13618.mo	Sanders et al	TRUE	FALSE	FALSE	0.28
13618.p1	Sanders et al	TRUE	FALSE	FALSE	0.33
13618.s1	Sanders et al	TRUE	FALSE	FALSE	0.28
13621.fa	Sanders et al	TRUE	FALSE	FALSE	0.33
13621.mo	Sanders et al	TRUE	FALSE	FALSE	0.41
13621.p1	Sanders et al	TRUE	FALSE	FALSE	0.32
13621.s1	Sanders et al	TRUE	FALSE	FALSE	0.30
13625.fa	Sanders et al	TRUE	FALSE	FALSE	0.29
13625.mo	Sanders et al	TRUE	FALSE	FALSE	0.29
13625.p1	Sanders et al	TRUE	FALSE	FALSE	0.29
13625.s1	Sanders et al	TRUE	FALSE	FALSE	0.29
13629.fa	O'Roak et al	TRUE	FALSE	FALSE	0.33
13629.mo	O'Roak et al	TRUE	FALSE	FALSE	0.37
13629.p1	O'Roak et al	TRUE	FALSE	FALSE	0.38
13629.s1	O'Roak et al	TRUE	FALSE	FALSE	0.32
13660.fa	Sanders et al	TRUE	FALSE	FALSE	0.29

13660.mo	Sanders et al	TRUE	FALSE	FALSE	0.29
13660.p1	Sanders et al	TRUE	FALSE	FALSE	0.32
13660.s1	Sanders et al	TRUE	FALSE	FALSE	0.33
13684.fa	lossifov et al	TRUE	FALSE	FALSE	0.44
13684.mo	lossifov et al	TRUE	FALSE	FALSE	0.47
13684.p1	lossifov et al	TRUE	FALSE	FALSE	0.45
13684.s1	lossifov et al	TRUE	FALSE	FALSE	0.45
13689.fa	lossifov et al	TRUE	FALSE	FALSE	0.43
13689.mo	lossifov et al	TRUE	FALSE	FALSE	0.44
13689.p1	lossifov et al	TRUE	FALSE	FALSE	0.42
13689.s1	lossifov et al	TRUE	FALSE	FALSE	0.44
13695.fa	lossifov et al	FALSE	FALSE	FALSE	0.48
13695.mo	lossifov et al	FALSE	FALSE	FALSE	0.46
13695.p1	lossifov et al	FALSE	FALSE	FALSE	0.46
13695.s1	lossifov et al	FALSE	FALSE	FALSE	0.45
13698.fa	lossifov et al	TRUE	FALSE	FALSE	0.40
13698.mo	lossifov et al	TRUE	FALSE	FALSE	0.45
13698.p1	lossifov et al	TRUE	FALSE	FALSE	0.49
13698.s1	lossifov et al	TRUE	FALSE	FALSE	0.42
13726.fa	O'Roak et al	TRUE	FALSE	FALSE	0.53
13726.mo	O'Roak et al	TRUE	FALSE	FALSE	0.50
13726.p1	O'Roak et al	TRUE	FALSE	FALSE	0.41
13726.s1	O'Roak et al	TRUE	FALSE	FALSE	0.32
13730.fa	Sanders et al	TRUE	FALSE	FALSE	0.28
13730.mo	Sanders et al	TRUE	FALSE	FALSE	0.28
13730.p1	Sanders et al	TRUE	FALSE	FALSE	0.33
13730.s1	Sanders et al	TRUE	FALSE	FALSE	0.29
13739.fa	Sanders et al	TRUE	FALSE	FALSE	0.30
13739.mo	Sanders et al	TRUE	FALSE	FALSE	0.29
13739.p1	Sanders et al	TRUE	FALSE	FALSE	0.28
13739.s1	Sanders et al	TRUE	FALSE	FALSE	0.29
13752.fa	Sanders et al	TRUE	FALSE	FALSE	0.24
13752.mo	Sanders et al	TRUE	FALSE	FALSE	0.24
13752.p1	Sanders et al	TRUE	FALSE	FALSE	0.24
13752.s1	Sanders et al	TRUE	FALSE	FALSE	0.24
13774.fa	Sanders et al	TRUE	FALSE	FALSE	0.29
13774.mo	Sanders et al	TRUE	FALSE	FALSE	0.32
13774.p1	Sanders et al	TRUE	FALSE	FALSE	0.27
13774.s1	Sanders et al	TRUE	FALSE	FALSE	0.30
13793.fa	O'Roak et al	TRUE	FALSE	FALSE	0.49
13793.mo	O'Roak et al	TRUE	FALSE	FALSE	0.52
13793.p1	O'Roak et al	TRUE	FALSE	FALSE	0.39
13793.s1	O'Roak et al	TRUE	FALSE	FALSE	0.36
13795.fa	Sanders et al	TRUE	FALSE	FALSE	0.28
13795.mo	Sanders et al	TRUE	FALSE	FALSE	0.29
13795.p1	Sanders et al	TRUE	FALSE	FALSE	0.27
13795.s1	Sanders et al	TRUE	FALSE	FALSE	0.30
13798.fa	O'Roak et al	TRUE	FALSE	FALSE	0.41
13798.mo	O'Roak et al	TRUE	FALSE	FALSE	0.37
13798.p1	O'Roak et al	TRUE	FALSE	FALSE	0.36
13798.s1	O'Roak et al	TRUE	FALSE	FALSE	0.39
13808.fa	Sanders et al	TRUE	FALSE	FALSE	0.21
13808.mo	Sanders et al	TRUE	FALSE	FALSE	0.21
13808.p1	Sanders et al	TRUE	FALSE	FALSE	0.20
13808.s1	Sanders et al	TRUE	FALSE	FALSE	0.25
13809.fa	Sanders et al	TRUE	FALSE	FALSE	0.27
13809.mo	Sanders et al	TRUE	FALSE	FALSE	0.27
13809.p1	Sanders et al	TRUE	FALSE	FALSE	0.27
13809.s1	Sanders et al	TRUE	FALSE	FALSE	0.27
13815.fa	O'Roak et al	TRUE	FALSE	FALSE	0.54
13815.mo	O'Roak et al	TRUE	FALSE	FALSE	0.40
13815.p1	O'Roak et al	TRUE	FALSE	FALSE	0.37
13815.s1	O'Roak et al	TRUE	FALSE	FALSE	0.48
13821.fa	Sanders et al	TRUE	FALSE	FALSE	0.28
13821.mo	Sanders et al	TRUE	FALSE	FALSE	0.23
13821.p1	Sanders et al	TRUE	FALSE	FALSE	0.26
13821.s1	Sanders et al	TRUE	FALSE	FALSE	0.26
13825.fa	Sanders et al	TRUE	FALSE	FALSE	0.29
13825.mo	Sanders et al	TRUE	FALSE	FALSE	0.29
13825.p1	Sanders et al	TRUE	FALSE	FALSE	0.31
13825.s1	Sanders et al	TRUE	FALSE	FALSE	0.31
13832.fa	Sanders et al	TRUE	FALSE	FALSE	0.28
13832.mo	Sanders et al	TRUE	FALSE	FALSE	0.29
13832.p1	Sanders et al	TRUE	FALSE	FALSE	0.27
13832.s1	Sanders et al	TRUE	FALSE	FALSE	0.28
13835.fa	O'Roak et al	FALSE	FALSE	FALSE	0.33
13835.mo	O'Roak et al	FALSE	FALSE	FALSE	0.37
13835.p1	O'Roak et al	FALSE	FALSE	FALSE	0.36
13835.s1	O'Roak et al	FALSE	FALSE	FALSE	0.38
13840.fa	Sanders et al	TRUE	FALSE	FALSE	0.25
13840.mo	Sanders et al	TRUE	FALSE	FALSE	0.27
13840.p1	Sanders et al	TRUE	FALSE	FALSE	0.26
13840.s1	Sanders et al	TRUE	FALSE	FALSE	0.24
13843.fa	Sanders et al	TRUE	FALSE	FALSE	0.22
13843.mo	Sanders et al	TRUE	FALSE	FALSE	0.24
13843.p1	Sanders et al	TRUE	FALSE	FALSE	0.24
13843.s1	Sanders et al	TRUE	FALSE	FALSE	0.24
13876.fa	Sanders et al	TRUE	FALSE	FALSE	0.21
13876.mo	Sanders et al	TRUE	FALSE	FALSE	0.24

13876.p1	Sanders et al	TRUE	FALSE	FALSE	0.24
13876.s1	Sanders et al	TRUE	FALSE	FALSE	0.30
13887.fa	Sanders et al	TRUE	FALSE	FALSE	0.29
13887.mo	Sanders et al	TRUE	FALSE	FALSE	0.27
13887.p1	Sanders et al	TRUE	FALSE	FALSE	0.29
13887.s1	Sanders et al	TRUE	FALSE	FALSE	0.30
13890.fa	O'Roak et al	TRUE	FALSE	FALSE	0.34
13890.mo	O'Roak et al	TRUE	FALSE	FALSE	0.34
13890.p1	O'Roak et al	TRUE	FALSE	FALSE	0.45
13890.s1	O'Roak et al	TRUE	FALSE	FALSE	0.37
13912.fa	Sanders et al	TRUE	FALSE	FALSE	0.24
13912.mo	Sanders et al	TRUE	FALSE	FALSE	0.24
13912.p1	Sanders et al	TRUE	FALSE	FALSE	0.27
13912.s1	Sanders et al	TRUE	FALSE	FALSE	0.23
13922.fa	Sanders et al	TRUE	FALSE	FALSE	0.24
13922.mo	Sanders et al	TRUE	FALSE	FALSE	0.25
13922.p1	Sanders et al	TRUE	FALSE	FALSE	0.26
13922.s1	Sanders et al	TRUE	FALSE	FALSE	0.26
13926.fa	O'Roak et al	TRUE	FALSE	FALSE	0.37
13926.mo	O'Roak et al	TRUE	FALSE	FALSE	0.38
13926.p1	O'Roak et al	TRUE	FALSE	FALSE	0.37
13926.s1	O'Roak et al	TRUE	FALSE	FALSE	0.38
13992.fa	Sanders et al	TRUE	FALSE	FALSE	0.23
13992.mo	Sanders et al	TRUE	FALSE	FALSE	0.27
13992.p1	Sanders et al	TRUE	FALSE	FALSE	0.24
13992.s1	Sanders et al	TRUE	FALSE	FALSE	0.26
14009.fa	Sanders et al	TRUE	FALSE	FALSE	0.27
14009.mo	Sanders et al	TRUE	FALSE	FALSE	0.23
14009.p1	Sanders et al	TRUE	FALSE	FALSE	0.25
14009.s1	Sanders et al	TRUE	FALSE	FALSE	0.24
14011.fa	O'Roak et al	FALSE	FALSE	FALSE	0.40
14011.mo	O'Roak et al	FALSE	FALSE	FALSE	0.35
14011.p1	O'Roak et al	FALSE	FALSE	FALSE	0.30
14011.s1	O'Roak et al	FALSE	FALSE	FALSE	0.33
14110.fa	Sanders et al	TRUE	FALSE	FALSE	0.29
14110.mo	Sanders et al	TRUE	FALSE	FALSE	0.30
14110.p1	Sanders et al	TRUE	FALSE	FALSE	0.31
14110.s1	Sanders et al	TRUE	FALSE	FALSE	0.36
14167.fa	Sanders et al	TRUE	FALSE	FALSE	0.30
14167.mo	Sanders et al	TRUE	FALSE	FALSE	0.31
14167.p1	Sanders et al	TRUE	FALSE	FALSE	0.30
14167.s1	Sanders et al	TRUE	FALSE	FALSE	0.30
14201.fa	O'Roak et al	FALSE	FALSE	FALSE	0.31
14201.mo	O'Roak et al	FALSE	FALSE	FALSE	0.33
14201.p1	O'Roak et al	FALSE	FALSE	FALSE	0.33
14201.s1	O'Roak et al	FALSE	FALSE	FALSE	0.37

Table S2

	Study	Quads (before QC)	Reads (36mers) mapped with mrsFAST to exome targets (per sample)		CoNIFER Standard Deviation Average	Accession Codes	
			Median	Minimum			
	O'Roak et al. (2012)	70 (70)	137,828,338	54,895,318	0.39	dbGAP: phs000482.v1.p1 or NDAR: NDARCOL0001878	
	Iossifov et al. (2012)	165 (165)	111,781,453	51,482,969	0.47	NDAR: NDARCOL0001936	
	Sanders et al. (2012)	176 (177)	160,294,934	48,431,443	0.29	SRA: SRP010920.1	
	All probands	411 (412)	142,557,586	48,431,443	0.38		
	All siblings	411 (412)	138,779,930	51,482,969	0.38		
	All mothers	411 (412)	134,381,120	51,504,433	0.38		
	All fathers	411 (412)	140,013,304	51,366,451	0.38		
	Discordant SRS	276	134,837,454	51,482,969	0.38		
	Concordant SRS	115	146,397,189	48,431,443	0.37		
	All	411 (412)	138,205,341	48,431,443	0.38		

Table S3

cnvId	calld	familyID	rel	Chromosome	Start (hg19)	Stop (hg19)	length (bp)	length (exons)	state	Frequency in 411 families	ESP Frequency	Inheritance	Genes	SRS discordant	Previously seen genes	de novo SNV summary in sample
16	5638	13698 p1	1	2106662	2234542	127880	13 dup			1	≤ 0.1% mo to both		C1orf86, PRKCZ, SKI	FALSE	PRKCZ	
16	8285	13698 s1	1	2118621	2940	5 dup				1	≤ 0.1% mo to both		PRKCZ	FALSE	PRKCZ	
19	11571 p1	1	2529645	2535808	8863	7 del				1	≤ 0.1% mo to p1		MMEL1	TRUE		
21	1548	12393 p1	1	3413218	3417328	4110	8 dup			3	≤ 0.5% fa to p1		MEGF6	FALSE		
21	1749	11336 p1	1	3424358	3432091	7733	9 del			2	≤ 0.1% mo to p1		MEGF6	TRUE		SLC26A5 [missense]
34	105	13798 p1	1	9063358	9115548	108190	27 dup			1	≤ 0.1% fa to p1		SLC2A7, SLC2A5, GPR157	FALSE		
39	5495	12618 p1	1	11134287	11155938	21651	16 del			1	≤ 0.1% mo to both		EXOSC10	TRUE		
39	5496	12618 s1	1	11134287	11155938	21651	16 del			1	≤ 0.1% mo to both		EXOSC10	TRUE		
45	5483	12498 p1	1	12020702	12030873	10771	8 dup			1	≤ 0.1% mo to both		PLOD1	TRUE	PLOD1	GPR82 [missense]
45	8286	12498 p1	1	12020702	12027148	6446	7 dup			1	≤ 0.1% mo to both		PLOD1	FALSE	PLOD1	CFAI [missense]
87	1622	13730 p1	1	32084793	32146688	61875	46 dup			1	≤ 0.1% mo to p1		HCRTR1, COL18A1, PEF1	FALSE		DICER1 [missense]
86	7978	12647 p1	1	32084793	32110465	25672	12 dup			2	≤ 0.1% mo to p1		HCRTR1, PEF1	TRUE		SLC30A5 [missense]
100	1561	11433 s1	1	40204572	40312869	108397	21 dup			1	≤ 0.1% fa to both		PPIE, TRIT1, BMP8B	TRUE		
100	1560	11433 p1	1	40204572	40313332	107496	21 dup			1	≤ 0.1% fa to both		PPIE, TRIT1, BMP8B	TRUE		
106	1581	11667 s1	1	42593553	42744343	50790	3 dup			7	≤ 0.1% fa to both		FOXJ3	TRUE		CDH3 [missense]
112	1522	11203 s1	1	47512138	47515846	3708	4 dup			7	≤ 2% mo to s1		CYP4X1	TRUE		
112	5458	12394 p1	1	47512138	47515846	3708	4 dup			7	≤ 2% fa to p1		CYP4X1	TRUE		
112	5463	12396 p1	1	47512138	47515846	3708	4 dup			7	≤ 2% mo to both		CYP4X1	TRUE		ZNF518A [missense]
112	5464	12396 s1	1	47512138	47515846	3708	4 dup			7	≤ 2% mo to both		CYP4X1	TRUE		MTHFS [frameshift], EFCAB5 [frameshift]
112	5622	13590 p1	1	47512138	47515846	3708	4 dup			7	≤ 2% fa to p1		CYP4X1	TRUE		
112	1831	13809 p1	1	47512138	47515846	3708	4 dup			7	≤ 2% mo to p1		CYP4X1	TRUE		
115	1508	11146 p1	1	48865022	48869554	4532	2 del			1	≤ 1% mo to both		SPATA6	TRUE		DIP2B [missense]
115	1509	11146 s1	1	48865022	48869554	4532	2 del			1	≤ 1% mo to both		SPATA6	TRUE		
123	88	13346 s1	1	57340621	57341882	1261	2 del			1	≤ 0.1% mo to s1		CBA	FALSE		
125	40	11872 p1	1	65730593	65831879	101286	4 dup			1	≤ 0.1% fa to p1		DNAJC6	TRUE		CACNA1D [missense], KATNAL2 [splice]
126	5555	13097 p1	1	65849875	65855310	5435	6 dup			2	≤ 2% mo to both		DNAJC6	FALSE		
126	5557	13097 s1	1	65849875	65855310	5435	6 dup			2	≤ 2% mo to both		DNAJC6	FALSE		
127	1499	11118 p1	1	66837995	67000051	162056	2 dup			1	≤ 0.1% fa to p1		PDE4B, SGP1	TRUE	PDE4B	SLCO1C1 [missense], A2M [missense], MYO7B
133	1792	12906 p1	1	74663962	74836076	172114	21 dup			1	≤ 0.1% mo to p1		PPGF-TNNI3K, PPGT	TRUE		
134	5487	12518 s1	1	84944064	84962053	17089	8 dup			1	≤ 0.1% mo to s1		RPF1	FALSE		
138	69	12810 p1	1	87029343	87038403	9060	6 del			5	≤ 2% mo to p1		CLCA4	TRUE		
138	5561	13144 p1	1	87029343	87040438	11095	7 del			5	≤ 2% mo to p1		CLCA4	TRUE		
138	5596	13387 s1	1	87029343	87040438	11095	7 del			5	≤ 2% mo to s1		CLCA4	TRUE		
141	7590	11000 p1	1	82262843	82373558	310715	37 dup			1	≤ 0.1% fa to p1		TGFBP3, BRDT, EPHX4, BTBD8	FALSE		
146	8296	13096 s1	1	10406692	10409402	25710	14 dup			2	≤ 0.1% mo to s1		RNP3	TRUE		
157	25	11610 s1	1	111724688	11183088	136280	37 dup			1	≤ 0.1% fa to s1		CHAC, CEPT1, CHRL2, DENND2D	FALSE		
160	8298	13296 s1	1	113245164	113264970	19786	13 dup			1	≤ 0.1% fa to s1		RHOA, PPP1U, FAM19A3	TRUE		
165	9628	13601 p1	1	115316891	115323228	6337	5 dup			1	≤ 0.1% fa to both		SIKE1	TRUE		AK1 [missense]
165	9629	13601 s1	1	115316891	115323236	5945	4 dup			1	≤ 0.1% fa to both		SIKE1	TRUE		
195	5571	13215 s1	1	145415278	145923443	508165	156 del			1	≤ 0.1% fa to s1		RNF115, GPR88A, RBM8A, PIAS3, CD160, HFE2, ANKRD34A, LIX1, POLR3GL, ANKRD35, ITGA10, PEX11B, NUDT17, TXNIP, GPR88C, POZK1, POLR3C	TRUE		
187	5513	12719 p1	1	146715494	146767190	51696	23 del			1	≤ 0.1% mo to p1		CHD1L	TRUE	CHD1L	
204	1809	13355 p1	1	150955581	150967175	11594	11 del			1	≤ 0.1% fa to both		ANKA9	TRUE		
204	1810	13355 s1	1	150955581	150967175	11594	11 del			1	≤ 0.1% fa to both		ANKA9	TRUE		
206	5509	12691 p1	1	151337018	151403177	11362	10 del			1	≤ 0.1% mo to p1		SELENBP1, POZ3, PSMB4	FALSE		
222	5627	13599 p1	1	156890595	156933386	42791	35 dup			1	≤ 0.1% mo to p1		ARHGAP11, LRRC71	TRUE		TMEM62 [missense]
225	1555	11411 s1	1	158323778	158326886	2908	6 del			1	≤ 0.1% fa to s1		CDIE	TRUE		DCLRE1B [missense]
252	55	12161 s1	1	182550359	182555941	5582	4 del			1	≤ 0.1% mo to s1		RNASL	FALSE		ASA42 [frameshift]
255	36	11715 p1	1	185097800	185130057	32257	11 dup			3	≤ 0.1% fa to both		TRMT1L, SWT1	TRUE		
255	37	11715 s1	1	185097800	185130057	32257	11 dup			3	≤ 0.1% fa to both		TRMT1L, SWT1	TRUE		
255	8312	12313 p1	1	185097800	185130057	32257	11 dup			3	≤ 0.5% mo to p1		TRMT1L, SWT1	TRUE		
255	5623	13590 p1	1	185097800	185130057	32257	11 dup			3	≤ 0.5% mo to both		TRMT1L, SWT1	TRUE		MTHFS [frameshift], EFCAB5 [frameshift]
255	5624	13590 s1	1	185106737	185121067	14330	9 dup			3	≤ 0.5% mo to both		TRMT1L	TRUE		
257	5488	12518 s1	1	190195211	190250880	55669	4 del			2	≤ 0.1% mo to s1		FAM5C	FALSE		
260	1863	14110 p1	1	197479589	197480283	2494	3 del			6	≤ 1% mo to p1		DENND1B	FALSE		PHF3 [missense]
277	1693	12175 p1	1	213003583	213009491	5908	2 del			6	≤ 2% fa to p1		C1orf227	TRUE		
279	116	13890 p1	1	220088790	220100447	11857	3 dup			1	≤ 0.1% fa to both		RNU5F	TRUE		DYRK1A [splice]
279	117	13890 s1	1	220088790	220100447	11857	3 dup			1	≤ 0.1% fa to both		RNU5F	TRUE		
286	7648	12403 p1	1	230313963	230339036	25073	2 del			1	≤ 0.1% fa to p1		GALNT2	TRUE		
295	1640	12162 p1	1	230313963	232094638	1780675	160 del			1	≤ 0.1% mo to p1		C1orf124, C1orf198, TSNAX-DISC1, GALNT2, C1orf131, GNPAT, FAM89A, ARV1, AGT, CAPN9, TRIM67	TRUE	AGT, GNPAT	
308	1725	12308 p1	1	245026918	245133745	106827	2 dup			1	≤ 0.1% mo to p1		HNRNPJ	TRUE		ZNF333 [missense], BRD1 [missense]
309	1544	11301 p1	1	245530135	245722314	192179	4 del			1	≤ 0.1% fa to both		KIF26B	TRUE		
309	1545	11301 s1	1	245530135	245722314	192179	4 del			1	≤ 0.1% fa to both		KIF26B	TRUE		
311	1597	11797 s1	1	245912864	246805339	892475	27 dup			2	≤ 0.1% mo to s1		TFB2M, CNST, SMYD3	TRUE		
311	1523	11203 s1	1	246490502	246811354	320852	21 dup			2	≤ 0.1% mo to s1		TFB2M, CNST, SMYD3	TRUE		
312	7633	11676 s1	1	246927547	246930602	3055	3 dup			1	≤ 0.1% mo to s1		SCOPDH	TRUE		
316	8327	12463 p1	1	247599271	247769817	170546	14 dup			1	≤ 0.1% fa to p1		OR2B11, NLRP3, OR2G3, OR2G2, OR2C3, OR2W5, C1orf150	TRUE	NLRP3	UNC80 [nonsense]
318	8328	12463 p1	1	248023918	248113098	89180	8 dup			1	≤ 0.5% fa to p1		OR2W5, C1orf150	TRUE		UNC80 [nonsense]
1605	5679	12837 p1	2	13968683	1546246	149363	28 dup			2	≤ 0.5% mo to p1		TPO	TRUE		UNC80 [nonsense], SH3RF3 [missense]

1605	5672	12733 s1	2	1426816	1520754	93938	19 dup	2	≤ 0.5% mo to s1	TPO	FALSE	RANBP9 [missense]
1604	5709	13296 s1	2	1437209	1479843	42634	8 dup	1	≤ 0.5% mo to s1	TPO	TRUE	
1628	2030	12729 p1	2	29117564	2916440	46876	15 dup	1	≤ 0.1% mo to both	WDR43	TRUE	
1628	2032	12729 s1	2	29117564	29169645	52081	18 dup	1	≤ 0.1% mo to both	WDR43	TRUE	
1632	5701	13176 s1	2	32631566	33246273	614707	87 dup	1	≤ 0.5% mo to s1	BIRC6, TTC27, LTBP1	TRUE	
1633	5674	12826 p1	2	44071645	44073450	1805	2 dup	1	≤ 0.1% mo to p1	ABCG8	TRUE	TROVE2 [frameshift]
1634	1888	11118 p1	2	44527109	44541090	13981	5 dup	2	≤ 0.1% fa to both	SLC3A1	TRUE	
1634	7946	11118 s1	2	44527109	44541090	13981	5 dup	2	≤ 0.1% fa to both	SLC3A1	TRUE	
1634	146	11472 p1	2	44527109	44541090	13981	5 dup	2	≤ 0.1% fa to both	SLC3A1	TRUE	KRT80 [missense], SP7 [missense]
1634	147	11472 s1	2	44527109	44541090	13981	5 dup	2	≤ 0.1% fa to both	SLC3A1	TRUE	MACC1 [missense], ZYG11A [missense]
1635	2031	12729 p1	2	45616448	45832580	216132	20 dup	1	≤ 0.5% fa to both	SRBD1	TRUE	
1635	2033	12729 s1	2	45616448	45879587	263139	21 dup	1	≤ 0.5% fa to both	SRBD1, PRKCE	TRUE	
1644	7950	11716 p1	2	61558410	62998527	1440117	76 dup	1	≤ 0.1% fa to p1	XPO1, FAM161A, EHBP1, TMEM17, BSGNT2, COMMD1, CCT4, USP34	TRUE	
1652	1985	12228 p1	2	74129486	74166149	36663	64 dup	4	≤ 0.1% mo to p1	ACTG2, DGUOK	TRUE	CYP4F3 [missense]
1661	175	11895 p1	2	86292396	86509365	216969	9 dup	2	≤ 0.1% fa to p1	REEP1, POLR1A, MRPL35, PTCO3, IMMT	FALSE	
1666	5682	12851 s1	2	96780544	97784254	1003710	255 dup	1	≤ 0.1% mo to s1	FER1L5, CIAO1, ANKRD36, SEMA4C, ANKRD39, SNRPB200, ASTL, ADRA28, NEURL3, LMAN2L, DUSP2, TMEM127, ITPRIPL1, ANKRD23, FAHD2B, CNM3, CNM4, ARID5A, LOC285093, STARD7, NCAPH, KIAA1310, FAM178B	TRUE	
1674	5735	13507 p1	2	98192841	98275940	83099	18 dup	2	≤ 0.5% fa to both	ANKRD36B, COX5B, ACTR1B	FALSE	SOOC [frameshift], LAMA1 [missense]
1674	5736	13507 s1	2	98192841	98274779	81938	15 dup	2	≤ 0.5% fa to both	ANKRD36B, COX5B, ACTR1B	FALSE	TMEM85 [missense], PPP2R1B [missense]
1674	1930	11484 p1	2	98262549	98275940	13391	12 dup	2	≤ 0.5% mo to both	ACTR1B, COX5B	TRUE	USP34 [missense]
1674	1931	11484 s1	2	98263529	98275940	12411	11 dup	2	≤ 0.5% mo to both	ACTR1B, COX5B	TRUE	
1675	5695	13166 s1	2	98866780	98872637	5857	2 del	1	≤ 2% fa to s1	VWA3B	TRUE	
1677	5740	13508 s1	2	99858826	99912133	53307	10 dup	2	≤ 0.5% mo to s1	LVG1, LVG2	FALSE	
1677	2057	13795 p1	2	99900855	99909103	8248	4 dup	2	≤ 0.5% mo to both	LYGI	TRUE	
1680	7964	11045 p1	2	10247181	102416105	8924	3 dup	3	≤ 0.1% mo to p1	MAPK4	FALSE	
1683	5683	12691 p1	2	111395540	111439066	42526	26 dup	1	≤ 0.1% mo to p1	BUB1	FALSE	
1688	5656	12579 p1	2	111395540	113090065	1694525	167 dup	1	≤ 0.1% fa to both	ZC3H8, BCL2L1, ACOXL, ZC3H6, MERTK, ANAPC1, BUB1, FBLN7, TMEM78	TRUE	
1687	5657	12579 s1	2	111395540	113090065	1694525	167 dup	1	≤ 0.1% fa to both	ZC3H8, BCL2L1, ACOXL, ZC3H6, MERTK, ANAPC1, BUB1, FBLN7, TMEM78	TRUE	
1692	2045	13355 p1	2	113251719	113275002	26393	4 dup	1	≤ 0.1% mo to p1	TTL	TRUE	
1693	2016	13552 s1	2	113346442	11340739	58297	2 dup	1	≤ 2% mo to s1	SLC20A1, CHCHD5	TRUE	IL1A
1694	5654	12526 s1	2	113537072	113541347	4275	4 del	1	≤ 0.1% mo to s1	IL1A	TRUE	
1699	5716	13418 p1	2	116535362	116699921	64559	12 dup	1	≤ 0.1% fa to p1	DPP10	FALSE	GNL1 [missense], DENND5B [missense], LRRC40 [missense]
1703	5749	13601 p1	2	128414976	128945188	530212	107 dup	1	≤ 0.1% fa to both	LIMS2, POLR2D, AMMECR1L, WDR33, SAP130, UGGT1	TRUE	AK1 [missense]
1703	5750	13601 s1	2	128463896	128945188	481292	104 dup	1	≤ 0.1% fa to both	POLR2D, AMMECR1L, SAP130, WDR33, UGGT1	TRUE	
1716	5675	12826 p1	2	135206219	135215744	9525	3 dup	1	≤ 0.1% mo to both	TMEM163, MGAT5	TRUE	TROVE2 [frameshift]
1716	8556	12826 s1	2	135206219	135215744	9525	3 dup	1	≤ 0.1% mo to both	TMEM163, MGAT5	TRUE	
1718	1938	11519 s1	2	159663566	159922483	258917	3 dup	1	≤ 0.1% mo to s1	TANC1, DAPL1	TRUE	TANC1
1719	164	11711 p1	2	160585533	160605414	19881	4 dup	1	≤ 0.1% mo to both	MARCH7	TRUE	KIAA0182 [missense]
1719	166	11711 s1	2	160585533	160605414	19881	4 dup	1	≤ 0.1% mo to both	MARCH7	TRUE	
1722	5707	13269 s1	2	165381513	165600385	218872	17 del	1	≤ 0.1% fa to s1	GRB14, COBL1	FALSE	
1743	5683	12851 s1	2	192711168	193059250	348082	11 del	1	≤ 0.1% mo to s1	TMEFF2, SDPR	TRUE	
1754	1946	11622 p1	2	211299210	211542709	243499	48 del	1	≤ 0.1% mo to both	LANCL1, CPS1	TRUE	CP51
1754	1947	11622 s1	2	211299210	211542709	243499	48 del	1	≤ 0.1% mo to both	LANCL1, CPS1	TRUE	CP51
1762	5644	12420 p1	2	224758950	224831751	72801	12 dup	1	≤ 0.1% fa to both	WDFY1, MRPL44	FALSE	HRH2 [missense], GOLGA4 [missense]
1762	5645	12420 s1	2	224758950	224831751	72801	12 dup	1	≤ 0.1% fa to both	WDFY1, MRPL44	FALSE	VPS18 [missense]
1764	5689	12997 p1	2	230632269	230724290	92021	39 dup	1	≤ 0.1% fa to p1	TRIPT2	TRUE	
1767	160	11659 s1	2	232070951	232072965	2014	2 del	1	≤ 0.1% mo to s1	ARMQ9	FALSE	
1770	5767	12175 s1	2	234181611	234293469	47858	16 dup	1	≤ 0.1% fa to s1	ATG16L1, SAG	TRUE	
1778	5748	13589 p1	2	241439374	241514045	74871	19 dup	1	≤ 0.1% fa to both	DUSP23, ANKMY1, RNPERL1	TRUE	WDR55 [missense]
1778	5747	13589 s1	2	241439368	241514045	74177	18 dup	1	≤ 0.1% fa to both	DUSP23, ANKMY1, RNPERL1	TRUE	
1780	5640	12394 p1	2	241598067	241709123	171056	42 dup	3	≤ 0.5% mo to p1	KIF1A, GPR35, AQP12B, AQP12A, CAPN10	TRUE	KIF1A
1782	1894	11107 p1	2	242371093	242375953	4860	4 del	1	≤ 0.1% mo to both	FARP2	TRUE	TROAP [missense]
1782	1895	11107 s1	2	242371093	242375953	4860	4 del	1	≤ 0.1% mo to both	FARP2	TRUE	
2232	5765	12368 p1	3	7594608	7782033	187285	6 del	1	≤ 0.1% mo to both	GRM7	TRUE	TEKT1 [missense]
2232	5766	12368 s1	3	7594608	7782033	187285	6 del	1	≤ 0.1% mo to both	GRM7	TRUE	NBP79 [missense]
2237	2179	11824 p1	3	8578670	8672615	93745	11 dup	1	≤ 0.1% fa to both	LMCO1, C3orf82	FALSE	EBA39 [missense]
2237	2180	11824 s1	3	8578670	8675624	96754	13 dup	1	≤ 0.1% fa to both	LMCO1, C3orf82	FALSE	
2238	5784	12851 p1	3	9867463	9874929	7446	4 dup	1	≤ 0.1% mo to both	APPC4-TTL3	TRUE	SLC25A29 [missense], HMGXB3 [missense], UBE3C [missense]
2238												
2238	8377	12851 s1	3	9867463	9871079	3596	3 dup	1	≤ 0.1% mo to both	APPC4-TTL3	TRUE	
2242	2182	11905 p1	3	12475396	12791331	315935	47 dup	2	≤ 0.1% mo to p1	PPARG, MKRN2, RAF1, TSEN2, TMEM40	TRUE	
2242	5761	12481 p1	3	12632296	12791331	159035	22 dup	2	≤ 0.1% mo to both	RAF1, TMEM40	FALSE	
2242	5762	12481 s1	3	12632296	12791331	159035	22 dup	2	≤ 0.1% mo to both	RAF1, TMEM40	FALSE	
2236	2239	12534 p1	3	12940688	12978197	37309	13 del	2	≤ 0.1% mo to p1	IQSEC1	TRUE	MTMR9 [missense], MMP8 [missense]
2236												
2236	5778	12727 p1	3	12940688	12983365	42477	14 dup	2	≤ 0.1% mo to both	IQSEC1	TRUE	ZMAT5 [missense]
2236	5779	12727 s1	3	12942481	12983365	40884	13 dup	2	≤ 0.1% mo to both	IQSEC1	TRUE	TCF12 [frameshift]
2262	2258	13322 s1	3	31917924	31921322	3398	2 del	1	≤ 0.1% mo to s1	OSBPL10	TRUE	HPS6 [missense]

2265	2163	11501 s1	3	35833873	35835450	1577	2 dup	3	≤ 0.5% fa.to.s1	ARPP21	FALSE	ANPEP [missense]
2265	2232	12383 s1	3	35833873	35835450	1577	2 dup	3	≤ 0.5% fa.to.s1	ARPP21	FALSE	
2265	2261	13393 p1	3	35833873	35835450	1577	2 dup	3	≤ 0.5% mo.to.both	ARPP21	FALSE	
2265	2265	13393 s1	3	35833873	35835450	1577	2 dup	3	≤ 0.5% mo.to.both	ARPP21	FALSE	
2266	5819	13512 p1	3	37095341	37096656	1315	2 dup	1	≤ 0.1% fa.to.p1	LRPP12	TRUE	
2301	8380	12683 p1	3	57143564	57144339	775	2 del	1	≤ 0.1% mo.to.p1	IL17FD	FALSE	KDM6B [frameshift]
2307	7798	13094 p1	3	7864062	78796050	147988	27 dup	1	≤ 0.1% fa.to.both	ROBO1	FALSE	WDFY3 [nonsense]
2307	5789	12092 s1	3	7864062	78796050	147988	26 dup	1	≤ 0.1% fa.to.both	ROBO1	FALSE	ROBO1 [nonsense]
2308	5785	12354 p1	3	81627075	81640315	13240	4 del	1	≤ 0.1% mo.to.both	GBE1	FALSE	CNO14 [missense]
2308	5766	12252 p1	3	81627075	81640315	13240	4 del	1	≤ 0.1% mo.to.both	GBE1	FALSE	
2309	5791	12099 p1	3	97486951	97934890	147929	19 del	1	≤ 0.1% fa.to.p1	Nr6a, ARL6	TRUE	RIMS1 [frameshift]
2317	5796	13162 p1	3	113586353	113619993	31640	5 del	2	≤ 0.1% mo.to.both	GRAMD1C	TRUE	
2317	5797	13162 s1	3	113586353	113619993	31640	5 del	2	≤ 0.1% mo.to.both	GRAMD1C	TRUE	
2326	260	13890 p1	3	129120435	129127701	7266	4 del	1	≤ 0.5% mo.to.p1	C3orf25	TRUE	DYRK1A [plice]
2329	2172	11676 p1	3	132277814	132280061	2247	3 del	7	≤ 1% mo.to.p1	NHPH3-ACAD11	TRUE	
2329	237	12304 p1	3	132277814	132280061	2247	3 del	7	≤ 1% mo.to.p1	NHPH3-ACAD11	FALSE	STIL [missense], PSEN1 [missense], PHF19 [missense]
2329	2266	13660 s1	3	132277814	132280061	2247	3 del	7	≤ 1% fa.to.s1	NHPH3-ACAD11	TRUE	ZNF423 [missense]
2329	255	13793 s1	3	132277814	132280061	2247	3 del	7	≤ 1% mo.to.s1	NHPH3-ACAD11	TRUE	
2331	2098	11067 p1	3	137781657	137816689	35032	14 dup	2	≤ 0.1% mo.to.both	D2JIP1L	TRUE	
2331	2099	11067 s1	3	137781657	137816689	35032	14 dup	2	≤ 0.1% mo.to.both	D2JIP1L	TRUE	
2331	230	12106 s1	3	137781657	137803095	21438	9 dup	2	≤ 0.1% mo.to.s1	D2JIP1L	TRUE	
2335	2096	11057 s1	3	141884463	142084208	199745	31 dup	6	≤ 0.5% fa.to.s1	GK5, XRN1	FALSE	
2335	2124	11108 p1	3	141884463	142075961	191498	28 dup	6	≤ 0.5% mo.to.both	GK5, XRN1	FALSE	KANK1 [missense]
2335	2126	11108 s1	3	141884463	142084021	199558	30 dup	6	≤ 0.5% mo.to.both	GK5, XRN1	FALSE	MAN2A1 [missense]
2335	2152	11304 s1	3	141884463	142084021	199558	30 dup	6	≤ 0.5% fa.to.s1	GK5, XRN1	TRUE	KLC2 [missense]
2335	247	13335 p1	3	141884463	142084208	199745	31 dup	6	≤ 0.5% fa.to.both	GK5, XRN1	FALSE	ZNF420 [missense]
2335	248	13335 s1	3	141884463	142084208	199745	31 dup	6	≤ 0.5% fa.to.both	GK5, XRN1	FALSE	
2335	5768	12631 p1	3	141889166	142075961	189795	27 dup	6	≤ 0.5% mo.to.both	GK5, XRN1	FALSE	
2335	5769	12631 s1	3	141889166	142084208	195042	30 dup	6	≤ 0.5% mo.to.both	GK5, XRN1	FALSE	GAPVD1 [missense], YF1A [missense]
2334	5900	13176 s1	3	141896323	141901891	5568	3 del	2	≤ 0.1% fa.to.both	GK5	TRUE	
2334	8388	13176 p1	3	141896323	141901891	5568	3 del	2	≤ 0.1% fa.to.both	GK5	TRUE	ZFYVE26 [frameshift]
2341	2218	12297 s1	3	150280328	150286079	5751	6 del	1	≤ 0.1% fa.to.s1	EIF2A	FALSE	
2342	8389	13487 p1	3	155481305	155493637	12332	3 dup	2	≤ 0.5% fa.to.p1	C3orf33	FALSE	GNAI4 [missense]
2342	2279	13922 p1	3	155481305	155493637	12332	3 dup	2	≤ 1% fa.to.s1	C3orf33	TRUE	CYP4Z1 [missense]
2342	2294	11798 p1	3	183829574	183923757	1183	3 del	2	≤ 0.1% mo.to.p1	HTB9E	TRUE	YTHDC2 [missense]
2356	2103	11075 p1	3	184104838	184299170	184332	7 del	1	≤ 0.1% mo.to.both	EPH83, CHRDL	FALSE	
2349	2104	11075 s1	3	184104838	184107210	2372	6 del	1	≤ 0.1% mo.to.both	CHRDL	FALSE	MED12L [missense], C4orf40 [missense]
2357	244	13169 s1	3	194947437	195453443	506006	38 dup	1	≤ 0.1% mo.to.s1	C3orf21, APOD, MUC20, ACAP2, PPP1R2	FALSE	
2362	245	13169 s1	3	197238765	197260432	21667	5 dup	1	≤ 0.1% mo.to.s1	BDH1	FALSE	
2347	2154	11336 p1	3	197553748	197556544	2796	2 del	1	≤ 0.1% mo.to.both	LHCH3	TRUE	SLC26A5 [missense]
2347	2155	11336 s1	3	197553748	197556544	2796	2 del	1	≤ 0.1% mo.to.both	LHCH3	TRUE	RGS7 [missense]
2364	2125	11108 p1	3	197574279	197640913	66634	14 dup	3	≤ 0.1% mo.to.p1	IQCG, LRCH3	FALSE	KANK1 [missense]
2386	5834	12645 p1	4	818279	845762	27483	5 dup	1	≤ 0.1% mo.to.p1	CPLX1, GAK	TRUE	ARHGAP21 [missense], ANK2 [nonsense], LRRC31 [missense]
2392	2389	12340 p1	4	1746674	1795770	49096	2 dup	1	≤ 0.1% fa.to.p1	TACC3, FGFR3	TRUE	PPM1D [nonsense], BCORL1 [missense], KRBA1 [missense], CACNA1E [missense]
2395	7474	11773 p1	4	2641461	2833561	194100	34 dup	1	≤ 0.1% mo.to.p1	TNIP2, FAM183A, SH3BP2	TRUE	
2398	2324	11219 p1	4	2906490	3201666	295176	93 dup	1	≤ 0.1% mo.to.both	ADD1, GRK4, C4orf10, MFSD10, HTT, NOP14	FALSE	
2398	2325	11219 s1	4	2906490	3201666	295176	93 dup	1	≤ 0.1% mo.to.both	ADD1, GRK4, C4orf10, MFSD10, HTT, NOP14	FALSE	
2399	2304	11060 p1	4	3445765	3450003	4238	9 del	4	≤ 0.5% fa.to.p1	HGFAC	FALSE	
2399	8152	11700 p1	4	3446991	3449762	2771	4 del	4	≤ 0.5% mo.to.p1	HGFAC	FALSE	
2403	278	12161 p1	4	6594899	6613005	18106	10 del	1	≤ 0.1% mo.to.both	MAN2B2	FALSE	UBR3 [frameshift], CARKO [nonsense]
2403	279	12161 s1	4	6594899	6613005	18106	10 del	1	≤ 0.1% mo.to.both	MAN2B2	FALSE	
2419	2310	11075 p1	4	40762450	40778276	15826	5 dup	1	≤ 0.5% mo.to.both	NSUN7	FALSE	MED12L [missense], C4orf40 [missense]
2419	2311	11075 s1	4	40762450	40778276	15826	5 dup	1	≤ 0.5% mo.to.both	NSUN7	FALSE	
2420	2305	11066 p1	4	41258993	41259143	150	2 dup	3	≤ 0.1% fa.to.p1	UCHL1	TRUE	UCHL1
2423	8403	12982 p1	4	47625587	47625965	378	2 del	1	≤ 0.1% mo.to.s1	CORIN	TRUE	
2425	2379	13524 p1	4	48165718	48173203	12485	6 dup	1	≤ 0.1% fa.to.p1	TEC	TRUE	
2429	2302	11014 p1	4	57361522	57322178	16056	9 dup	1	≤ 0.1% mo.to.p1	ARL9, SRP72, HOPX	FALSE	
2445	5824	12420 p1	4	81201478	82092958	885480	26 dup	1	≤ 0.1% mo.to.both	FGF5, C4orf22, BMP3, PRKG2	FALSE	
2445	5825	12420 s1	4	81201478	82092958	885480	26 dup	1	≤ 0.1% mo.to.both	FGF5, C4orf22, BMP3, PRKG2	FALSE	HRH2 [missense], GOLGA4 [missense]
2448	5821	12409 p1	4	89978088	90035703	57615	2 del	1	≤ 0.1% mo.to.both	TIGD2, FAM13A	TRUE	VPS18 [missense]
2448	5822	12409 s1	4	89978088	90035703	57615	2 del	1	≤ 0.1% mo.to.both	TIGD2, FAM13A	TRUE	LRP2 [nonsense]
2450	292	13533 p1	4	90655960	90874569	18609	3 del	2	≤ 0.1% fa.to.p1	MMRN1	FALSE	
2453	2341	11622 p1	4	100738070	100782782	44712	5 dup	1	≤ 0.1% fa.to.both	DAPP1	TRUE	
2453	2342	11622 s1	4	100738070	100774505	36435	4 dup	1	≤ 0.1% fa.to.both	DAPP1	TRUE	
2457	263	11190 p1	4	106535404	106871579	336175	22 dup	1	≤ 0.1% fa.to.both	PAPSS1, SGMS2, CYP2U1	TRUE	
2457	264	11190 s1	4	106535404	106871579	336175	22 dup	1	≤ 0.1% fa.to.both	PAPSS1, SGMS2, CYP2U1	TRUE	
2457	5839	12770 p1	4	134071295	135122348	1051053	6 dup	1	≤ 0.1% fa.to.both	PCDH10	TRUE	PCDH10
2465	5841	12770 s1	4	134071295	135122348	1051053	6 dup	1	≤ 0.1% fa.to.both	PCDH10	TRUE	
2473	275	11959 s1	4	151727422	152070713	343291	49 dup	1	≤ 0.1% fa.to.s1	SH3D19, RPS3A, LRBA	FALSE	
2459	2373	12370 p1	4	159590764	159616795	26031	9 dup	1	≤ 0.1% mo.to.p1	C4orf46, ETFDH	TRUE	
2458	2390	13385 p1	4	169083678	169086477	2799	3 del	1	≤ 0.5% fa.to.p1	ANXA10	TRUE	
2462	2348	12100 p1	4	187192762	187476519	283757	15 dup	1	≤ 0.1% fa.to.p1	F11, MTRN1A	TRUE	

2463	8168	11429 s1	4	189018214	189068526	50312	12 dup	1	≤ 0.1% fa.to.s1	TRIM2L, TRIM1L TPPP ZDHC11, CEP72	FALSE	ILBR [missense]
2483	341	12373 p1	5	619104	801279	182175	14 dup	2	≤ 0.1% mo.to.both	TPPP CEP72	TRUE	TRUE
2483	342	12373 s1	5	619104	678175	59071	15 dup	2	≤ 0.1% mo.to.both	CEP72	TRUE	DNAH10 [missense], KIAA0240 [missense], MUC4 [missense]
2483	8408	13293 p1	5	619104	644540	25436	9 dup	2	≤ 0.1% fa.to.p1		TRUE	SEMA5A SEMA5A
2490	5918	13396 p1	5	9136615	9238002	101387	8 del	1	≤ 0.1% mo.to.both	SEMA5A	TRUE	SEMA5A
2490	5919	13396 p1	5	9136615	9337924	201309	10 del	1	≤ 0.1% mo.to.both	C7	TRUE	CDG123 [missense]
2503	313	11659 p1	5	40931165	4093792	6627	4 del	1	≤ 0.1% fa.to.both		FALSE	FALSE
2503	314	11659 p1	5	40931165	4093792	6627	4 del	1	≤ 0.1% fa.to.both		FALSE	FALSE
2537	2442	11466 s1	5	81283399	81354491	71032	2 dup	2	≤ 0.1% mo.to.s1	ATG10	TRUE	EPG5 [missense]
2537	5890	12851 s1	5	81283399	81354491	71032	2 dup	2	≤ 0.1% fa.to.s1	ATG10	TRUE	EPG5 [missense]
2545	3876	12368 s1	5	110427966	110446977	18991	15 dup	6	≤ 0.5% mo.to.s1	WR36	TRUE	NBP79 [nonsense]
2545	2345	12370 p1	5	110430617	110446977	16360	14 dup	6	≤ 0.5% mo.to.both	WR36	TRUE	HIST1H2AE [missense]
2545	2492	12370 s1	5	110430617	110446977	16360	14 dup	6	≤ 0.5% mo.to.both	WR36	TRUE	
2545	2508	12690 p1	5	110430617	110446977	16360	14 dup	6	≤ 0.5% mo.to.p1	WR36	TRUE	
2545	5939	13684 p1	5	110430617	110446977	16360	14 dup	6	≤ 0.5% mo.to.p1	WR36	TRUE	
2545	2533	13876 s1	5	110430617	110446977	16360	14 dup	6	≤ 0.5% fa.to.s1	WR36	FALSE	
2545	2520	13322 p1	5	110432776	110446977	14201	13 dup	6	≤ 0.5% fa.to.p1	WR36	TRUE	
2546	303	11469 p1	5	112915282	112920980	13798	6 dup	6	≤ 0.5% fa.to.both	YTHDC2	TRUE	
2546	304	11469 s1	5	112915282	112920980	13798	6 dup	6	≤ 0.5% fa.to.both	YTHDC2	TRUE	
2549	5911	13312 s1	5	121297727	12130365	32638	4 del	2	≤ 0.1% mo.to.s1	SFRBP1	FALSE	TRIM37 [missense] UBR3 [frameshift], CARKO [nonsense]
2549	336	12161 p1	5	121309890	121358102	48212	6 dup	2	≤ 0.1% mo.to.both	SFRBP1	FALSE	
2549	337	12161 s1	5	121309890	121362821	52931	7 dup	2	≤ 0.1% mo.to.both	SFRBP1	FALSE	
2550	2527	13825 p1	5	122161744	122163341	1597	3 del	1	≤ 0.5% fa.to.p1	SNX2	FALSE	
2551	5916	13387 s1	5	123973590	124036962	63372	8 dup	1	≤ 0.1% mo.to.s1	ZNF608	TRUE	
2566	346	12390 s1	5	141312823	141314151	1328	3 del	1	≤ 0.1% mo.to.s1	KIAA0141	TRUE	
2568	5937	13601 s1	5	147498548	147498699	1151	2 dup	2	≤ 0.1% mo.to.s1	SPINK5	TRUE	
2570	8183	13625 p1	5	150175002	150277746	102744	4 del	3	≤ 0.1% mo.to.p1	C5orf62, ZNF300	TRUE	
2572	2454	11828 p1	5	156456743	156479665	22922	6 dup	1	≤ 0.1% fa.to.both	HAVCR1	TRUE	
2572	2455	11828 p1	5	156456743	156479665	22922	6 dup	1	≤ 0.1% fa.to.both	HAVCR1	TRUE	
2574	2540	13922 p1	5	175992353	175995818	3465	3 del	1	≤ 0.1% mo.to.both	CDHR2	TRUE	
2574	2541	13922 s1	5	175992353	175995818	3465	3 del	1	≤ 0.1% mo.to.both	CDHR2	TRUE	
2577	2514	12906 s1	5	176314002	176318192	4190	10 dup	3	≤ 0.5% mo.to.s1	HK3	TRUE	
2577	5929	13504 s1	5	176314451	176317942	3491	7 dup	3	≤ 0.5% mo.to.both	HK3	TRUE	
2577	8436	13504 p1	5	176314451	176317942	3491	7 dup	3	≤ 0.5% mo.to.both	HK3	TRUE	
2611	2571	11115 p1	6	2949147	3017219	68072	10 del	2	≤ 0.1% fa.to.p1	NOO2, SERPINB6	TRUE	TUBGCP5 [missense]
2617	2698	13876 p1	6	20781375	20846409	65034	2 dup	1	≤ 0.1% mo.to.p1	CDKAL1	FALSE	ALDH5A1
2618	5978	12758 p1	6	24454242	24523153	66911	20 dup	2	≤ 0.1% mo.to.p1	ALDH5A1, GPLD1	FALSE	
2622	2584	11220 s1	6	33584673	33585067	214	2 dup	2	≤ 0.1% mo.to.s1	CUTA	TRUE	
2625	5986	13153 p1	6	35959435	35967885	8450	4 dup	4	≤ 0.1% mo.to.p1	SLC26A8	FALSE	MAPK13 [missense]
2657	5974	12743 p1	6	41243662	41318602	74740	9 dup	1	≤ 0.1% fa.to.p1	THUM1, NCR2	FALSE	
2658	7509	13335 s1	6	41709533	41710227	694	2 del	1	≤ 0.1% mo.to.s1	PGC	FALSE	
2660	5960	12637 s1	6	42983907	42992011	8204	3 del	1	≤ 0.1% fa.to.s1	PICHA	FALSE	CDC7 [missense]
2664	5955	12467 s1	6	43494986	43501744	5848	6 del	1	≤ 0.1% mo.to.both	XPO5	FALSE	
2664	5953	12467 p1	6	43494986	43501744	5179	5 del	1	≤ 0.1% mo.to.both	XPO5	FALSE	
2666	2685	13825 s1	6	43603612	43604383	771	2 dup	2	≤ 0.1% fa.to.s1	MAD2L1BP	FALSE	
2672	2585	11220 s1	6	47471015	47567179	96764	12 del	1	≤ 0.1% fa.to.s1	CD2AP	FALSE	
2674	378	11013 s1	6	49426794	49440571	13777	5 dup	2	≤ 2% mo.to.s1	MUT, CENPO	TRUE	
2679	383	11229 p1	6	51747890	51752043	4153	3 del	1	≤ 0.1% fa.to.p1	PKHD1	FALSE	KIAA1586
2679	394	12106 p1	6	56915571	56919661	4090	2 dup	2	≤ 0.5% fa.to.both	KIAA1586	TRUE	KIAA1586
2682	2661	12869 p1	6	65523270	66205303	682033	19 dup	1	≤ 0.5% fa.to.both	EYS	TRUE	PRCP [missense]
2683	2624	11828 s1	6	65596589	65622636	26047	4 del	1	≤ 0.1% mo.to.s1		TRUE	
2688	2612	11551 p1	6	88315634	88318947	3313	3 del	1	≤ 0.5% mo.to.p1	ORC3	TRUE	
2689	380	11459 p1	6	88317390	88366700	49310	10 del	1	≤ 0.1% fa.to.p1	ORC3	TRUE	DEPDC7 [missense]
2692	5987	13153 s1	6	99998868	99999771	903	2 dup	3	≤ 2% fa.to.s1	CNC	TRUE	
2700	5941	12334 p1	6	118786567	118953774	167207	13 dup	1	≤ 0.1% mo.to.both	C6orf204	FALSE	CSNK1G3 [frameshift], PDS5A [nonsense]
2700	5942	12334 s1	6	118786567	119215686	429119	14 dup	1	≤ 0.1% mo.to.both	C6orf204	FALSE	
2703	6010	13487 p1	6	123573556	123964483	12927	5 del	3	≤ 0.1% mo.to.both	TRDN	FALSE	CYP4Z1 [missense]
2712	6021	13513 p1	6	139206642	139208967	325	2 del	1	≤ 0.1% mo.to.both	ECTZL	FALSE	ZMYND11 [splice]
2712	6022	13513 s1	6	139206642	139208967	325	2 del	1	≤ 0.1% mo.to.both	ECTZL	FALSE	FGD5 [missense]
2713	5988	13183 p1	6	142396784	142400040	3256	2 del	1	≤ 0.1% mo.to.both	NMBR	TRUE	BCI11A [frameshift], CNOT6 [missense]
2713	5990	13183 s1	6	142396784	142400040	3256	2 del	1	≤ 0.1% mo.to.both	NMBR	TRUE	
2715	5982	11581 p1	6	146870599	146875741	5142	2 del	3	≤ 0.5% fa.to.p1	RAB32	TRUE	
2715	5982	12655 s1	6	146870599	146875741	5142	2 del	3	≤ 0.5% mo.to.s1	RAB32	TRUE	CD151 [missense]
2715	2677	13739 p1	6	146870599	146875741	5142	2 del	3	≤ 0.5% mo.to.both	RAB32	TRUE	COL11A1 [missense]
2715	2678	13739 s1	6	146870599	146875741	5142	2 del	3	≤ 0.5% mo.to.both	RAB32	FALSE	
2717	2608	11509 s1	6	15108973	151153341	63568	13 del	1	≤ 0.1% mo.to.s1	PLEKHG1	TRUE	
2718	6003	13396 p1	6	15108973	151153341	3918	2 del	1	≤ 0.1% fa.to.both	C6orf97	TRUE	
2718	6004	13396 p1	6	151865706	151869624	3918	2 del	1	≤ 0.1% fa.to.both	C6orf97	TRUE	CDC123 [missense]
2722	5998	13327 s1	6	160575829	160577106	1277	2 del	1	≤ 0.1% mo.to.s1	SLC22A1	TRUE	
2723	2655	12650 s1	6	162683556	162664505	180949	2 dup	2	≤ 0.1% mo.to.s1	PARK2	FALSE	
2726	2643	12317 p1	6	168187925	168227435	39510	3 dup	3	≤ 0.1% mo.to.s1	C6orf124	TRUE	
2730	2610	11519 p1	6	169617915	169646376	28461	19 dup	1	≤ 0.1% mo.to.p1	THBS2	FALSE	SMC3 [missense], SUV420H1 [missense]
2732	6016	13504 p1	6	170844307	170892835	48528	18 dup	1	≤ 0.1% mo.to.both	TBP, PDCD2, PSMB1	TRUE	TBP
2732	6017	13504 p1	6	170844307	170892835	48528	18 dup	1	≤ 0.1% mo.to.both	TBP, PDCD2, PSMB1	TRUE	TBP
2739	6031	12334 p1	7	4050597	4119216	68619	8 del	1	≤ 0.1% mo.to.p1	SDK1	FALSE	SDK1
2758	488	13335 s1	7	7338263	7495743	97480	20 del	1	≤ 0.1% mo.to.s1	COL28A1	FALSE	

3002	2989	13825 s1	8	144391610	144400256	8646	6 del	1	≤ 0.1% mo to s1	TOPIMT	FALSE		
3004	2964	11252 s1	8	144511514	144548018	36504	6 dup	1	≤ 0.1% fa to both	MAFA, ZC3H3	FALSE	NT5E [missense]	
3019	2963	12534 p1	8	145947028	146033780	86752	18 dup	1	≤ 0.1% mo to p1	ZNF251, ZNF34, ZNF517, RPL8	TRUE	MTMR9 [missense], MMP8 [missense]	
3020	551	13815 p1	8	146029025	146029623	598	2 del	1	≤ 0.5% mo to both	ZNF517	TRUE		
3020	553	13815 s1	8	146029025	146029623	598	2 del	1	≤ 0.5% mo to both	ZNF517	TRUE		
3023	591	12106 p1	9	172080	340321	186241	19 dup	4	≤ 0.5% fa to both	DOCK8, C9orf66, CBWD1	TRUE	DOCK8	
3023	6239	13366 s1	9	214508	312166	97658	6 dup	4	≤ 0.1% fa to s1	DOCK8, C9orf66	FALSE	DOCK8	
3022	583	12106 s1	9	214508	340321	125913	14 dup	4	≤ 0.5% fa to both	DOCK8, C9orf66	TRUE	DOCK8	
3023	3046	11316 p1	9	271626	407069	135443	27 dup	4	≤ 0.5% mo to p1	DOCK8	TRUE	DOCK8	
3023	3048	11363 p1	9	286460	407069	130609	26 dup	4	≤ 0.5% mo to p1	DOCK8	TRUE	DOCK8	
3026	6194	12665 p1	9	368017	677009	308992	35 dup	1	≤ 0.5% mo to p1	DOCK8, KANK1	TRUE	DOCK8, KANK1	
3031	597	12161 s1	9	5968018	6015607	47589	4 dup	1	≤ 0.1% mo to s1	RANBP8, KIAA2026	FALSE		
3035	3103	12512 s1	9	12693996	12760026	82030	8 del	1	≤ 0.1% mo to s1	TYRP1, C9orf150	FALSE		
3036	6208	13094 p1	9	14675550	14866875	11425	4 dup	4	≤ 0.1% fa to p1	FREMT1	FALSE	WDFY3 [nonsense]	
3037	6187	12480 p1	9	15564086	15623411	59325	6 del	1	≤ 0.1% mo to p1	C9orf93	TRUE		
3044	633	13926 s1	9	27217685	27229230	11545	5 dup	1	≤ 0.1% fa to s1	TEK	TRUE		
3047	6176	12252 p1	9	33135186	33261167	125981	11 dup	1	≤ 0.5% fa to both	BGALT1, None, SPINK4, BAG1	FALSE	BGALT1	
3047	6214	13129 s1	9	33166755	33261167	94412	10 dup	1	≤ 0.5% fa to both	BGALT1, None, SPINK4, BAG1	FALSE	BGALT1	
3050	8504	13129 s1	9	33888878	33948585	61707	29 dup	1	≤ 0.1% mo to both	UBAP2, UBE2R2	TRUE		
3050	8504	13129 p1	9	33888878	33948585	61707	29 dup	1	≤ 0.1% mo to both	UBAP2, UBE2R2	TRUE		
3052	6251	13589 p1	9	34286614	34290387	3773	2 del	2	≤ 0.1% mo to both	KIF24	TRUE	WDR5 [missense]	
3054	606	12741 p1	9	35228011	35237823	9812	4 del	1	≤ 0.1% mo to both	UNC13B	TRUE	UNC13B	
3054	607	12741 s1	9	35228011	35237823	9812	4 del	1	≤ 0.1% mo to both	UNC13B	TRUE	UNC13B	
3057	3087	12308 p1	9	35662942	35664489	1547	4 del	8	≤ 1% mo to p1	C9orf100	TRUE		
3057	3108	12534 p1	9	35662942	35664489	1547	4 del	8	≤ 1% fa to p1	C9orf100	TRUE	MTMR9 [missense], MMP8 [missense]	
3057	599	12578 p1	9	35662942	35664489	1547	4 del	8	≤ 1% fa to p1	C9orf100	TRUE		
3057	6191	12637 p1	9	35662942	35664489	1547	4 del	8	≤ 1% fa to p1	C9orf100	FALSE	RNPEPL1 [missense], RHOT2 [missense]	
3057	6200	12829 p1	9	35662942	35664489	1547	4 del	8	≤ 1% mo to both	C9orf100	FALSE		
3057	6201	12829 s1	9	35662942	35664489	1547	4 del	8	≤ 1% mo to both	C9orf100	FALSE		
3057	8507	13099 s1	9	35664004	35664489	485	2 del	8	≤ 1% fa to s1	C9orf100	TRUE		
3073	572	11479 s1	9	89923343	89968114	44771	21 dup	1	≤ 0.1% fa to s1	ZCCHC6	TRUE		
3075	6183	12334 p1	9	94984801	94985771	970	2 del	1	≤ 0.5% mo to p1	JARS	FALSE		
3087	6291	13296 p1	9	115407929	115580095	172166	8 dup	2	≤ 1% mo to p1	SNX30, KIAA1958, C9orf80	TRUE		
3088	3098	12507 p1	9	115811641	115912152	51	2 del	2	≤ 0.5% mo to both	ZFP37	FALSE		
3088	3098	12507 s1	9	115811641	115912152	51	2 del	2	≤ 0.5% mo to both	ZFP37	FALSE		
3088	623	13793 p1	9	115811641	115818668	7327	3 del	2	≤ 0.5% mo to p1	ZFP37	TRUE	PCDH4 [missense]	
3090	618	13629 p1	9	11612786	116132422	9636	4 dup	2	≤ 0.5% mo to p1	BSPRY	TRUE		
3090	621	13629 s1	9	116124648	116132422	9636	4 dup	2	≤ 0.5% fa to both	BSPRY	TRUE		
3093	3057	11437 s1	9	123898063	123921297	23214	17 del	1	≤ 0.1% fa to s1	CNTRL	FALSE		
3097	6232	13296 p1	9	125562401	125589066	26665	4 dup	1	≤ 0.1% mo to both	ORH1K1, PDCL	TRUE		
3097	6233	13296 s1	9	125562401	125589066	26665	4 dup	1	≤ 0.1% mo to both	ORH1K1, PDCL	TRUE	CGNL1 [missense], DENND5B [missense], LRRCC40 [missense]	
3119	6240	13418 p1	9	136259417	136262421	3004	3 del	1	≤ 0.1% mo to p1	C9orf96	FALSE		
3127	3050	11356 p1	9	139634401	139651044	16643	16 dup	1	≤ 0.1% mo to p1	LCN8, LCN10, LCN8	TRUE		
3129	6216	13129 s1	9	139887376	139890995	3619	11 dup	1	≤ 0.1% fa to s1	CUL3, C9orf142	TRUE		
3135	6182	12334 p1	9	140126154	140127856	1702	6 dup	1	≤ 0.1% fa to p1	SLC34A3	FALSE		
3138	3126	13808 p1	9	140458885	140458606	721	3 del	1	≤ 0.1% mo to p1	WDR85	TRUE		
3139	8279	12647 s1	9	140622800	140733293	110493	25 dup	1	≤ 0.1% mo to both	EHMT1, MIR602	TRUE	EHMT1	
3139	3113	12647 p1	9	140646782	140733293	86511	22 del	1	≤ 0.1% mo to both	EHMT1, MIR602	TRUE	EHMT1	
3141	3129	13825 s1	9	141000150	141124256	124106	15 dup	3	≤ 0.1% fa to s1	CACNA1B	FALSE	CACNA1B	
323	662	11895 s1	10	225952	532470	306518	50 dup	1	≤ 0.1% fa to s1	DIP2C, ZMYND11	FALSE		
324	3213	11611 p1	10	858865	910210	51345	15 dup	1	≤ 0.1% mo to both	LARP4B	TRUE		
324	3214	11611 s1	10	858865	910210	51345	15 dup	1	≤ 0.1% mo to both	LARP4B	TRUE		
325	8330	13144 s1	10	3124579	3178030	53451	20 dup	1	≤ 0.1% mo to s1	PKPK	TRUE		
326	3215	11622 p1	10	3175394	3176774	1380	2 del	1	≤ 0.5% fa to p1	PKPK	TRUE	CACNA1H [frameshift]	
330	6333	13162 p1	10	5203384	5260723	57339	12 del	1	≤ 0.1% mo to p1	AKR1C4, AKR1C1	TRUE		
340	3249	12317 p1	10	18276407	18292287	15880	6 dup	1	≤ 0.1% fa to both	SLC39A12	FALSE	RIMS1 [frameshift]	
340	3250	12317 s1	10	18289595	18331762	42167	3 dup	1	≤ 0.1% fa to both	SLC39A12	FALSE		
344	3150	11094 p1	10	27687222	27700858	13636	3 del	3	≤ 1% fa to p1	PICHD3	TRUE	CHST2 [missense]	
344	3245	12303 p1	10	27687222	27700858	13636	3 del	3	≤ 1% mo to p1	PICHD3	TRUE	CDC14 [missense]	
346	6331	12313 s1	10	28940408	28945596	5188	3 del	2	≤ 0.1% fa to s1	PICHD3	TRUE	NCAPD2 [missense]	
346	684	13793 s1	10	28940408	28945596	5188	3 del	2	≤ 0.1% mo to s1	PICHD3	TRUE	PGM3 [missense]	
348	679	12741 p1	10	35299238	35303322	6084	4 dup	1	≤ 0.1% mo to p1		TRUE		
370	6362	13465 p1	10	48370532	51980838	3520306	320 dup	1	≤ 0.1% fa to both	CUL2	TRUE	EH2 [missense]	
										PAR6, AGAP7, ERC06, CHAT, VSTM4, NCOA4, GDF2, ARHGAP22, C10orf71, MSMB, FAM21A, MAPK8, OGDHL, C10orf53, None, DRGX, FRMPD2P1, RBP3, FRMPD2, ZNF488, C10orf128, WDFY4, PTPN20B, GDF10, TIMM23			
370	6365	13465 s1	10	48370532	51980838	3520306	320 dup	1	≤ 0.1% fa to both	PAR6, AGAP7, ERC06, CHAT, VSTM4, NCOA4, GDF2, ARHGAP22, C10orf71, MSMB, FAM21A, MAPK8, OGDHL, C10orf53, None, DRGX, FRMPD2P1, RBP3, FRMPD2, ZNF488, C10orf128, WDFY4, PTPN20B, GDF10, TIMM23	TRUE	CHAT, ERC06	
374	656	11696 p1	10	54527896	54531395	3499	4 del	5	≤ 2% mo to p1		FALSE		
374	8347	12997 p1	10	54527896	54531395	3499	4 del	5	≤ 2% fa to both	MBL2	TRUE	INCENP [missense]	

374	8348	12997 s1	10	54527896	54531395	3499	4 del	5	≤ 2% fa to both	MBL2	TRUE	PCDH15
376	673	12304 s1	10	56287571	56424022	136451	4 del	1	≤ 0.1% mo to s1	PCDH15	FALSE	PCDH15
379	8349	12307 s1	10	64425888	64430061	4173	2 del	1	≤ 0.5% fa to s1	ZNF365	FALSE	ZNF365
380	658	11711 p1	10	62830374	68381542	101168	2 del	1	≤ 0.1% fa to both	CTNNA3	TRUE	CTNNA3
380	659	11711 p1	10	62830374	68381542	101168	2 del	1	≤ 0.1% fa to both	CTNNA3	TRUE	CTNNA3
382	3174	11267 p1	10	72604229	72645686	41457	16 del	1	≤ 0.1% mo to both	SGP11, PCBD1	TRUE	SGP11, PCBD1
382	3175	11267 p1	10	72604229	72645686	41457	16 del	1	≤ 0.1% mo to both	SGP11, PCBD1	TRUE	SGP11, PCBD1
384	3178	11298 p1	10	75010609	75018174	5565	5 dup	1	≤ 0.5% mo to p1	C10orf103	TRUE	SLC6A13 [missense]
395	6270	12510 p1	10	90703553	90707143	3590	2 dup	4	≤ 2% fa to both	ACTA2	TRUE	ACTA2
395	6271	12510 p1	10	90703553	90707143	3590	2 dup	4	≤ 2% fa to both	ACTA2	TRUE	ACTA2
395	6281	12518 p1	10	90703553	90707143	3590	2 dup	4	≤ 2% fa to both	ACTA2	FALSE	ACTA2
395	6286	12518 p1	10	90703553	90707143	3590	2 dup	4	≤ 2% fa to both	ACTA2	FALSE	ACTA2
394	3194	11474 p1	10	95371155	95357560	20425	6 dup	1	≤ 0.1% mo to both	LG11	FALSE	LG11
403	3195	11474 s1	10	95371155	95357560	20425	6 dup	1	≤ 0.1% mo to both	LG11	FALSE	ZNF10 [missense]
404	3167	11180 p1	10	96466540	96495201	28661	5 del	1	≤ 0.1% mo to both	CYP2C18	FALSE	CYP2C18
404	3168	11180 p1	10	96466540	96495201	28661	5 del	1	≤ 0.1% mo to both	CYP2C18	FALSE	CYP2C18
400	670	12285 s1	10	101594136	101596047	1911	2 dup	5	≤ 2% mo to both	ABCC2	FALSE	ABCC2
400	671	12285 s1	10	101594136	101596047	1911	2 dup	5	≤ 2% mo to both	ABCC2	FALSE	ABCC2
400	3253	12383 s1	10	101594136	101596047	1911	2 dup	5	≤ 2% mo to both	ABCC2	FALSE	ABCC2
401	652	11629 p1	10	103290993	103310617	19624	8 dup	1	≤ 0.5% fa to both	BTRC	TRUE	BTRC
401	653	11629 p1	10	103290993	103310617	19624	8 dup	1	≤ 0.5% fa to both	BTRC	TRUE	BTRC
396	6295	12644 p1	10	114427976	114496847	68871	3 dup	1	≤ 0.1% mo to both	VT11A	TRUE	VT11A
396	6296	12644 s1	10	114427976	114496847	68871	3 dup	1	≤ 0.1% mo to both	VT11A	TRUE	VT11A
402	3201	11519 p1	10	122218817	122348014	132197	7 del	1	≤ 0.1% mo to p1	PPAPDC1A	TRUE	PPAPDC1A
413	8359	13512 p1	10	124672277	124697675	25398	3 del	1	≤ 0.1% mo to p1	C10orf88, FAM24A	TRUE	C10orf88, FAM24A
414	6375	13512 p1	10	124800724	124924572	123848	19 dup	1	≤ 0.1% mo to p1	HMX3, HMX2, BUB3, ACADSB	TRUE	HMX3, HMX2, BUB3, ACADSB
416	3209	11561 p1	10	134734122	134793304	59182	20 dup	2	≤ 0.5% mo to both	C10orf93	FALSE	C10orf93
416	3210	11561 p1	10	134734122	134793304	59182	20 dup	2	≤ 0.5% mo to both	C10orf93	FALSE	C10orf93
397	3291	14110 s1	10	135032334	135032594	260	2 dup	2	≤ 0.1% mo to both	KNDCC1	FALSE	KNDCC1
420	8356	13162 s1	10	135168873	135179599	10726	9 dup	1	≤ 0.1% mo to both	ECHS1, C10orf125	TRUE	ECHS1, C10orf125
420	8356	13162 s1	10	135168873	135179599	10726	9 dup	1	≤ 0.1% mo to both	ECHS1, C10orf125	TRUE	ECHS1, C10orf125
449	3356	11810 p1	11	193099	249057	55958	36 dup	1	≤ 0.5% fa to both	OPF3, SIRT3, RIC8A, PSMD13, SCGB1C1, BET1L	FALSE	OPF3, SIRT3, RIC8A, PSMD13, SCGB1C1, BET1L
449	3358	11810 s1	11	193099	249057	55958	36 dup	1	≤ 0.5% fa to both	OPF3, SIRT3, RIC8A, PSMD13, SCGB1C1, BET1L	FALSE	OPF3, SIRT3, RIC8A, PSMD13, SCGB1C1, BET1L
450	3350	11786 p1	11	244040	244469	429	2 dup	10	≤ 2% mo to p1	PSMD13	FALSE	PSMD13
464	3321	11180 p1	11	4869466	4904113	34647	2 del	1	≤ 0.5% fa to both	OR51T1, OR51S1	FALSE	OR51T1, OR51S1
464	3322	11180 p1	11	4869466	4904113	34647	2 del	1	≤ 0.5% fa to both	OR51T1, OR51S1	FALSE	OR51T1, OR51S1
465	6393	12683 s1	11	5009441	5021160	11719	7 del	1	≤ 0.1% mo to s1	OR51L1, MMP26	FALSE	OR51L1, MMP26
467	697	11569 p1	11	5862185	5878932	16747	2 del	4	≤ 2% fa to both	OR52E6, OR52E8	TRUE	OR52E6, OR52E8
467	698	11569 s1	11	5862185	5878932	16747	2 del	4	≤ 2% fa to both	OR52E6, OR52E8	TRUE	OR52E6, OR52E8
473	721	12581 s1	11	794264	7961067	11603	2 del	1	≤ 0.1% fa to s1	OR10A3, OR10A6	TRUE	OR10A3, OR10A6
476	714	11964 p1	11	1480588	14902314	21726	7 dup	1	≤ 0.1% fa to p1	CYP2R1, PDE3B	TRUE	CYP2R1, PDE3B
479	3345	11667 p1	11	18273363	1829842	2479	4 del	1	≤ 0.1% mo to both	IGSF22	TRUE	IGSF22
479	3347	11667 p1	11	18273363	1829842	2479	4 del	1	≤ 0.1% mo to both	IGSF22	TRUE	IGSF22
480	3408	12512 s1	11	20805225	20869299	64074	2 del	1	≤ 0.1% fa to s1	NELL1	FALSE	NELL1
481	6387	12523 s1	11	22215038	22833562	618524	42 del	1	≤ 0.1% mo to both	ANO5, SLC17A6, GAS2, FANCF	FALSE	FANCF, SLC17A6
481	6386	12523 p1	11	22232809	22833562	600753	40 del	1	≤ 0.1% mo to both	ANO5, SLC17A6, GAS2, FANCF	FALSE	FANCF, SLC17A6
482	744	13926 p1	11	31312193	31541638	229445	16 del	1	≤ 0.1% mo to both	IMMP1L, DNAJC24, DDCD1, ELP4	TRUE	IMMP1L, DNAJC24, DDCD1, ELP4
482	745	13926 s1	11	31312193	31541638	229445	16 del	1	≤ 0.1% mo to both	IMMP1L, DNAJC24, DDCD1, ELP4	TRUE	IMMP1L, DNAJC24, DDCD1, ELP4
483	6436	13296 p1	11	32608557	32623945	15388	10 dup	1	≤ 1% mo to p1	EIF3M	TRUE	EIF3M
484	728	12810 p1	11	32897110	32781789	84679	8 del	1	≤ 0.1% mo to p1	CDCD73	TRUE	CDCD73
485	6383	12445 s1	11	32987845	33113886	126041	30 dup	1	≤ 0.1% fa to s1	TCPI1L1, QSER1, DEPDC7, CSTF3	TRUE	TCPI1L1, QSER1, DEPDC7, CSTF3
490	3439	13843 p1	11	43772460	43775671	3211	2 del	1	≤ 1% fa to p1	HSD17B12	TRUE	HSD17B12
498	6411	13097 s1	11	55432642	55595630	162988	6 del	1	≤ 1% mo to s1	OR5L1, OR5L2, OR5D18, OR5D13, OR4C6, OR5D14	FALSE	OR5L1, OR5L2, OR5D18, OR5D13, OR4C6, OR5D14
501	3327	11219 p1	11	57256388	57327905	71517	25 dup	1	≤ 0.1% fa to p1	UBE2L6, Nore, TIMM10, SLC43A1, SMTNL1	FALSE	UBE2L6, Nore, TIMM10, SLC43A1, SMTNL1
500	3393	12303 p1	11	5924902	59283330	38428	3 dup	3	≤ 0.1% mo to p1	OR4D9, OR4D10, OR4D11	TRUE	OR4D9, OR4D10, OR4D11
503	6402	12936 s1	11	59620447	59623531	3084	4 del	1	≤ 0.5% mo to both	TCN1	FALSE	TCN1
503	6401	12936 p1	11	59620675	59622308	1633	2 del	1	≤ 0.5% mo to both	TCN1	FALSE	TCN1
511	6379	12334 s1	11	62600423	62807042	6619	11 dup	1	≤ 0.5% mo to s1	WDR74	FALSE	WDR74
512	710	11798 s1	11	63057637	63059115	1478	2 dup	1	≤ 0.5% mo to s1	SLC22A10	TRUE	SLC22A10
514	3451	14110 s1	11	64569050	64569229	179	2 dup	1	≤ 0.1% fa to both	MAR4K2	FALSE	MAR4K2
529	6415	13148 s1	11	70954738	70956547	1809	2 del	1	≤ 0.1% mo to both	GDPD4	TRUE	GDPD4
529	6415	13148 s1	11	70954738	70956547	1809	2 del	1	≤ 0.1% mo to both	GDPD4	TRUE	GDPD4
534	701	11610 p1	11	85429632	85468768	38936	10 del	1	≤ 0.5% mo to both	SYTL2	FALSE	SYTL2
534	702	11610 s1	11	85429632	85468768	38936	10 del	1	≤ 0.5% mo to both	SYTL2	FALSE	SYTL2
549	7715	11115 s1	11	100558409	100831720	273311	14 dup	8	≤ 0.1% fa to s1		TRUE	
549	7466	11229 p1	11	100558409	100895532	301123	24 dup	8	≤ 0.1% fa to p1		FALSE	
550	6461	13502 p1	11	102708028	102709441	1413	2 del	1	≤ 0.1% mo to both	MMP3	TRUE	MMP3
550	6462	13502 s1	11	102708028	102709441	1413	2 del	1	≤ 0.1% mo to both	MMP3	TRUE	MMP3
554	3346	11667 s1	11	108256646	108264105	7459	2 dup	3	≤ 0.5% mo to both	C11orf65	TRUE	C11orf65
554	7717	11667 p1	11	108256646	108264105	7459	2 dup	3	≤ 0.5% mo to both	C11orf65	TRUE	C11orf65
554	7467	11773 s1	11	108256646	108264105	7459	2 dup	3	≤ 0.5% mo to s1	C11orf65	TRUE	C11orf65
554	3433	13808 s1	11	108256646	108264105	7459	2 dup	3	≤ 0.5% mo to s1	C11orf65	TRUE	C11orf65
558	725	12630 s1	11	112046324	112084584	38260	9 del	2	≤ 0.1% fa to s1	BOO2	TRUE	BOO2
559	6439	13307 s1	11	113669681	113725057	55096	24 dup	1	≤ 1% mo to s1	SPY28	TRUE	SPY28
571	3357	11810 p1	11	124508437	124509763	1326	2 del	1	≤ 0.1% mo to p1	SLAE	FALSE	SLAE
576	3430	13730 p1	11	134151270	134215023	63753	21 del	2	≤ 0.5% mo to p1	GLB1L2, GLB1L3	FALSE	DICER1 [missense]

576	6429	13239 s1	11	134162025	134214349	52324	16 dup	2	≤ 0.1% fa.to s1	GLB1L2, GLB1L3	FALSE	
575	6406	12851 p1	11	134177017	134257553	80536	34 dup	1	≤ 0.1% fa.to both	GLB1L2, GLB1L3, B3GAT1	TRUE	SLC25A29 [missense], HMCXB3 [missense], UBE3C [missense]
575	6407	12851 s1	11	134177017	134257553	80536	34 dup	1	≤ 0.1% fa.to both	GLB1L2, GLB1L3, B3GAT1	TRUE	B3GAT1
579	7722	11282 p1	12	4651049	4668159	17110	8 dup	3	≤ 0.5% mo.to both	RAD51AP1	TRUE	B3GAT1
579	7723	11519 p1	12	4651049	4668159	17110	8 dup	3	≤ 0.5% fa.to p1	RAD51AP1	TRUE	SMC3 [missense], SUV420H1 [missense]
579	3495	11282 s1	12	4655474	4668159	12685	6 dup	3	≤ 0.5% mo.to both	RAD51AP1	TRUE	
579	3621	13809 p1	12	4655474	4668159	12685	6 dup	3	≤ 0.5% fa.to p1	RAD51AP1	TRUE	
596	5957	12802 s1	12	8609707	8689862	81155	14 del	1	≤ 0.1% mo.to both	CLEC4E, CLEC4D, CLEC6A	TRUE	
596	7730	12802 p1	12	8609707	8689862	81155	14 del	1	≤ 0.1% mo.to both	CLEC4E, CLEC4D, CLEC6A	TRUE	
602	3594	12736 p1	12	10337928	10342739	4811	3 dup	1	≤ 0.1% mo.to both	C12orf69	TRUE	
602	7731	13125 p1	12	10337928	10342739	369812	40 del	2	≤ 0.1% mo.to both	C12orf69	TRUE	
614	6513	13125 s1	12	21007961	21377773	369812	40 del	2	≤ 0.1% mo.to both	LS1-3MT12, SLC01B1, SLC01B3	TRUE	RAD21L1 [nonsense]
614	6514	13125 p1	12	21007961	21377773	369812	40 del	2	≤ 0.1% mo.to both	LS1-3MT12, SLC01B1, SLC01B3	TRUE	
619	8313	12997 p1	12	25264705	25267804	3099	2 del	1	≤ 0.1% fa.to both	CASO1	TRUE	
619	8314	12997 s1	12	25264705	25267804	3099	2 del	1	≤ 0.1% fa.to both	CASO1	TRUE	
628	3618	13739 s1	12	33532766	38712280	5179494	11 dup	2	≤ 0.5% mo.to s1	ALG10, SYT10, ALG10B	TRUE	COL11A1 [missense]
640	3463	11090 p1	12	49888983	49891056	2073	5 dup	5	≤ 2% mo.to p1	PRPH	TRUE	
640	6506	12836 s1	12	49888983	49891056	2073	5 dup	5	≤ 2% fa.to s1	PRPH	FALSE	
642	6494	12719 s1	12	50475327	50484373	9046	9 dup	1	≤ 0.1% mo.to s1	ACCN2, SMARCD1	TRUE	
643	3540	11828 p1	12	51203238	51213562	10324	4 dup	1	≤ 0.1% mo.to p1	ATF1	TRUE	
647	3473	11114 p1	12	52774863	52779369	4506	6 dup	1	≤ 0.1% mo.to both	KRT84	TRUE	SCN2A [nonsense]
647	7739	11114 s1	12	52774863	52779369	4506	6 dup	1	≤ 0.1% mo.to both	KRT84	TRUE	DZANK1 [missense]
658	3479	11180 s1	12	56619862	56620203	241	2 dup	2	≤ 0.1% fa.to s1	OBFC2B	FALSE	
663	8395	12723 p1	12	57345812	57351246	5434	4 dup	4	≤ 1% fa.to both	RDH16	TRUE	
663	8396	12723 s1	12	57345812	57351246	5434	4 dup	4	≤ 1% fa.to both	RDH16	TRUE	SERAC1 [missense]
663	807	13835 p1	12	57345812	57348948	3136	3 del	4	≤ 1% mo.to both	RDH16	TRUE	
663	808	13835 s1	12	57345812	57351246	5434	4 del	4	≤ 1% mo.to both	RDH16	TRUE	
671	6474	12334 p1	12	69742188	69744052	1864	2 del	1	≤ 0.1% fa.to p1	LYZ	FALSE	
672	7751	12763 p1	12	75436092	75905377	469285	49 dup	1	≤ 0.1% mo.to p1	KCNQ2, GLIPR1, KRR1, CAPS2, GLIPR1L2, GLIPR1L1	TRUE	
674	6492	12716 s1	12	80190561	80211298	20737	9 dup	1	≤ 0.1% mo.to s1	PPPIR12A	TRUE	
675	3599	12869 p1	12	80632630	80730784	98154	31 del	2	≤ 0.5% mo.to both	OTOGL	TRUE	PRGP [missense]
675	3601	12869 s1	12	80632630	80730335	97705	30 del	2	≤ 0.5% mo.to both	OTOGL	TRUE	
675	802	13798 p1	12	80632630	80730784	98154	32 del	2	≤ 0.5% fa.to p1	OTOGL	FALSE	
680	8398	13215 p1	12	96704829	96707232	2403	2 dup	2	≤ 0.1% mo.to p1	CDK17	TRUE	
684	3629	13832 s1	12	104171666	104174142	2536	2 del	1	≤ 0.1% fa.to s1	NTSDC3	TRUE	
685	774	11895 s1	12	109290761	10939251	2470	3 del	1	≤ 0.1% mo.to s1	DAO	FALSE	DAO
693	786	12361 p1	12	112162446	112303664	126538	28 dup	2	≤ 0.5% mo.to p1	ACAD10, MAPKAPK5, C12orf47, ALDH2	TRUE	
697	3607	13322 p1	12	114296595	114404032	107437	22 dup	1	≤ 0.1% mo.to p1	RBM19	TRUE	
703	3489	11241 p1	12	120875929	12088632	8703	7 dup	2	≤ 0.5% mo.to p1	GATC, COX6A1, TRIAP1	TRUE	C18orf26 [missense]
703	6505	12836 p1	12	120875929	12088632	8703	7 dup	2	≤ 0.5% mo.to p1	GATC, COX6A1, TRIAP1	FALSE	TBC1D2B [missense]
716	7488	12285 p1	12	133533328	132686946	3618	7 dup	1	≤ 0.1% fa.to p1	NOC4L	FALSE	EF4G1 [missense]
720	3483	11196 p1	12	133618053	133781116	163063	16 dup	1	≤ 0.5% mo.to p1	ZNF140, ZNF268, ZNF10, ZNF84	TRUE	
724	3678	11509 s1	13	20425494	20426320	826	2 dup	5	≤ 1% fa.to s1	ZMYM5	FALSE	FAM45A [missense]
724	7476	12011 p1	13	20425494	20426320	826	2 dup	5	≤ 1% fa.to both	ZMYM5	TRUE	
724	7477	12011 s1	13	20425494	20426320	826	2 dup	5	≤ 1% fa.to both	ZMYM5	TRUE	NLRX1 [missense], ADAM33 [nonsense]
724	3718	12561 p1	13	20425494	20426320	826	2 dup	5	≤ 1% fa.to p1	ZMYM5	FALSE	
724	3739	13322 p1	13	20425494	20426320	826	2 dup	5	≤ 1% fa.to p1	ZMYM5	TRUE	
730	3683	11676 p1	13	21032318	21311944	8726	3 dup	1	≤ 0.1% mo.to p1	NSAMT2	TRUE	
736	3656	11154 s1	13	25276072	25285660	9588	10 dup	3	≤ 0.1% fa.to s1	ATP12A	TRUE	MTHFS [frameshift], EFCAB5 [frameshift]
742	6576	13590 p1	13	33306237	34405491	1099254	43 dup	1	≤ 0.1% mo.to both	KL-STARD13, RFC3, PDG5B	TRUE	
742	6577	13590 s1	13	33309308	34540262	1230954	45 dup	1	≤ 0.1% mo.to both	KL-STARD13, RFC3, PDG5B	TRUE	
754	3749	13808 p1	13	49084764	49086310	1546	3 del	4	≤ 0.1% fa.to both	RCBTB2	TRUE	
754	3750	13808 s1	13	49084764	49086310	1546	3 del	4	≤ 0.1% fa.to both	RCBTB2	TRUE	
755	3666	11298 s1	13	50123593	50134220	10627	5 dup	1	≤ 0.5% mo.to s1	RCBTB1	TRUE	RCBTB1
759	3747	13774 p1	13	76111799	76123988	12199	3 dup	9	≤ 2% mo.to p1	COMMD6, UCHL3	TRUE	
764	3671	11412 p1	13	96508411	96515968	7557	4 del	1	≤ 0.5% fa.to both	UGT2	FALSE	DNAH11 [missense]
764	3672	11412 s1	13	96508411	96515968	7557	4 del	1	≤ 0.5% fa.to both	UGT2	FALSE	C16orf82 [missense]
779	817	11190 s1	13	11176011	11191920	143809	21 dup	1	≤ 0.1% mo.to s1	RAB20, CARS2, Nones, CARKD	TRUE	
785	3687	11808 p1	13	113979925	113984006	431	2 dup	1	≤ 0.1% mo.to p1	GRTP1	FALSE	
786	6555	12829 s1	13	114138154	114175048	36894	8 dup	1	≤ 0.1% mo.to s1	DCUJH1D2, TMCO3	FALSE	
788	3732	13063 p1	13	114514708	114535461	20753	7 dup	2	≤ 0.5% mo.to both	GAS6, FAM70B	FALSE	TPK1 [missense]
788	3733	13063 s1	13	114514708	114535461	16976	6 dup	2	≤ 0.5% mo.to both	GAS6, FAM70B	FALSE	
790	3660	11196 p1	13	115002273	115091756	69463	26 dup	1	≤ 0.5% mo.to both	UPF3A, CDC16, ZNF828	TRUE	UPF3A
790	3661	11196 s1	13	115007595	115091756	84161	23 dup	1	≤ 0.5% mo.to both	UPF3A, CDC16, ZNF828	TRUE	UPF3A
798	6586	12460 p1	14	21215739	21424416	208677	6 dup	4	≤ 0.1% mo.to p1	RNASE1, RNASE3, RNASE2, RNASE6, EDM9A, EDM3B	TRUE	
794	3811	11766 s1	14	21238309	21270227	31918	3 dup	1	≤ 0.1% fa.to s1	RNASE1, EDM3B, RNASE6	FALSE	SYTL3 [nonsense]
795	7780	11094 p1	14	21359845	21621184	264339	73 dup	1	≤ 0.1% fa.to p1	RNASE3, RNASE7, RNASE13, RNASEB, ZNF219, OR5A1, TPPP2, NDR32, SLC39A2, METTL17, ARHGEF40	TRUE	CCDC14 [missense]
811	3788	11242 p1	14	22133296	22410148	276852	19 dup	1	≤ 0.1% fa.to p1	Nones, OR4E2	TRUE	
810	6604	13139 p1	14	22180235	22192868	12633	2 del	1	≤ 0.1% fa.to p1	Nones	FALSE	
809	3851	13912 s1	14	22891209	22928469	37260	8 del	5	≤ 0.1% mo.to both	Nones	TRUE	
845	6606	13153 p1	14	63757595	63860646	103051	10 dup	2	≤ 0.5% fa.to p1	RHOJ, GPHB5, PPP2R5E	TRUE	MAPK13 [missense]
846	3843	13195 p1	14	64006246	64066660	60414	2 dup	1	≤ 0.1% fa.to both	PPP2R5E, WDR89	TRUE	TSPYL5 [3n-non-frameshifting], SCAR2 [missense]

851	865	12304 p1	14	65016715	65019579	2864	3 dup	3	≤ 2% fa_to_p1	C14orf50	FALSE	STIL [missense], PSEN1 [missense], PHF19 [missense]
851	877	12810 p1	14	65016715	65019579	2864	3 dup	3	≤ 2% fa_to_p1	C14orf50	TRUE	EMID2 [missense]
854	8420	12582 p1	14	69919957	69969596	49639	10 dup	1	≤ 0.5% mo_to_p1	SLC39A9	TRUE	
856	888	13533 p1	14	74522286	74541720	19434	15 del	1	≤ 0.1% fa_to_both	ALDH6A1, C14orf45	FALSE	
856	891	13533 s1	14	74522286	74541720	19434	15 del	1	≤ 0.1% fa_to_both	ALDH6A1, C14orf45	FALSE	ENOX2 [missense]
860	6596	12687 p1	14	78374145	78392304	18159	3 del	1	≤ 0.1% mo_to_p1	ADCK1	TRUE	
863	3814	11810 p1	14	90767592	90784457	16875	6 dup	1	≤ 0.1% mo_to_p1	C14orf102	FALSE	
870	3782	11118 p1	14	99182528	99183611	1083	2 del	4	≤ 2% fa_to_p1	C14orf177	TRUE	
870	3795	11366 p1	14	99182528	99183611	1083	2 del	4	≤ 2% mo_to_p1	C14orf177	TRUE	NAPRT1 [splice], SY2B [missense]
870	879	13116 p1	14	99182528	99183611	1083	2 del	4	≤ 2% fa_to_both	C14orf177	TRUE	
870	881	13116 s1	14	99182528	99183611	1083	2 del	4	≤ 2% fa_to_both	C14orf177	TRUE	SRRM5 [missense]
870	904	14201 s1	14	99182528	99183611	1083	2 del	4	≤ 2% mo_to_s1	EML1	TRUE	
871	6590	11894 p1	14	100402365	100405664	3299	4 dup	3	≤ 0.5% mo_to_p1	EML1	FALSE	EML1
871	3826	12317 s1	14	100402365	100405664	3299	4 dup	3	≤ 0.5% fa_to_s1	EML1	FALSE	CHST2 [missense]
874	6593	12386 p1	14	105836177	105861009	24832	17 dup	1	≤ 0.1% fa_to_p1	PACS2	TRUE	
888	6675	13493 p1	15	20739496	23445762	2706266	97 del	6	≤ 0.5% fa_to_p1	None, CYFIP1, NIPA2, LOC727924, GOLGA6L1, OR4N4, POTE8, NIPA1, GOLGA8E, TUBGCP5	TRUE	NIPA2, NIPA1, CYFIP1, TUBGCP5
888	4002	13393 p1	15	22739533	23062319	322786	66 dup	6	≤ 0.5% fa_to_both	NIPA2, NIPA1, CYFIP1, GOLGA6L1, TUBGCP5	FALSE	NIPA2, NIPA1, CYFIP1, TUBGCP5
888	6640	12735 p1	15	22835915	23062319	226404	62 del	6	≤ 0.5% fa_to_p1	NIPA2, NIPA1, CYFIP1, TUBGCP5	TRUE	NIPA2, NIPA1, CYFIP1, TUBGCP5
888	4003	13393 s1	15	22835915	23062319	226404	62 dup	6	≤ 0.5% fa_to_both	NIPA2, NIPA1, CYFIP1, TUBGCP5	FALSE	NIPA2, NIPA1, CYFIP1, TUBGCP5
8437	13493 p1	13493 p1	15	22835915	23062319	226404	62 del	6	≤ 0.5% fa_to_p1	NIPA2, NIPA1, CYFIP1, TUBGCP5	TRUE	NIPA2, NIPA1, CYFIP1, TUBGCP5
888	6672	13487 s1	15	31356910	31369124	12214	6 del	4	≤ 0.5% fa_to_s1	TRPM1	FALSE	
897	8440	12933 p1	15	32323100	32404100	81000	3 dup	2	≤ 1% fa_to_both	CHRNA7	TRUE	CHRNA7
897	8441	12933 s1	15	32323100	32404100	81000	3 dup	2	≤ 1% fa_to_both	CHRNA7	TRUE	CHRNA7
905	936	13890 p1	15	36983895	37002162	18277	5 del	1	≤ 0.1% fa_to_s1	C15orf41	FALSE	
909	3954	12317 s1	15	40993261	41001314	8653	3 del	1	≤ 0.1% mo_to_s1	RAD51	FALSE	CHST2 [missense]
915	934	13629 s1	15	42967473	42147943	80470	75 dup	1	≤ 0.1% mo_to_s1	JMJD7, PLA2G4B, SFTBNS, MAPKBP1	TRUE	
918	4013	13823 p1	15	42807434	42836322	28888	5 dup	1	≤ 0.1% mo_to_p1	LRRCS1, SNAP23	FALSE	
920	6655	13148 p1	15	43577011	43585152	8141	5 del	1	≤ 0.1% fa_to_both	TGM7	TRUE	KIAA1244 [missense]
920	6656	13148 s1	15	43577011	43584295	7284	4 del	1	≤ 0.1% fa_to_both	TGM7	TRUE	POOLCE [frameshift]
921	6651	13018 p1	15	43827893	43844143	16250	9 dup	1	≤ 0.1% mo_to_p1	ADAM	TRUE	MCOLN3 [missense]
922	914	11479 p1	15	43969610	43701284	4684	5 dup	1	≤ 0.1% mo_to_p1	TF53BP1, TUBGCP4	TRUE	
929	6664	13216 s1	15	50731270	51535109	803839	113 dup	1	≤ 0.1% fa_to_s1	USP8, TNFAIP8L3, CYP19A1, USP90, AP4E1, SPPL2A, TRPM7	TRUE	AP4E1
932	6670	13330 p1	15	53991946	54025346	33400	12 del	2	≤ 0.1% mo_to_p1	WDR72	FALSE	
934	4015	13843 p1	15	55475512	55497903	22391	6 dup	2	≤ 0.1% mo_to_p1	RAB27A, RSL24D1	TRUE	AGK [missense]
936	3938	12295 p1	15	57730197	57754090	23893	7 dup	2	≤ 0.5% mo_to_both	CGNL1	FALSE	
936	6644	12837 p1	15	57730197	57754090	23893	7 dup	2	≤ 0.5% fa_to_p1	CGNL1	FALSE	SH3RF3 [missense]
936	3939	12295 s1	15	57732574	57746016	13442	5 dup	2	≤ 0.5% mo_to_both	CGNL1	FALSE	
963	906	11229 s1	15	76426604	76641077	214473	28 dup	1	≤ 0.1% fa_to_s1	ETFA, C15orf27, ISL2, SCAPER	FALSE	SNRNP200 [missense]
974	3878	11241 s1	15	85461754	85488431	26677	10 del	3	≤ 0.1% fa_to_s1	SLC28A1	TRUE	HRH2 [missense], GOLGA4 [missense]
974	6628	12420 p1	15	85461754	85488431	26677	10 del	3	≤ 0.1% mo_to_p1	SLC28A1	FALSE	
974	6633	12445 p1	15	85461754	85488431	26677	10 del	3	≤ 0.1% mo_to_p1	SLC28A1	TRUE	
976	3867	11092 p1	15	86028951	86129054	100103	7 dup	1	≤ 0.5% mo_to_p1	AKAP13	TRUE	
981	3977	12651 s1	15	91486141	91500955	14814	15 del	2	≤ 0.1% fa_to_s1	RCCD1, UNC45A	TRUE	
981	4005	13543 p1	15	91488121	91520001	31880	25 del	2	≤ 0.1% mo_to_p1	RCCD1, PRC1, UNC45A	TRUE	
983	6658	13197 s1	15	94857510	94888393	30983	6 dup	1	≤ 0.1% mo_to_s1	MCTP2	TRUE	GMPPA [3n-non-frameshifting]
984	911	11452 s1	15	101970191	101983665	13674	4 dup	1	≤ 0.1% mo_to_s1	PCSK6	TRUE	
985	3892	11356 s1	15	102224277	102501099	276822	15 dup	2	≤ 0.1% mo_to_s1	OR4F6, TARS12, OR4F4, OR4F15	TRUE	DLGAP3 [missense], STEAP3 [missense], DHX30 [missense]
985	3969	12375 p1	15	102224277	10261525	37248	10 dup	2	≤ 0.1% fa_to_p1	TARS12	TRUE	FAM208A [missense]
982	6741	12826 p1	16	454955	461578	6623	7 del	1	≤ 0.1% mo_to_both	DECIP2	TRUE	THOYE2 [frameshift]
992	6742	12826 s1	16	454955	461578	6623	7 del	1	≤ 0.1% mo_to_both	DECIP2	TRUE	
1015	6609	13398 s1	16	3014521	3100547	86026	53 dup	1	≤ 0.1% mo_to_both	HCFOT1R1, PAQR4, PKMYT1, CODC64B, CLDN9, MIMP25, KREMEN2, TNFRSF12A, THOC6, CLDN6, HCFOT1R1, PAQR4, PKMYT1, CODC64B, CLDN9, MIMP25, KREMEN2, TNFRSF12A, THOC6, CLDN6	TRUE	POGZ [frameshift], PYHIN1 [missense], TTN [missense]
1015	6608	13398 p1	16	3016325	3100547	84222	50 dup	1	≤ 0.1% mo_to_both	HCFOT1R1, PAQR4, PKMYT1, CODC64B, CLDN9, MIMP25, KREMEN2, TNFRSF12A, THOC6, CLDN6	TRUE	
1017	6736	12735 p1	16	3107033	3119356	12323	12 del	1	≤ 0.1% fa_to_p1	MIMP25, IL32	TRUE	
1020	6925	13512 p1	16	4390319	4391505	1186	3 dup	1	≤ 0.1% fa_to_p1	COROT-PAM16	TRUE	
1022	6752	13018 s1	16	4621561	4642368	20807	10 del	1	≤ 0.1% mo_to_both	GLYR1	TRUE	POOLCE [frameshift]
1022	6751	13018 p1	16	4624972	4642368	17396	8 del	1	≤ 0.1% mo_to_both	GLYR1	TRUE	MCOLN3 [missense]
1023	4043	11118 p1	16	4871380	4871598	218	2 dup	2	≤ 1% mo_to_p1	USP7, METTL22, TMEM186, ABAT, PMM2, CARHSP1	TRUE	PMIM2, ABAT
1029	6796	13327 s1	16	8722586	8997259	274673	53 dup	1	≤ 0.1% mo_to_both	METTL22	TRUE	
1028	6794	13327 p1	16	8728895	8740000	11105	8 dup	1	≤ 0.1% mo_to_both	NTAN1, PDXDC1, RRN3	TRUE	
1039	6719	12652 s1	16	15125591	15168718	43127	24 dup	3	≤ 1% fa_to_s1		TRUE	CTSB [missense]

1039	6756	13125 p1	16	15125591	15135533	9942	12 dup	3	≤ 1% fa.to both	NTAN1, PDXDC1	TRUE	RAD21L1 [nonsense]
1039	6757	13125 s1	16	15126717	15135533	8616	11 dup	3	≤ 1% fa.to both	NTAN1, PDXDC1	TRUE	
1045	6772	13215 p1	16	15596178	15609285	13107	6 del	1	≤ 0.1% fa.to p1	C16orf45	TRUE	
1046	6829	13689 s1	16	16146580	16276445	129865	36 dup	1	≤ 0.1% mo.to s1	ABCC6, ABCC1	TRUE	QOZ4 [missense], FBN1 [nonsense], IARS [nonsense]
1054	989	13593 s1	16	19603111	19663412	60301	19 dup	1	≤ 0.1% fa.to s1	C16orf62	FALSE	
1057	6760	13139 p1	16	21623965	21742251	118286	30 del	5	≤ 1% mo.to both	METTL9, OTOA	FALSE	
1057	6761	13139 s1	16	21623965	21763826	139961	35 del	5	≤ 1% mo.to both	METTL9, OTOA	FALSE	C7orf71 [missense]
1057	6996	12441 p1	16	21655610	21763826	108216	29 dup	5	≤ 1% mo.to p1	METTL9, OTOA	TRUE	
1057	6727	12697 p1	16	21655610	21768598	112988	31 del	5	≤ 1% fa.to p1	METTL9, OTOA	TRUE	
1062	4115	12100 s1	16	23079366	23117808	38442	14 dup	3	≤ 0.1% fa.to s1	USP91	TRUE	
1080	7516	11629 p1	16	23675049	30199897	524646	192 dup	7	≤ 0.1% mo.to p1	DOCA, ASPHD1, CORO1A, TBX6, KIF22, CDIPT, QPRT, YPEL3, PPP4C, MAPK3, SPN, WWP, FAM57B, ALDOA, INO80E, SEZ6L2, TAOX2, KCTD13, MAZ, PRRT2, GDDP3, C16orf52, C16orf53, TMEM219, C16orf54, HIRIP3	TRUE	ALDOA, MAPK3, SEZ6L2
1080	4203	13509 p1	16	29675049	30199408	524359	191 dup	7	≤ 0.1% fa.to p1	DOCA, ASPHD1, CORO1A, TBX6, KIF22, CDIPT, QPRT, YPEL3, PPP4C, MAPK3, SPN, WWP, FAM57B, ALDOA, INO80E, SEZ6L2, TAOX2, KCTD13, MAZ, PRRT2, GDDP3, C16orf52, C16orf53, TMEM219, C16orf54, HIRIP3	TRUE	ALDOA, MAPK3, SEZ6L2
1085	966	12152 s1	16	30793038	30849705	56667	5 dup	1	≤ 0.1% mo.to s1	ZNF629	TRUE	
1088	943	11229 p1	16	31477170	31488897	11727	11 dup	1	≤ 0.5% mo.to both	TGFB111, ARMC5	FALSE	
1088	945	11229 s1	16	31477170	31487883	10713	10 dup	1	≤ 0.5% mo.to both	TGFB111, ARMC5	FALSE	
1088	6806	13366 p1	16	57596005	57604447	8442	9 del	1	≤ 0.1% mo.to both	GPR114	FALSE	
1088	8471	13366 s1	16	57596005	57604447	8442	9 del	1	≤ 0.1% mo.to both	GPR114	FALSE	
1101	8472	13227 s1	16	66967889	66972144	4255	4 del	1	≤ 0.1% mo.to s1	CES2, FAM98B	TRUE	LOOR [missense]
1105	4217	13825 p1	16	68224670	68283349	58679	19 dup	1	≤ 0.1% mo.to both	ESRP2, NFATC3, PLA2G15	FALSE	
1105	4218	13825 s1	16	68224670	68283349	58679	19 dup	1	≤ 0.1% mo.to both	ESRP2, NFATC3, PLA2G15	FALSE	
1107	996	14201 p1	16	68710287	68713877	3590	5 dup	2	≤ 0.1% mo.to p1	CDH3	TRUE	
1114	979	13169 s1	16	70286623	70296427	9804	10 dup	2	≤ 0.1% fa.to s1	AARS	FALSE	
1109	4191	12790 p1	16	70363194	70405466	42272	16 dup	1	≤ 0.1% fa.to p1	DDX19A, DDX19B	TRUE	
1110	4113	12100 p1	16	70714696	70714928	232	2 dup	3	≤ 0.5% fa.to p1	MTSS1L	TRUE	
1112	6715	12628 s1	16	71997103	72001906	4803	3 del	2	≤ 0.5% fa.to both		TRUE	GSDE1 [missense], OTUD7A [3'-non-frameshifting]
1112	6714	12628 p1	16	72001035	72001906	871	2 del	2	≤ 0.5% fa.to both		TRUE	
1112	986	13447 s1	16	72001035	72001906	871	2 del	2	≤ 0.5% fa.to s1		FALSE	
1125	4208	13621 s1	16	75276367	75690509	414142	55 dup	1	≤ 0.1% fa.to s1	KARS, TMEM170A, BCAR1, GABARAPL2, TMEM231, ADAT1, TERP2IR, CFDP1, CHST5, CHST6	TRUE	CHST5
1128	4091	11581 p1	16	77322004	77769883	537879	36 dup	1	≤ 0.1% mo.to p1	ADAMTS18, MON1B, NUDT7	TRUE	
1129	6814	13493 p1	16	77866644	77918699	22055	4 del	1	≤ 0.1% fa.to both	VANTL	TRUE	
1129	6815	13493 s1	16	77866644	77918699	22055	4 del	1	≤ 0.1% fa.to both	VANTL	TRUE	
1131	4104	11797 p1	16	78056523	78062087	5564	2 del	2	≤ 0.5% fa.to both	CLEC3A	TRUE	
1131	4105	11797 s1	16	78056523	78062087	5564	2 del	2	≤ 0.5% fa.to both	CLEC3A	TRUE	
1131	6774	13216 p1	16	78056523	78064738	8215	3 del	2	≤ 0.5% fa.to both	CLEC3A	TRUE	ITGA7 [missense]
1131	6776	13216 s1	16	78056523	78064738	8215	3 del	2	≤ 0.5% fa.to both	CLEC3A	TRUE	
1135	970	12373 p1	16	81314461	81396216	81755	10 dup	2	≤ 0.1% fa.to p1	GAN, BCOM1	TRUE	GAN
1142	962	11964 p1	16	84438733	84459407	20674	9 del	2	≤ 0.1% mo.to p1	ATP2C2	TRUE	
1143	7863	11412 p1	16	84693365	84812688	119323	15 del	9	≤ 2% mo.to p1	KULH36, USP10	FALSE	
1160	1036	12578 p1	17	6010	611981	5971	3 dup	3	≤ 0.1% mo.to p1		TRUE	
1162	6906	13263 s1	17	6010	636421	630411	38 dup	2	≤ 0.1% mo.to s1	RPHSAL, FAM101B, C17orf97, FAM157A, VPSS3	FALSE	
1161	6898	13165 s1	17	295626	729318	433692	48 dup	1	≤ 0.1% fa.to s1	RNMTL1, FAM57A, GLOD4, GEMIN4, VPSS3, NXN	TRUE	
1159	1061	13335 p1	17	673108	695309	22201	12 del	2	≤ 0.1% fa.to both	RNMTL1, GLOD4	FALSE	ZNF420 [missense]
1159	1062	13335 s1	17	673108	695309	22201	12 del	2	≤ 0.1% fa.to both	RNMTL1, GLOD4	FALSE	
1159	6959	13599 s1	17	673108	695309	22201	12 del	2	≤ 0.1% mo.to s1	RNMTL1, GLOD4	TRUE	
1166	6903	13196 s1	17	3458025	3486724	28699	9 del	2	≤ 0.1% fa.to s1	TRPV1, TRPV3	TRUE	TRPV3
1168	4278	11285 s1	17	3518630	3561469	42839	14 del	2	≤ 0.1% mo.to s1	CTNS, SHPK	FALSE	
1168	4306	11532 p1	17	3518630	3561469	42839	14 del	2	≤ 0.1% mo.to p1	CTNS, SHPK	TRUE	
1167	4390	13271 p1	17	3550737	3552225	1468	2 del	2	≤ 0.1% mo.to both	CTNS	TRUE	
1178	4275	11267 p1	17	7258523	7259951	1428	7 dup	1	≤ 0.1% mo.to both	TMEM95	TRUE	
1178	4276	11267 s1	17	7258523	7259951	1428	7 dup	1	≤ 0.1% mo.to both	TMEM95	TRUE	
1189	4234	11067 s1	17	11668766	11713686	46930	10 dup	1	≤ 0.1% mo.to s1	DNAAH9	TRUE	
1202	6919	13328 p1	17	14085305	15456878	136373	31 dup	1	≤ 0.5% fa.to p1	RNP22, CDBT15, TEK13, Nore, FAM1882-CDRT4, HSS3TB1, COX10	FALSE	SMAD2 [missense]
1207	6859	12579 p1	17	16320982	16351274	30292	18 dup	1	≤ 0.1% fa.to both	NCRNA00188, TRPV2	TRUE	
1207	6860	12579 s1	17	16320982	16347445	26463	17 dup	1	≤ 0.1% fa.to both	NCRNA00188, TRPV2	TRUE	
1218	4412	13843 p1	17	26918715	26920084	1369	2 del	1	≤ 0.1% fa.to both	SPAG5	TRUE	AGK [missense]
1218	4413	13843 s1	17	26918715	26920084	1369	2 del	1	≤ 0.1% fa.to both	SPAG5	TRUE	DOCK10 [missense]
1225	4295	11484 s1	17	29311634	29324349	12715	3 dup	3	≤ 0.1% fa.to s1	RNF135	TRUE	
1237	4269	11196 p1	17	33679374	33769305	89931	9 del	4	≤ 1% mo.to p1	SLFN11, SLFN12, SLFN13	TRUE	
1237	6948	13507 s1	17	33679374	33769305	89931	9 del	4	≤ 1% mo.to s1	SLFN11, SLFN12, SLFN13	FALSE	SOOC [frameshift], LAMA1 [missense]
1239	1006	11459 p1	17	34182066	34183809	1743	4 del	1	≤ 0.1% fa.to both	C17orf66	TRUE	DEPC7 [missense]
1239	1007	11459 s1	17	34182066	34183809	1743	4 del	1	≤ 0.1% fa.to both	C17orf66	TRUE	
1264	4388	13195 s1	17	39671723	39680792	9069	9 dup	1	≤ 0.5% fa.to s1	KRT15, KRT19	TRUE	
1300	1009	11479 s1	17	5731852	57430887	119035	11 dup	1	≤ 0.1% mo.to s1	GDPD1, YPEL2	TRUE	
1313	4256	11107 p1	17	67075113	67087403	12290	14 del	2	≤ 0.1% mo.to both	ABCA6	TRUE	TROAP [missense]
1313	4257	11107 s1	17	67075113	67087403	12290	14 del	2	≤ 0.1% mo.to both	ABCA6	TRUE	
1320	4352	12561 p1	17	72874459	72877385	2926	3 del	1	≤ 0.5% fa.to both	FAD56	FALSE	NLRX1 [missense], ADAM33 [nonsense]

1320	4353	12561 s1	17	72874459	72877385	2926	3 del	1	≤ 0.5% fa.to both	FADS6	FALSE	ROB03 [missense]
1330	4376	13000 s1	17	74621411	7287793	4382	8 dup	1	≤ 0.1% mo.to s1	STEGAL NAC1	TRUE	
1331	4396	13608 p1	17	74865111	7593785	533674	38 del	1	≤ 0.1% fa.to p1	SEPT9, NCRNA00338, MGAT5B, SEC14L1	FALSE	CSD01 [nonsense]
1333	6962	13601 p1	17	76198579	76210508	11929	10 dup	1	≤ 0.1% fa.to both	BIRC5, AFID	TRUE	AK1 [missense]
1336	6963	13601 p1	17	76198783	76202131	3348	7 dup	1	≤ 0.1% fa.to both	AFID	TRUE	
1336	4272	11220 p1	17	78188413	78188937	524	2 del	1	≤ 0.1% fa.to both	SGSH	FALSE	SGSH
1336	7895	11220 p1	17	78188413	78188937	524	2 del	1	≤ 0.1% fa.to both	SGSH	FALSE	SGSH
1337	6867	12691 s1	17	78221928	78323706	101778	36 dup	1	≤ 0.1% mo.to s1	SLC36A11, RNF213	FALSE	
1345	6881	12937 s1	17	79827048	79827282	234	2 del	1	≤ 0.1% fa.to s1	APHGDIA	TRUE	PNPLA6 [missense]
1351	6938	12424 p1	17	80151631	80153240	1809	3 del	1	≤ 0.1% mo.to p1	CCDC57	FALSE	OR2T3 [missense]
1356	4423	11008 p1	18	580408	645097	64689	8 dup	1	≤ 0.1% mo.to both	CLU11, CETN1	FALSE	KATNAL2 [splice]
1356	4424	11008 p1	18	580408	645097	64689	8 dup	1	≤ 0.1% mo.to both	CLU11, CETN1	FALSE	
1357	4435	11252 p1	18	688573	697355	8762	7 del	4	≤ 0.1% mo.to p1	ENOSF1	FALSE	LCN10 [missense]
1373	4442	11429 p1	18	18531341	18650574	119233	32 dup	1	≤ 0.1% fa.to both	ROCK1	FALSE	IL6R [missense]
1373	4443	11429 p1	18	18531341	18650574	119233	32 dup	1	≤ 0.1% fa.to both	ROCK1	FALSE	
1375	6973	12384 s1	18	21056969	21059388	2419	3 dup	1	≤ 0.1% fa.to s1	RIOK3	TRUE	
1379	6981	12697 p1	18	24367174	24628467	192293	10 dup	1	≤ 0.1% fa.to p1	CHST9, AQP4, CHST9-AS1	TRUE	
1380	6979	12523 p1	18	28683329	29049312	80983	25 dup	1	≤ 0.5% fa.to p1	DSG3, DSG4	FALSE	G3BP2 [missense]
1384	1079	11942 p1	18	32953419	32954256	837	2 del	2	≤ 0.5% mo.to p1	ZNF396	FALSE	
1384	6997	13443 p1	18	32953419	32954256	837	2 del	2	≤ 0.5% mo.to p1	ZNF396	FALSE	CPZ [missense]
1385	4483	12656 p1	18	33936930	33946648	11718	4 dup	1	≤ 0.1% mo.to p1	MOCOS	TRUE	ZNF780A [missense]
1392	1082	12250 s1	18	47008714	47010141	1427	2 dup	1	≤ 0.1% fa.to s1	C18orf82	FALSE	VP518 [missense]
1397	6975	12420 s1	18	47801346	47801814	468	3 dup	1	≤ 2% fa.to s1	MBD1	FALSE	
1406	1087	13726 p1	18	64211207	64212140	933	2 del	4	≤ 0.1% mo.to both	CDH19	TRUE	
1406	1088	13726 p1	18	64211207	64212140	933	2 del	4	≤ 0.1% mo.to both	CDH19	TRUE	MAPK8 [missense]
1412	4484	12656 p1	18	70416262	70417855	1593	2 dup	1	≤ 0.1% fa.to both	NETO1	TRUE	CPZ [missense]
1412	4485	12656 p1	18	70416262	70417855	1593	2 dup	1	≤ 0.1% fa.to both	NETO1	TRUE	
1414	4489	12869 p1	18	72229281	72251798	22517	8 dup	1	≤ 0.1% fa.to p1	CNDP1	TRUE	PRCP [missense]
1415	7002	13508 p1	18	74817166	74980858	163692	4 dup	1	≤ 0.1% mo.to both	GALR1, MBP	FALSE	
1415	7003	13508 s1	18	74968113	74980858	12745	2 dup	1	≤ 0.1% mo.to both	GALR1	FALSE	
1417	4438	11356 p1	18	77470345	77891075	420730	28 dup	2	≤ 0.1% fa.to p1	KCNQ2, RBFA, CTDP1, ADNP2, TXNL4A, PQLC1	TRUE	NAPRT1 [splice], SV2B [missense]
1417	4461	11810 p1	18	77663975	78003231	341256	23 dup	2	≤ 0.1% fa.to both	PARP6G, LOC100130522, RBFA, ADNP2, TXNL4A, PQLC1	FALSE	
1417	4462	11810 s1	18	77693968	78003231	311263	21 dup	2	≤ 0.1% fa.to both	PARP6G, LOC100130522, RBFA, ADNP2, TXNL4A, PQLC1	FALSE	RSRC1 [nonsense]
1423	1143	12358 p1	19	1783026	1784945	1919	2 del	2	≤ 0.5% fa.to p1	ATPRB3	TRUE	TSRP [missense]
1424	1180	13815 p1	19	1799945	1802644	2899	4 del	2	≤ 0.5% fa.to p1	ATPRB3	TRUE	
1429	4530	11288 p1	19	2410255	2418136	7881	5 del	1	≤ 0.5% fa.to both	TNFRSS9	TRUE	SLC6A13 [missense]
1429	4531	11288 p1	19	2410255	2418136	7881	5 del	1	≤ 0.5% fa.to both	TNFRSS9	TRUE	
1431	1132	12161 s1	19	3805946	3834961	29035	17 dup	2	≤ 0.1% fa.to both	ZFR2	FALSE	
1431	1131	12161 p1	19	3807169	3833776	26607	15 dup	2	≤ 0.1% fa.to both	ZFR2	FALSE	UBR3 [frameshift], CARKD [nonsense]
1431	7051	12837 p1	19	3816671	3825405	8734	7 del	2	≤ 0.1% fa.to both	ZFR2	TRUE	SH3RF3 [missense]
1431	7052	12837 s1	19	3816671	3825405	8734	7 del	2	≤ 0.1% fa.to both	ZFR2	TRUE	
1442	8514	13296 p1	19	6881951	6886913	4962	8 dup	1	≤ 0.1% mo.to p1	C3	TRUE	
1448	7030	12626 p1	19	6890491	6991143	100852	27 dup	6	≤ 0.5% fa.to both	EMR1, EMR4P	TRUE	
1448	4659	12780 p1	19	6890491	7083755	193264	41 dup	6	≤ 0.5% fa.to both	EMR1, MBD3L2, EMR4P, ZNF557	TRUE	
1448	4660	12780 s1	19	6890491	7083755	193264	41 dup	6	≤ 0.5% fa.to both	EMR1, MBD3L2, EMR4P, ZNF557	TRUE	RABEP1 [missense]
1448	1088	11364 s1	19	6896408	6991143	187347	27 dup	6	≤ 0.5% fa.to s1	EMR1, MBD3L2, EMR4P, ZNF557	FALSE	
1448	7020	12462 p1	19	6896408	7083755	187347	40 dup	6	≤ 0.5% mo.to s1	EMR1, MBD3L2, EMR4P, ZNF557	TRUE	EPB41L3 [nonsense]
1448	7028	12626 p1	19	6896408	7083755	187347	26 dup	6	≤ 0.5% fa.to both	EMR1, EMR4P	TRUE	GALNT9 [missense]
1448	7112	13504 p1	19	6897159	6989173	92014	24 dup	6	≤ 0.5% fa.to both	EMR1, EMR4P	TRUE	
1447	7029	12626 p1	19	7075085	7083755	8670	6 dup	4	≤ 0.5% fa.to both	ZNF557	TRUE	GALNT9 [missense]
1447	7031	12626 s1	19	7075085	7083755	8670	6 dup	4	≤ 0.5% fa.to both	ZNF557	TRUE	RABEP1 [missense]
1447	4661	12780 s1	19	7075085	7083755	8670	6 dup	4	≤ 0.5% fa.to both	ZNF557	TRUE	
1447	7111	13504 p1	19	7081370	7083755	2385	3 dup	4	≤ 0.5% fa.to both	ZNF557	TRUE	
1453	4644	12616 p1	19	8158329	8168571	9242	9 dup	1	≤ 0.1% fa.to both	FN3	FALSE	
1453	4645	12616 s1	19	8158329	8168571	9242	9 dup	1	≤ 0.1% fa.to both	FN3	FALSE	RAB3D
1465	7006	12252 s1	19	11438073	11448075	12002	4 dup	1	≤ 0.1% mo.to s1	RAB3D, TSPAN16	FALSE	
1466	1108	14779 p1	19	11541718	11548945	7227	6 dup	1	≤ 0.1% mo.to p1	PRKCSH, CDCG151	TRUE	
1472	1125	12106 p1	19	14142684	14153601	10917	6 dup	4	≤ 1% fa.to p1	IL27FA	TRUE	
1472	7116	13512 s1	19	14143197	14153601	10404	5 dup	4	≤ 1% fa.to s1	IL27FA	TRUE	
1472	7101	13487 s1	19	14153354	14153601	937	2 del	4	≤ 1% mo.to s1	IL27FA	FALSE	
1481	4549	11288 p1	19	18704375	18704917	542	2 dup	1	≤ 0.1% mo.to p1	CLF1	TRUE	SLC6A13 [missense]
1482	7054	12636 s1	19	19030031	19033553	3822	7 dup	1	≤ 0.1% mo.to s1	DDX43, COPE	TRUE	PAM49a [frameshift]
1519	4580	11724 p1	19	40199836	40225713	29877	4 dup	4	≤ 0.1% fa.to p1	CLC, LGALS14	TRUE	TUBA1A [missense]
1528	7010	12384 s1	19	43920032	43922549	2517	5 del	3	≤ 0.1% fa.to s1	TEX101	TRUE	
1528	4678	13509 p1	19	43920032	43922549	2517	5 del	3	≤ 0.1% mo.to both	TEX101	TRUE	
1528	4679	13509 p1	19	43920032	43922549	2517	5 del	3	≤ 0.1% mo.to both	TEX101	TRUE	
1530	1191	14201 s1	19	45284161	45296877	12716	7 dup	1	≤ 0.1% fa.to s1	CBL	TRUE	
1531	1091	11029 p1	19	45494107	45495658	1551	3 del	1	≤ 0.1% mo.to both	CLPTM1	TRUE	
1531	1092	11029 s1	19	45494107	45495658	1551	3 del	1	≤ 0.1% mo.to both	CLPTM1	TRUE	
1532	4510	11085 s1	19	45848799	45899694	50895	44 dup	2	≤ 0.5% mo.to s1	ERCC2, KLC3, PPP1R13L	TRUE	POGK [missense]
1532	1158	13116 p1	19	45848799	45901576	52777	47 dup	2	≤ 0.5% fa.to both	ERCC2, KLC3, PPP1R13L	TRUE	ERCC2
1532	1160	13116 p1	19	45848799	45901576	52777	47 dup	2	≤ 0.5% fa.to both	ERCC2, KLC3, PPP1R13L	TRUE	ERCC2
1534	7057	12840 p1	19	46119059	46120111	1052	3 del	1	≤ 0.1% mo.to both	EML2	TRUE	ATP1B1 [nonsense], TM4SF19 [splice]
1534	7058	12840 s1	19	46119059	46120111	1052	3 del	1	≤ 0.1% mo.to both	EML2	TRUE	
1544	4554	11304 s1	19	49481180	49497180	16000	11 dup	1	≤ 0.1% fa.to s1	GY31, RUVBL2	TRUE	
1546	7060	12843 p1	19	49836257	49875365	39108	37 dup	1	≤ 0.5% mo.to p1	TRPM4, PPIFA3, HRC	FALSE	

1547	7087	13296 s1	19	49671173	49675365	4192	6 del	1	≤ 0.5% mo_to_s1	TRPM4	TRUE	
1557	7043	12705 p1	19	52227176	52543836	316660	29 dup	4	≤ 0.1% fa_to_both	ZNF613, ZNF615, ZNF614, FPR1, FPR2, FPR3, ZNF432, ZNF350, ZNF649, ZNF577	FALSE	DIP2C [frameshift]
1557	7044	12705 s1	19	52271911	52550121	278210	30 dup	4	≤ 0.1% fa_to_both	ZNF613, ZNF615, ZNF614, FPR2, FPR3, ZNF432, ZNF350, ZNF649, ZNF577	FALSE	PLXNA4 [missense]
1557	7082	13215 s1	19	52271911	52579425	307514	33 dup	4	≤ 0.1% mo_to_s1	ZNF613, ZNF615, ZNF614, FPR2, FPR3, ZNF432, ZNF350, ZNF649, ZNF577	TRUE	
1568	4697	14110 p1	19	54629902	54631752	1850	3 del	1	≤ 0.1% mo_to_p1	PRPF31	FALSE	PHF3 [missense]
1568	4674	13171 p1	19	55872060	55879739	7679	10 del	1	≤ 0.1% mo_to_p1	IL11	TRUE	
1596	7074	13162 s1	19	57286055	57293489	7434	3 dup	1	≤ 0.1% fa_to_s1	ZNF2	TRUE	
1596	1184	13815 p1	19	57835049	57932849	97900	15 del	1	≤ 0.1% mo_to_p1	ZNF547, ZNF304, ZNF17, ZNF548, ZNF543	TRUE	
1601	7045	12716 p1	19	58639244	58652113	12869	2 dup	1	≤ 0.1% fa_to_both	ZNF529	TRUE	
1788	8293	12409 s1	20	3734704	3739325	4621	4 dup	5	≤ 0.5% mo_to_s1	C22orf27	TRUE	
1788	1219	13335 s1	20	3734704	3739325	1550	3 dup	5	≤ 0.5% mo_to_s1	C22orf27	FALSE	
1788	4741	12650 p1	20	3735043	3739325	4282	3 dup	5	≤ 0.5% mo_to_both	C22orf27	FALSE	
1789	4742	12650 s1	20	3735043	3739325	4282	3 dup	5	≤ 0.5% mo_to_both	C22orf27	FALSE	
1792	7131	13101 s1	20	3835271	3893281	59010	12 dup	1	≤ 0.1% fa_to_s1	MAVS, PANK2	TRUE	PANK2
1792	7192	12383 p1	20	3835271	3893456	3185	2 dup	1	≤ 0.5% fa_to_both	MAVS	FALSE	
1791	7993	12383 s1	20	3835271	3893456	3185	2 dup	1	≤ 0.5% fa_to_both	MAVS	FALSE	
1797	1193	11013 p1	20	7864239	7990962	12673	15 dup	1	≤ 0.1% mo_to_both	HAO1, TMX4	TRUE	TMFRSS2 [missense]
1797	1194	11013 s1	20	7864239	7990962	12673	15 dup	1	≤ 0.1% mo_to_both	HAO1, TMX4	TRUE	CGNL1 [missense], DENND5B [missense], LRRC40 [missense]
1799	7136	13418 p1	20	13856683	13869158	12475	7 dup	1	≤ 0.1% mo_to_both	SEL1L2	FALSE	
1799	7137	13418 s1	20	13858165	13869158	10993	6 dup	1	≤ 0.1% mo_to_both	SEL1L2	FALSE	
1806	7143	13589 s1	20	25469879	25472161	2282	3 del	1	≤ 0.1% fa_to_s1	NNL	TRUE	
1808	1207	11872 p1	20	30533609	30538190	4581	4 dup	1	≤ 0.5% mo_to_p1	PDRG1	TRUE	CACNA1D [missense], KATNAL2 [splice]
1824	4726	11501 p1	20	44351006	44354275	3269	3 del	6	≤ 2% mo_to_p1	SPINT4	FALSE	
1824	1224	13926 s1	20	44351006	44354275	3269	3 del	6	≤ 2% mo_to_s1	SPINT4	TRUE	
1828	7995	11304 p1	20	45228609	45235221	6612	2 dup	2	≤ 0.1% fa_to_p1	SLC13A3	TRUE	
1828	7122	12473 p1	20	45228609	45235221	6612	2 dup	2	≤ 0.1% mo_to_p1	SLC13A3	TRUE	
1830	1822	13815 s1	20	48124456	48166726	42270	11 del	2	≤ 0.5% fa_to_s1	PTGS	TRUE	
1842	4712	11196 s1	20	62953245	62940953	51708	15 dup	1	≤ 0.5% fa_to_both	PCMTD2, MYT1	TRUE	
1851	7175	13396 p1	21	19628825	19632603	3778	3 del	1	≤ 0.1% mo_to_p1	CHOLD	TRUE	
1854	4778	11581 p1	21	2596371	26317408	21037	7 dup	1	≤ 0.1% fa_to_both	ADAMTS	TRUE	
1854	4779	11581 s1	21	2596371	26317408	21037	7 dup	1	≤ 0.1% fa_to_both	ADAMTS	TRUE	
1864	4808	12507 p1	21	35742777	35899047	156270	8 dup	2	≤ 0.5% mo_to_both	KCNJ2, RCAN1, KONE1, FAM165B	FALSE	
1864	7170	13327 p1	21	35742777	35899047	156270	8 dup	2	≤ 0.5% mo_to_both	KCNJ2, RCAN1, KONE1, FAM165B	TRUE	
1864	4810	12507 s1	21	35890391	35899047	8666	4 dup	2	≤ 0.5% fa_to_p1	RCAN1	FALSE	
1868	1233	13593 p1	21	37649318	37665869	16551	10 dup	2	≤ 0.1% mo_to_both	DOPEY2	FALSE	RGS22 [missense]
1868	1234	13593 s1	21	37649318	37665869	16551	10 dup	2	≤ 0.1% mo_to_both	DOPEY2	FALSE	
1874	4784	11716 p1	21	42603997	42815068	14071	9 dup	1	≤ 0.1% mo_to_both	MX1	TRUE	
1880	4799	12308 p1	21	44836621	44837654	1033	2 dup	2	≤ 2% mo_to_p1	SIK1	TRUE	
1880	7149	12481 p1	21	45724911	45826648	101737	48 dup	1	≤ 0.1% fa_to_p1	PKL, C2orf2, TRPM2	FALSE	
1887	4824	13795 p1	21	46318981	46323450	4469	4 del	1	≤ 0.1% mo_to_both	ITGB2	TRUE	ITGB2
1887	4825	13795 s1	21	46318981	46323450	4469	4 del	1	≤ 0.1% mo_to_both	ITGB2	TRUE	ITGB2
1897	4768	11196 p1	21	48019275	48084239	64964	12 dup	5	≤ 0.1% mo_to_both	PRMT2, S100B	TRUE	
1898	7216	12997 p1	22	17264508	17268963	24455	3 dup	3	≤ 1% mo_to_p1	XKR3	TRUE	
1898	7230	13418 p1	22	17264508	17268963	24455	3 dup	5	≤ 1% mo_to_p1	XKR3	FALSE	CGNL1 [missense], DENND5B [missense], LRRC40 [missense]
1910	7191	12481 p1	22	21369463	21473538	104075	17 dup	1	≤ 0.1% fa_to_both	SLC7A4, P2RX6	FALSE	
1910	7192	12481 s1	22	21369463	21384640	15177	15 dup	1	≤ 0.1% fa_to_both	SLC7A4, P2RX6	FALSE	
1917	8016	12680 p1	22	22123492	22127271	3779	2 dup	1	≤ 0.1% fa_to_both	MAPK1	TRUE	MAPK1
1927	4841	11090 p1	22	23401639	24967945	1566306	242 dup	1	≤ 0.1% mo_to_p1	GUSBP11, MMP11, SLC2A11, SPECC1L, UPB1, CHCHD10, ZNF70, VPRED3, RAB36, CABIN1, RTDR1, DDTL, C22orf13, DERL3, C22orf15, BCR, SUSD2, SMARCB1, ADORA2A, LOC284889, DDT, GGT5, SNRPD3, GSTT2, GSTT1, C22orf43, GLL1	TRUE	UPB1, ADORA2A, SMARCB1
1930	8025	11090 p1	22	24034217	24325797	291580	80 dup	1	≤ 0.1% mo_to_p1	GUSBP11, SMARCB1, MMP11, LOC284889, DDTL, SLC2A11, DDT, DERL3, C22orf15, CHCHD10, ZNF70, VPRED3, GSTT2	TRUE	SMARCB1
1931	4854	11242 s1	22	24034217	24037206	2989	6 dup	1	≤ 0.5% fa_to_s1	GUSBP11	TRUE	
1927	8024	11090 p1	22	24431961	24967945	535984	98 dup	1	≤ 0.1% mo_to_p1	ADORA2A, SNRPD3, CABIN1, SPECC1L, C22orf13, SUSD2, GGT5, UPB1	TRUE	UPB1, ADORA2A
1938	7194	12498 s1	22	29092888	29095925	3037	2 del	1	≤ 0.1% mo_to_both	CHEK2	TRUE	
1938	8522	12498 p1	22	29092888	29095925	3037	2 del	1	≤ 0.1% mo_to_both	CHEK2	TRUE	GPR82 [missense]
1947	1263	12810 p1	22	32545739	32651316	105577	26 dup	1	≤ 0.1% fa_to_p1	C22orf42, Nme, RPLP2, SLC5A4	TRUE	CFA4 [missense]
1950	1283	14011 p1	22	32786226	32794087	5861	5 del	1	≤ 0.1% fa_to_both	C22orf28	TRUE	
1950	1284	14011 s1	22	32786226	32794087	5861	5 del	1	≤ 0.1% fa_to_both	C22orf28	TRUE	
1952	1240	11364 p1	22	35713969	35742965	29096	13 dup	1	≤ 0.1% mo_to_both	TOM1	FALSE	
1952	1241	11364 s1	22	35713969	35742965	29096	13 dup	1	≤ 0.1% mo_to_both	TOM1	FALSE	
1954	4856	11247 p1	22	36701975	36907833	205958	29 dup	1	≤ 0.1% fa_to_p1	TXN2, FOXRED2, MYH9, EIF3D	TRUE	
1958	4963	13992 s1	22	38477895	38483271	5376	5 dup	1	≤ 0.1% fa_to_s1	SLC16A8, BAIAP2L2	TRUE	
1966	7215	12937 s1	22	40754867	40762526	7659	9 del	1	≤ 0.1% fa_to_s1	ADSL	TRUE	OR2T3 [missense]
1979	4893	11828 p1	22	42781166	42899592	118416	8 dup	2	≤ 0.1% mo_to_p1	NFAM1, SERHL	TRUE	
1978	8034	11654 p1	22	42956191	42962344	6153	2 dup	8	≤ 2% fa_to_both	SERHL2	TRUE	
1978	1267	13116 p1	22	42956191	42962344	6153	2 del	8	≤ 2% fa_to_both	SERHL2	TRUE	
1978	4948	13625 p1	22	42956191	42962344	6153	2 del	8	≤ 2% fa_to_both	SERHL2	TRUE	
1978	4948	13625 s1	22	42956191	42962344	6153	2 del	8	≤ 2% fa_to_both	SERHL2	TRUE	
1984	4821	12375 s1	22	44322814	44377341	54527	19 dup	1	≤ 0.1% fa_to_s1	PNPLA3, SAMM50	TRUE	
1987	1255	12011 s1	22	44883503	44893436	9933	2 dup	1	≤ 0.1% fa_to_both	LDOC1L	TRUE	

1987	7445	12011 p1	22	44883503	44893436	9633	2 dup	1	≤ 0.1% fa to both	LDOC1L	TRUE	FAM45A [nonsense]
1988	1273	13606 s1	22	45121123	45198044	76921	8 del	1	≤ 0.1% fa to s1	PRR-ARHGAP8, PRR5	TRUE	ZFYVE26 [nonsense]
1994	7186	12409 p1	22	47287161	47308084	20823	3 del	1	≤ 0.1% mo to both	TBC1D22A	TRUE	LRP2 [nonsense]
1994	8318	12409 s1	22	47287161	47308084	20823	3 del	1	≤ 0.1% mo to both	TBC1D22A	TRUE	
1998	4887	11716 p1	22	51016203	51018700	2497	6 del	2	≤ 0.1% fa to both	CHKB-CPT1B, CPT1B	TRUE	
1998	4888	11716 s1	22	51016203	51018700	2497	6 del	2	≤ 0.1% fa to both	CHKB-CPT1B, CPT1B	TRUE	

Table S4

cnvID	callID	familyID	rel	Chromosome	Start (hg19)	Stop (hg19)	length (bp)	length (exons)	stat e (exons)	Freq uen cy in 411 qua ds	Genes
2363	2111	11079 p1		3	195778812	197273314	1,494,502	161	del	1	PCYT1A, FBXO45, LRR33, WDR53, NCBP2, TM4SF19, TCETX1D2, DLG1, MF12, SENP5, RNF168, ZDHHC19, OSTalpha, TFRC, C3orf43, CEP19, FIGX, TCETX1D2, FIGZ, UBXN7, PAK2, BDH1
1080	4033	11090 p1		16	29675049	30199897	524,848	187	del	7	DOC2A, ASPHD1, CORO1A, TBX6, KIF22, CDIPT, QPRT, YPEL3, PPP4C, MAPK3, SPN, MVP, FAM57B, ALDOA, INO80E, SEZ6L2, TAOK2, KCTD13, MAZ, PRRT2, GPPD3, C16orf92, C16orf53, TMEM219, C16orf54, HIRIP3
2810	2737	11154 p1		7	72848337	74120764	1,272,427	209	dup	1	STX1A, WBSCR27, WBSCR22, LAT2, LIMK1, WBSCR28, RFC2, FZD9, VPS37D, ABHD11, CLIP2, CLDN3, CLDN4, BCL7B, ELN, MLXIPL, DNAJC30, GTF2IRD1, BAZ1B, TBL2, EIF4H, GTF2I
888	3883	11265 p1		15	22845761	23062319	216,558	57	del	6	NIP2A, NIP1A, CYFIP1, TUBGCP5
1251	4289	11353 p1		17	34892950	36104875	1,211,925	159	del	1	LHX1, DUSP14, MRM1, ACACA, DDX52, DHR511, SYNRG, HNF1B, AATF, FIGW, TADA2A, GGNBP2
1080	4069	11433 p1		16	29808213	30199897	391,684	175	del	7	DOC2A, ASPHD1, CORO1A, TBX6, KIF22, CDIPT, YPEL3, PPP4C, MAPK3, MVP, FAM57B, ALDOA, INO80E, SEZ6L2, TAOK2, KCTD13, MAZ, PRRT2, GPPD3, C16orf92, C16orf53, TMEM219, HIRIP3
1180	4305	11532 p1		17	6329934	7352107	1,022,173	366	dup	1	CLEC10A, ACADVL, APL1, C17orf74, C17orf100, SPEM1, FAM64A, LOC100506713, EIF5A, GPS2, DLG4, CTDNEP1, MIR497HG, TEK11, RNASEK-C17ORF49, GABARAP, DVL2, SLC16A11, SLC16A13, NEURL4, FGF11, CLDN7, C17orf81, NLGN2, C17orf61-PLSCR3, TNK1, MED31, SLC13A5, ACAP1, ASGR1, ASGR2, TMEM102, TMEM95, KIAA0753, PHF23, XAF1, PLSCR3, TXNDC17, SLC2A4, CHRN1B, KCTD11, PITPNM3, BCL6B, FBXO39, YBX2
1080	4114	12100 p1		16	29675049	30199897	524,848	193	del	7	DOC2A, ASPHD1, CORO1A, TBX6, KIF22, CDIPT, QPRT, YPEL3, PPP4C, MAPK3, SPN, MVP, FAM57B, ALDOA, INO80E, SEZ6L2, TAOK2, KCTD13, MAZ, PRRT2, GPPD3, C16orf92, C16orf53, TMEM219, C16orf54, HIRIP3
1057	4123	12162 p1		16	21636251	21737979	101,728	26	dup	5	METTL9, OTOA
1967	4904	12224 p1		22	40521821	40762526	240,705	38	del	1	ADSL, TNRC8B
1080	4144	12308 p1		16	29675049	30199897	524,848	193	del	7	DOC2A, ASPHD1, CORO1A, TBX6, KIF22, CDIPT, QPRT, YPEL3, PPP4C, MAPK3, SPN, MVP, FAM57B, ALDOA, INO80E, SEZ6L2, TAOK2, KCTD13, MAZ, PRRT2, GPPD3, C16orf92, C16orf53, TMEM219, C16orf54, HIRIP3
749	3705	12343 p1		13	41301637	41910892	609,255	53	dup	1	KBTD7, KBTD6, MTRF1, MRPS31, NAA16, MIR320D1, ELF1, SLC25A15, WBP4
1120	6725	12691 p1		16	69971020	72923861	2,952,841	487	del	1	WWP2, HP, SF3B3, DHX38, ZNF19, CLEC18A, CLEC18C, FUK, PMFBP1, HYDIN, TAT, ATXN1L, PDPR, MARVELD3, DDX19A, DDX19B, VAC14, ST3GAL2, None, MTSS1L, HPR, CALB2, KIAA0174, COG4, IL34, ZFXH3, ZNF23, PHLPP2, CHST4, AP1G1, FTSJD1, AARS, DHODH, TXNL4B
1682	5673	12735 p1		2	105438016	109293154	3,855,138	163	del	1	POU3F3, SULT1C3, ST6GAL2, NCK2, MRPS9, UXS1, SULT1C2, SULT1C4, C2orf40, FHL2, LIMS1, C2orf49, GCC2, SLC5A7, GPR45, TGFBAP1
1079	4186	12736 p1		16	29464953	30212614	747,661	222	dup	3	DOC2A, ASPHD1, CORO1A, TBX6, PRRT2, CDIPT, QPRT, YPEL3, PPP4C, SLX1B, MAPK3, SPN, BOLA2B, MVP, FAM57B, ALDOA, INO80E, SEZ6L2, TAOK2, KCTD13, SLX1A-SULT1A3, MAZ, KIF22, GPPD3, C16orf92, C16orf53, TMEM219, C16orf54, HIRIP3

568	6409	12975 p1	11	121348826	122852379	1,503,553	86 del	1	SORL1, CRTAM, UBASH3B, BSX, MIR100HG, C11orf63
1475	7065	13018 p1	19	15052300	16038149	985,849	234 dup	1	EPHX3, CYP4F8, PGLYRP2, CASP14, RASAL3, CYP4F3, CYP4F2, CCDC105, CYP4F12, CYP4F11, AKAP8L, OR111, ILVBL, SYDE1, BRD4, OR7C2, CYP4F22, NOTCH3, OR10H2, OR10H3, OR10H1, AKAP8, OR10H5, SLC1A6, WIZ
143	84	13346 p1	1	93545080	95712426	2,167,346	229 del	1	TMED5, GCLM, RWDD3, CNN3, DR1, ALG14, MTF2, TMEM56-RWDD3, DNTTIP2, SLC44A3, ABCA4, F3, MIR760, TMEM56, ARHGAP29, CCDC18, ABCD3, BCAR3, FNBP1L
150	85	13346 p1	1	99358627	100983836	1,625,209	162 del	1	CCDC76, LRRC39, AGL, HIAT1, CDC14A, PALMD, LPPR5, LPPR4, SAS56, SLC35A3, FRRS1, MIR548D1, DBT, RTCD1
161	86	13346 p1	1	112795166	113471930	676,764	87 del	1	WNT2B, CAPZA1, MOV10, FAM19A3, SLC16A1, PPM1J, RHOC, ST7L, CTTNBP2NL
891	3997	13355 p1	15	27017548	27188573	171,025	12 dup	1	GABRA5, GABRB3
504	736	13726 p1	11	56949367	60233630	3,284,263	348 del	1	DTX4, OR5B21, LPXN, SSRP1, OR5B2, LRRC55, UBE2L6, OR1S1, OR1S2, SERPING1, PRG3, PRG2, MS4A7, P2RX3, MS4A5, MS4A2, MS4A3, MS4A1, YPEL4, OR5A1, MIR130A, OR10Q1, OR4D10, OR4D11, MS4A14, TIMM10, OR10V1, OR5B3, MED19, PATL1, MPEG1, None, OR4D6, OR10W1, MRPL16, OR4D9, OR5B12, OR5B17, MS4A6A, ZFP91-CNTF, TNKS1BP1, TMX2, ZDHHC5, MS4A4A, SLC43A3, SMTNL1, SLC43A1, TMX2-CTNND1, OR9Q2, OSBP, OR9Q1, APLNR, PLACL1, GIF, CLP1, OR5A2, STX3, GLYATL1, TCN1, GLYATL2, GLYAT, OR5AN1, MS4A6E, FAM111A, FAM111B, RTN4RL2
1127	991	13815 p1	16	76311560	76513435	201,875	12 del	1	CNTNAP4
2962	2993	13876 p1	8	36641928	43054712	6,412,784	598 dup	1	IDO2, TM2D2, IDO1, STAR, BAG4, ASH2L, AP3M2, LETM2, IKBKB, ADAM18, ADAM32, C8orf4, HTRA4, HGSNAT, RNFI70, ANK1, ZNF703, CHRN3, DDHD2, PPAPDC1B, WHSC1L1, NKX6-3, GPR124, KCNU1, GOLGA7, MYST3, SGK196, POLB, FNTA, HOOK3, LSM1, PLAT, CHRNA6, BRF2, ADAM2, C8orf40, C8orf86, FGFR1, SLC20A2, THAP1, RAB11FIP1, SFRP1, GOT1L1, GINS4, ERLIN2, TACC1, DKK4, ADAM9, VDACC3, AGPAT6, ADRB3, EIF4EBP1, PLEKHA2, PROSC, ZMAT4
2964	2994	13876 p1	8	43197329	49986892	6,789,563	142 dup	1	None, CEBPD, PRKDC, KIAA0146, MCM4, SNAI2, UBE2V2, EFCAB1, POTEA, C8orf22
1780	1911	11241 s1	2	241615961	241709123	93,162	40 dup	3	AQP12B, AQP12A, KIF1A
1195	6843	12480 s1	17	10347920	10356645	8,725	16 del	1	MYH4
114	93	13447 s1	1	48688408	48705209	16,801	12 dup	1	SLC5A9
3059	6253	13601 s1	9	35649844	35754126	104,282	134 del	1	CA9, CCDC107, MSMP, CREB3, TPM2, TLN1, C9orf100, SIT1, GBA2, RGP1
1643	208	13629 s1	2	61505299	61528589	23,290	14 dup	1	USP34

Table S5

callID	sampleID	chr	start	stop	iM_start	iM_stop	probes	aCHG_mean	iM_PASS	aCGH_pass	Notes
1749	12383.p1	chr1	3413218	3417328			26	0.048704			False Positive
5495	12618.p1	chr1	11134287	11155938			9	-0.762897		YES	
1561	11433.s1	chr1	40204572	40312969			26	0.442855		YES	
1581	11667.s1	chr1	42693553	42744343	42693597	42837441	25	0.404631	YES	YES	
40	11872.p1	chr1	65730593	65831879	65666329	65823558	69	0.466047	YES	YES	
5555	13097.p1	chr1	65849875	65855310			43	0.530438		YES	
1499	11118.p1	chr1	66837995	67000051			30	-0.057537		False Positive	
69	12810.p1	chr1	87029343	87038403	87028669	87038695	72	-0.519149	YES	YES	
8299	13296.s1	chr1	113245184	113264970			25	0.355266		YES	
55	12161.s1	chr1	182550359	182555941	182549019	182564062	43	-0.571282	YES	YES	
37	11715.s1	chr1	185097800	185130057			45	0.507937		YES	
662	11895.s1	chr10	225952	532470			60	0.54439		YES	
3250	12317.s1	chr10	18289595	18331762	18276761	18415963	38	0.401112	YES	YES	
8347	12997.p1	chr10	54527896	54531395	54524658	54537447	26	-0.577977	YES	YES	
8348	12997.s1	chr10	54527896	54531395	54524658	54536839	26	-0.2795	YES	YES	
3175	11267.s1	chr10	72604229	72645686	72601813	72824456	20	-0.629198	YES	YES	
6270	12510.p1	chr10	90703553	90707143			27	0.00287		PPG, ACTA2	
6271	12510.s1	chr10	90703553	90707143			27	0.084178		PPG, ACTA2	
3253	12383.s1	chr10	101594136	101596047			0			FP/Not Tested due to lack of aCGH probes	
3201	11519.p1	chr10	122216817	122349014	121814553	122500375	1	-0.887886	YES	YES	
6334	13162.s1	chr10	135168873	135179599			0			YES	
697	11569.p1	chr11	5862185	5878932			23	-0.261248		YES	
3345	11667.p1	chr11	18727363	18729842			19	-0.433224		YES	
3347	11667.s1	chr11	18727363	18729842			19	-0.224613		YES	
728	12810.p1	chr11	32697110	32781789	32699987	32815580	42	-0.566457	YES	YES	
6402	12836.s1	chr11	59620447	59623531			23	-0.879108		YES	
6401	12836.p1	chr11	59620675	59622308			11	-0.839071		YES	
710	11788.s1	chr11	63057637	63059115			12	0.456157		YES	
7466	11229.p1	chr11	100558409	100859532			108	0.047092			PPG, ARHGAP42L
7715	11115.s1	chr11	100558409	100831720			94	-0.060677			PPG, ARHGAP42L
3346	11667.s1	chr11	108256646	108264105			58	0.369384		YES	
7717	11667.p1	chr11	108256646	108264105			58	0.518831		YES	
6407	12851.s1	chr11	134177017	134257553	134164618	134346119	39	0.494821	YES	YES	
7722	11282.p1	chr12	4651049	4668159			36	0.1066			PPG, RAD51AP1
7723	11519.p1	chr12	4651049	4668159			36	-0.003573			PPG, RAD51AP1
3495	11282.s1	chr12	4655474	4668159			29	0.03143			PPG, RAD51AP1
8313	12997.p1	chr12	25264705	25267804			23	-0.992749		YES	
8314	12997.s1	chr12	25264705	25267804			23	-0.563687		YES	
6506	12836.s1	chr12	49688983	49691056			17	0.150829			Confirmed w/ manual inspection
3463	11090.p1	chr12	49688983	49691056			17	0.274504		YES	
3540	11828.p1	chr12	51203238	51213562			14	-0.075046			PPG, ATF1
774	11895.s1	chr12	109290781	109293251			19	-0.361741		YES	
3489	11241.p1	chr12	120875929	120884632			65	0.508616		YES	
6505	12836.p1	chr12	120875929	120884632	120873726	120888619	65	0.390249	YES	YES	
7477	12011.s1	chr13	20425494	20426320	20420175	20451410	6	0.482955	YES	YES	
3671	11412.p1	chr13	96508411	96515968			59	-0.863395		YES	
3672	11412.s1	chr13	96508411	96515968			59	-0.582202		YES	
6555	12829.s1	chr13	114138154	114175048			18	0.532308		YES	
3661	11196.s1	chr13	115007595	115091756			170	0.201715			False Positive
865	12304.p1	chr14	65016715	65019579			22	0.321413		YES	
877	12810.p1	chr14	65016715	65019579			22	0.399589		YES	
8420	12582.p1	chr14	69919957	69969596	69922552	69978718	23	0.504025	YES	YES	

3782	11118.p1	chr14	99182528	99183611	99181386	99193587	9	-0.694542	YES	
3826	12317.s1	chr14	100402365	100405664	100398318	100407548	25	0.274271	YES	
3954	12317.s1	chr15	40993261	41001314			45	-0.759746	YES	
6651	13018.p1	chr15	43627893	43644143			19	0.140645		PPG, ADAL
906	11229.s1	chr15	76426604	76641077	76394419	76640658	42	0.400188	YES	
6751	13018.p1	chr16	4624972	4642368	4614859	4643452	21	-0.710982	YES	
4043	11118.p1	chr16	4871380	4871598			2	-0.019551		FP/Not Tested due to lack of aCGH probes
6696	12441.p1	chr16	21655610	21763826			25	0.513189	YES	
943	11229.p1	chr16	31477170	31488897	31478711	31489033	12	0.403372	YES	
945	11229.s1	chr16	31477170	31487883	35183650	35284399	11	0.25197	YES	
7863	11412.p1	chr16	84693365	84812688			0			FP/Not Tested due to lack of aCGH probes
1036	12578.p1	chr17	6010	11981	8547	34203	46	0.124371	YES	
4306	11532.p1	chr17	3518630	3561469	3503527	3561396	68	-0.731878	YES	
4276	11267.s1	chr17	7258523	7259951			10	0.094905		False Positive
4295	11484.s1	chr17	29311634	29324349	29301608	29319956	17	0.592237	YES	
4353	12561.s1	chr17	72874459	72877385			24	-0.200099	YES	
6838	12424.p1	chr17	80151631	80153240			13	-0.592442	YES	
4435	11252.p1	chr18	688573	697355	691173	695030	70	-0.693998	YES	
1079	11942.p1	chr18	32953419	32954256			6	-0.445434	YES	
4485	12656.s1	chr18	70416262	70417855	70402753	70412150	11	0.121603	YES	
1132	12161.s1	chr19	3805946	3834981			99	0.136737		Confirmed w/ manual inspection
1131	12161.p1	chr19	3807169	3833776			98	0.206042		Confirmed w/ manual inspection
7052	12837.s1	chr19	3816671	3825405			69	-0.449766	YES	
7006	12252.s1	chr19	11436073	11448075			14	0.008268		PPG, RAB3D
7054	12838.s1	chr19	19030031	19033553			28	0.176093		Confirmed w/ manual inspection
1091	11029.p1	chr19	45494107	45495658			12	-0.270947	YES	
1092	11029.s1	chr19	45494107	45495658			12	-0.421898	YES	
4554	11304.s1	chr19	49481180	49497180			76	0.262305	YES	
7087	13296.s1	chr19	49671173	49675365			32	-0.303428	YES	
7074	13162.s1	chr19	57286055	57293489			58	0.0126		False Positive
5709	13296.s1	chr2	1437209	1479843	1426346	1520676	23	0.44676	YES	
1898	11118.p1	chr2	44527109	44541090	44519142	44545576	18	0.414627	YES	
1985	12228.p1	chr2	74129486	74166149			128	0.007104		False Positive
5682	12851.s1	chr2	96780544	97784254	97845100	98202258	84	0.401272	YES	
5682	12851.s1	chr2	96780544	97784254	96731109	97577661	84	0.401272	YES	
1931	11484.s1	chr2	98263529	98275940			18	0.240511	YES	
7964	11045.p1	chr2	102407181	102416105			70	0.437141	YES	
2016	12552.s1	chr2	113346442	113404739			30	0.008806		False Positive
5683	12851.s1	chr2	192711168	193059250	192711162	193252329	70	-0.634688	YES	
5689	12997.p1	chr2	230632269	230724290	230630874	230712174	135	0.486573	YES	
160	11659.s1	chr2	232070951	232072965			16	-0.557958	YES	
8293	12409.s1	chr20	3734704	3739325			35	0.021807		PPG, C20orf27
4742	12650.s1	chr20	3735043	3739325			33	0.081261		PPG, C20orf27
7992	12383.p1	chr20	3835271	3838456			24	0.270166		
7993	12383.s1	chr20	3835271	3838456	3822148	3830474	24	0.113595	YES	
1194	11013.s1	chr20	7864239	7990962	7549585	8317018	23	0.636108	YES	
1207	11872.p1	chr20	30533609	30538190			36	0.2861	YES	
4712	11196.s1	chr20	62853245	62904953			26	0.368368	YES	
7149	12481.p1	chr21	45724911	45826648			22	0.512218	YES	
7216	12997.p1	chr22	17264508	17288963			12	0.529071	YES	
7191	12481.p1	chr22	21369463	21473538	21374550	21465780	2	-0.040647	YES	
4841	11090.p1	chr22	23401639	24967945	24182500	24999104	401	0.385952	YES	
4841	11090.p1	chr22	23401639	24967945	23648009	24163081	401	0.385952	YES	

8025	11090.p1	chr22	24034217	24325797	24182500	24999104	9	0.526062	YES	YES	
8025	11090.p1	chr22	24034217	24325797	23648009	24163081	9	0.526062	YES	YES	
8024	11090.p1	chr22	24431961	24967945	24182500	24999104	9	0.462384	YES	YES	
8024	11090.p1	chr22	24431961	24967945	23648009	24163081	9	0.462384	YES	YES	
1263	12810.p1	chr22	32545739	32651316	32530256	32703072	75	0.289551	YES	YES	
4893	11828.p1	chr22	42781166	42899582	42788316	42896474	0		YES		
1255	12011.s1	chr22	44883503	44893436			80	0.420071	YES	YES	
8318	12409.s1	chr22	47287161	47308084	47285396	47443386	10	-0.673402	YES	YES	
5766	12588.s1	chr3	7594808	7782093	7516145	7802705	1	-0.841887	YES	YES	
8377	12851.s1	chr3	9867483	9871079			0				FP/Not Tested due to lack of aCGH probes
5761	12481.p1	chr3	12632296	12791331	12633293	12806142	0		YES		
2232	12383.s1	chr3	35833873	35836540	35811601	35938795	12	0.366609	YES	YES	
5756	12252.s1	chr3	81627075	81640315	81594581	81644911	18	-0.836112	YES	YES	
5797	13162.s1	chr3	113588353	113619993	113576614	113619872	15	-0.708712	YES	YES	
237	12304.p1	chr3	132277814	132280061			16	-0.510444	YES	YES	
230	12106.s1	chr3	137781657	137803095			9	0.462725	YES	YES	
2152	11304.s1	chr3	141884463	142084021	141821227	142085310	108	0.430274	YES	YES	
5769	12631.s1	chr3	141889166	142084208	141821227	142065934	108	0.44434	YES	YES	
278	12161.p1	chr4	6594899	6613005	6594947	6611813	25	-0.279563	YES	YES	
279	12161.s1	chr4	6594899	6613005	6594947	6611813	25	-0.454313	YES	YES	
5822	12409.s1	chr4	89978088	90035703	89967292	90149589	28	-0.69108	YES	YES	
2373	12370.p1	chr4	159590764	159616795			13	0.487415	YES	YES	
314	11659.s1	chr5	40931165	40937792	40927429	40943442	53	-0.685554	YES	YES	
2442	11456.s1	chr5	81283389	81354421			35	0.013013		PPG, ATG10	
5890	12851.s1	chr5	81283389	81354421			35	0.001057		PPG, ATG10	
5876	12588.s1	chr5	110427986	110446977			26	0.488264	YES	YES	
2490	12370.p1	chr5	110430617	110446977	110417428	110441533	22	0.550374	YES	YES	
2492	12370.s1	chr5	110430617	110446977			22	0.355151	YES	YES	
336	12161.p1	chr5	121309890	121358102			25	0.260764	YES	YES	
337	12161.s1	chr5	121309890	121362821	121297115	121366608	27	0.424875	YES	YES	
346	12390.s1	chr5	141312823	141314151			9	-0.277458	YES	YES	
2454	11828.p1	chr5	156456743	156479665			11	0.50319	YES	YES	
7512	11013.s1	chr6	49426794	49440571			41	0.415165	YES	YES	
378	11229.p1	chr6	51747890	51752043			33	-0.798963	YES	YES	
394	12106.s1	chr6	56915571	56919661	56917538	56954550	31	0.496498	YES	YES	
2612	11551.p1	chr6	88315634	88318947			26	-0.759539	YES	YES	
5962	12655.p1	chr6	146870599	146875741	146870930	146875322	40	-0.510074	YES	YES	
2655	12650.s1	chr6	162683556	162864505	162632171	163059714	0		YES		
2610	11519.p1	chr6	169617915	169646376	169243539	169795975	0		YES		
465	12304.p1	chr7	98628206	98633339			0				FP/Not Tested due to lack of aCGH probes
470	12578.p1	chr7	150706017	150725697	150705842	150723467	0		YES		
6067	12837.s1	chr7	151833916	152027824	151826268	152055852	1	0.483768	YES	YES	
2918	11196.s1	chr8	190895	382935			116	0.250819	YES	YES	
8490	12631.s1	chr8	27378399	27380025			13	-0.520584	YES	YES	
530	11715.s1	chr8	57078801	57080828	57055054	57098250	16	0.498051	YES	YES	
2922	11252.p1	chr8	144295142	144450815			147	0.374638	YES	YES	
593	12106.s1	chr9	214508	340321	174447	359712	62	0.929548	YES	YES	
6194	12655.p1	chr9	368017	677009	788730	864122	63	0.485291	YES	YES	
6194	12655.p1	chr9	368017	677009	347559	699065	63	0.485291	YES	YES	
597	12161.s1	chr9	5968018	6015607	5966898	6135411	23	0.429288	YES	YES	
6177	12252.s1	chr9	33166755	33261167	33141752	33260632	49	0.519623	YES	YES	
607	12741.s1	chr9	35228011	35237823	35171905	35238095	77	-0.957764	YES	YES	
599	12578.p1	chr9	35662942	35664489			12	-0.196435	YES	YES	

6191	12637.p1	chr9	35662942	35664489			12	-0.457227	YES	
6201	12829.s1	chr9	35662942	35664489			12	-0.39486	YES	
6233	13296.s1	chr9	125562401	125589066			13	0.484615	YES	

Table S6

False Positive Rate							
False Positive Rates							
	False Pos.	Validated	Total Tested	FPR	Chi Square/Fisher Exact	P value	
Dataset	lossifov et al.	1	51	0.02			
	O'Roak et al.	0	44	0.00	$\chi^2 = 7.26$	0.026	
	Sanders et al.	6	53	0.10			
Affected	Probands	3	68	0.04	OR=0.882	1	
	Siblings	4	80	0.05			
Size	2 exons	2	26	0.07			
	3+ exons	5	122	0.04	OR=0.533	0.611	
Del/Dup	Deletions	1	54	0.02			
	Duplications	6	89	0.06	OR=0.275	0.432	
Total		7	148	0.05			
False Negative Rate							
	Missed	Found	Total	FNR	Chi Square/Fisher Exact	P value	
Dataset	lossifov et al.	76	111	0.41			
	O'Roak et al.	21	77	0.21	$\chi^2 = 22.3$	< 1*10 ⁻⁴	
	Sanders et al.	55	201	0.21			
Offspring	Probands	93	220	0.29			
	Siblings	59	169	0.26	OR=1.21	0.355	
Size	2 exons	39	24	0.62			
	3+ exons	113	365	0.24	OR=0.19	< 1*10 ⁻⁸	
Total		152	389	0.28			

Table S7

Sample	SRS Score	FSIQ	SRS Discordant	Sex	# of rare CNVs	# of de novo SNVs	# of de novo CNVs
11000.p1	70	65	FALSE	male	1	0	0
11000.s1	48		FALSE	female	0	0	0
11008.p1	65	129	FALSE	male	1	1	0
11008.s1			FALSE	male	1	0	0
11010.p1	76	65	TRUE	male	0	0	0
11010.s1	40		TRUE	male	0	0	0
11013.p1	90	132	TRUE	male	1	0	0
11013.s1	47		TRUE	male	2	0	0
11014.p1	75	148	FALSE	male	1	0	0
11014.s1	48		FALSE	male	0	0	0
11029.p1	90	29	TRUE	female	1	0	0
11029.s1	41		TRUE	male	1	0	0
11045.p1	60	79	FALSE	female	1	0	0
11045.s1	53		FALSE	male	0	0	0
11057.p1	64	89	FALSE	male	0	0	0
11057.s1	34		FALSE	male	1	0	0
11060.p1	66	100	FALSE	male	1	0	0
11060.s2	40		FALSE	male	0	0	0
11066.p1	82	88	TRUE	male	1	0	0
11066.s2	37		TRUE	male	0	0	0
11067.p1	85	130	TRUE	male	1	0	0
11067.s1	40		TRUE	female	2	0	0
11074.p1	63	68	FALSE	male	0	1	0
11074.s1	35		FALSE	male	0	0	0
11075.p1	67	39	FALSE	male	2	0	0
11075.s1	36		FALSE	male	2	0	0
11077.p1	79	33	TRUE	male	0	1	0
11077.s1	37		TRUE	male	0	0	0
11079.p1	83	53	TRUE	male	0	0	1
11079.s1	42		TRUE	female	0	0	0
11085.p1	82	61	TRUE	male	0	0	0
11085.s1	39		TRUE	female	1	0	0
11089.p1	84	40	TRUE	male	0	0	0
11089.s1	57		TRUE	male	0	0	0
11090.p1	83	56	TRUE	male	4	0	1
11090.s1	42		TRUE	male	0	0	0
11092.p1	76	109	TRUE	male	1	0	0
11092.s1	35		TRUE	female	0	0	0

11094.p1	76	87	TRUE	male		2	0	0
11094.s1	46		TRUE	male		0	0	0
11107.p1	90	30	TRUE	male		3	0	0
11107.s1	39		TRUE	male		2	0	0
11108.p1	75	104	FALSE	male		2	0	0
11108.s1	51		FALSE	male		1	0	0
11114.p1	90	40	TRUE	female		1	1	0
11114.s1	39		TRUE	female		1	0	0
11115.p1	90	37	TRUE	female		1	0	0
11115.s1	50		TRUE	female		1	0	0
11117.p1	89	121	TRUE	female		0	0	0
11117.s1	41		TRUE	female		1	0	0
11118.p1	80	93	TRUE	female		4	0	0
11118.s1	51		TRUE	male		1	0	0
11132.p1	90	47	FALSE	male		0	1	0
11132.s1	64		FALSE	male		0	0	0
11146.p1	90	85	TRUE	male		2	0	0
11146.s1	40		TRUE	male		2	0	0
11154.p1	90	90	TRUE	male		0	0	1
11154.s1	45		TRUE	female		1	0	0
11172.p1	71	63	FALSE	male		0	0	0
11172.s1	40		FALSE	female		0	0	0
11180.p1		27	FALSE	female		2	0	0
11180.s1	43		FALSE	male		3	0	0
11190.p1	80	69	TRUE	male		1	0	0
11190.s1	55		TRUE	female		2	0	0
11196.p1	80	112	TRUE	male		5	0	0
11196.s1	53		TRUE	male		3	0	0
11203.p1	84	86	TRUE	female		0	0	0
11203.s1	53		TRUE	female		2	0	0
11219.p1	51	99	FALSE	male		2	0	0
11219.s1	36		FALSE	female		1	0	0
11220.p1	75	80	FALSE	female		1	0	0
11220.s1	41		FALSE	female		3	0	0
11229.p1	73	63	FALSE	male		3	0	0
11229.s1	48		FALSE	male		2	0	0
11241.p1	90	76	TRUE	male		1	0	0
11241.s1	38		TRUE	male		1	0	1
11242.p1	79	94	TRUE	male		1	0	0
11242.s1	48		TRUE	male		1	0	0

11247.p1	87	128	TRUE	male	1	0	0
11247.s1	56		TRUE	female	0	0	0
11252.p1	86	78	FALSE	male	2	0	0
11252.s1	62		FALSE	male	2	0	0
11265.p1	90	106	TRUE	male	0	0	1
11265.s1	47		TRUE	female	0	0	0
11267.p1	90	85	TRUE	female	2	0	0
11267.s1	54		TRUE	male	2	0	0
11282.p1	89	75	TRUE	male	1	0	0
11282.s1	42		TRUE	male	1	0	0
11285.p1		13	FALSE	male	0	0	0
11285.s1	44		FALSE	male	1	0	0
11290.p1	65	119	FALSE	male	0	0	0
11290.s1	39		FALSE	female	0	0	0
11291.p1	68	86	FALSE	male	0	1	0
11291.s1	51		FALSE	female	0	0	0
11298.p1	90	141	TRUE	male	3	0	0
11298.s1	37		TRUE	male	2	0	0
11301.p1	82	101	TRUE	male	1	0	0
11301.s1	44		TRUE	female	1	0	0
11304.p1	90	53	TRUE	male	1	0	0
11304.s1	51		TRUE	female	2	0	0
11316.p1	90	47	TRUE	female	1	0	0
11316.s1	54		TRUE	female	0	0	0
11336.p1	81	123	TRUE	male	3	0	0
11336.s1	45		TRUE	female	3	0	0
11353.p1	90	79	TRUE	female	1	0	1
11353.s1	40		TRUE	male	0	0	0
11356.p1	90	72	TRUE	female	3	1	0
11356.s1	42		TRUE	male	1	0	0
11364.p1	72	106	FALSE	male	1	0	0
11364.s1	40		FALSE	female	3	0	0
11382.p1	90	76	FALSE	male	0	0	0
11382.s1	63		FALSE	female	0	1	0
11390.p1	90	66	TRUE	female	0	0	0
11390.s1	45		TRUE	female	0	0	0
11411.p1	90	61	TRUE	male	0	0	0
11411.s1	41		TRUE	female	1	0	0
11412.p1	70	107	FALSE	male	2	0	0
11412.s1	41		FALSE	female	1	0	0

11429.p1	72	99	FALSE	male		1	0	0
11429.s1			FALSE	male		2	0	0
11433.p1	89	78	TRUE	male		1	0	1
11433.s1	45		TRUE	female		1	0	0
11437.p1	75	82	FALSE	male		1	0	0
11437.s1	35		FALSE	male		1	0	0
11452.p1	90	80	TRUE	male		0	1	0
11452.s1	35		TRUE	male		1	0	0
11456.p1	79	75	TRUE	male		0	0	0
11456.s1	54		TRUE	male		1	0	0
11459.p1	90	80	TRUE	male		2	0	0
11459.s1	42		TRUE	male		2	0	0
11462.p1	74	114	FALSE	male		0	0	0
11462.s1	50		FALSE	male		0	0	0
11469.p1	79	109	TRUE	male		1	0	0
11469.s1	49		TRUE	female		1	0	0
11472.p1	90	30	TRUE	female		2	0	0
11472.s1	41		TRUE	female		1	0	0
11474.p1	89	116	FALSE	male		1	0	0
11474.s1			FALSE	male		1	0	0
11479.p1	79	133	TRUE	male		3	0	0
11479.s1	46		TRUE	female		2	0	0
11484.p1	76	106	TRUE	male		1	0	0
11484.s1	43		TRUE	male		2	0	0
11490.p1	83	84	TRUE	male		0	0	0
11490.s1	50		TRUE	female		0	1	0
11491.p1	74	53	FALSE	male		0	0	0
11491.s1	42		FALSE	male		0	0	0
11501.p1	64	78	FALSE	male		2	0	0
11501.s1	48		FALSE	male		2	0	0
11509.p1	90	80	FALSE	male		0	0	0
11509.s1			FALSE	male		2	0	0
11519.p1	78	50	TRUE	male		3	0	0
11519.s1	47		TRUE	female		1	0	0
11524.p1	74	113	FALSE	male		0	1	0
11524.s1	40		FALSE	male		0	0	0
11532.p1	77	59	TRUE	male		1	0	1
11532.s1	39		TRUE	female		0	0	0
11551.p1	90	98	TRUE	male		1	0	0
11551.s1	39		TRUE	female		0	0	0

11561.p1	71	109	FALSE	male		1	0	0
11561.s1	42		FALSE	male		1	0	0
11569.p1	90	59	TRUE	female		1	0	0
11569.s1	45		TRUE	male		1	0	0
11571.p1	90	100	TRUE	male		1	0	0
11571.s1	51		TRUE	female		0	0	0
11581.p1	78	64	TRUE	male		3	0	0
11581.s1	50		TRUE	male		1	0	0
11610.p1	82	127	FALSE	male		1	1	0
11610.s1	60		FALSE	male		2	0	0
11611.p1	90	32	TRUE	female		1	0	0
11611.s1	34		TRUE	male		1	0	0
11622.p1	90	97	TRUE	male		3	0	0
11622.s1	50		TRUE	female		3	0	0
11629.p1	90	50	TRUE	male		3	0	0
11629.s1	47		TRUE	female		2	0	0
11638.p1	90	55	TRUE	male		0	0	0
11638.s1	46		TRUE	male		0	0	0
11641.p1	80	93	TRUE	male		0	0	0
11641.s1	41		TRUE	male		0	0	0
11654.p1	89	40	TRUE	female		1	0	0
11654.s1	57		TRUE	female		0	0	0
11659.p1	74	88	FALSE	female		1	0	0
11659.s1	55		FALSE	female		2	0	0
11667.p1	78	53	TRUE	male		2	0	0
11667.s1	45		TRUE	male		3	0	0
11676.p1	90	78	TRUE	female		2	0	0
11676.s1	40		TRUE	female		1	0	0
11691.p1	90	59	TRUE	male		0	0	0
11691.s1	47		TRUE	male		0	0	0
11696.p1	71	95	FALSE	male		2	0	0
11696.s1	56		FALSE	male		0	0	0
11700.p1	65	88	FALSE	male		1	0	0
11700.s1	41		FALSE	female		0	0	0
11711.p1	88	94	TRUE	male		2	0	0
11711.s1	48		TRUE	male		2	0	0
11715.p1	90	96	TRUE	male		1	1	0
11715.s1	40		TRUE	female		2	0	0
11716.p1	90	49	TRUE	male		4	0	0
11716.s1	42		TRUE	male		1	0	0

11720.p1	59	68	FALSE	male		0	0	0
11720.s1	39		FALSE	female		0	0	0
11722.p1	81	97	TRUE	male		2	0	0
11722.s1	42		TRUE	male		0	0	0
11724.p1	84	59	TRUE	male		1	0	0
11724.s1	35		TRUE	female		0	0	0
11740.p1	75	98	FALSE	male		1	0	0
11740.s1	42		FALSE	male		0	0	0
11766.p1	57	104	FALSE	male		1	0	0
11766.s1	42		FALSE	female		1	1	0
11773.p1	90	43	TRUE	male		1	0	0
11773.s1	42		TRUE	male		1	0	0
11788.p1	84	84	TRUE	male		1	0	0
11788.s1	41		TRUE	male		1	0	0
11797.p1	76	118	TRUE	male		1	0	0
11797.s1	51		TRUE	female		2	0	0
11808.p1	71	84	FALSE	female		1	0	0
11808.s1	36		FALSE	female		0	0	0
11809.p1	89	92	TRUE	male		0	0	0
11809.s1	53		TRUE	female		0	0	0
11810.p1	90	92	FALSE	male		4	0	0
11810.s1	86		FALSE	female		2	1	0
11824.p1	79	81	FALSE	male		1	0	0
11824.s1	68		FALSE	male		1	0	0
11828.p1	88	74	TRUE	male		3	0	0
11828.s1	41		TRUE	female		2	0	0
11872.p1	88	62	TRUE	female		2	1	0
11872.s1	49		TRUE	female		0	0	0
11892.p1	70	55	FALSE	male		0	1	0
11892.s1	41		FALSE	female		0	0	0
11894.p1	70	114	FALSE	male		1	0	0
11894.s1	40		FALSE	male		0	0	0
11895.p1	70	86	FALSE	male		1	0	0
11895.s1	47		FALSE	male		2	0	0
11905.p1	90	53	TRUE	male		1	0	0
11905.s1	34		TRUE	male		0	0	0
11942.p1	65	50	FALSE	male		1	0	0
11942.s1	36		FALSE	male		0	0	0
11959.p1	70	64	FALSE	male		0	0	0
11959.s1	44		FALSE	female		1	0	0

11962.p1	75	81	FALSE	male	0	0	0
11962.s1	49		FALSE	male	0	0	0
11964.p1	90	40	TRUE	female	2	0	0
11964.s1	44		TRUE	female	0	0	0
12011.p1	82	82	TRUE	male	2	0	0
12011.s1	43		TRUE	male	2	0	0
12051.p1	90	62	TRUE	male	0	0	0
12051.s1	39		TRUE	male	0	0	0
12100.p1	90	71	TRUE	male	2	0	1
12100.s1	38		TRUE	female	1	0	0
12106.p1	88	112	TRUE	male	3	0	0
12106.s1	56		TRUE	female	3	0	0
12152.p1	87	114	TRUE	male	0	0	0
12152.s1	42		TRUE	male	1	0	0
12153.p1	67	60	FALSE	male	0	0	0
12153.s1	38		FALSE	female	0	1	0
12161.p1	55	106	FALSE	female	3	2	0
12161.s1	44		FALSE	female	5	0	0
12162.p1	77	67	TRUE	male	1	0	1
12162.s1	35		TRUE	female	0	0	0
12175.p1	83	71	TRUE	male	1	0	0
12175.s1	41		TRUE	female	1	0	0
12187.p1	67	109	FALSE	male	0	0	0
12187.s1	42		FALSE	female	0	0	0
12224.p1	89	80	TRUE	male	1	1	1
12224.s1	41		TRUE	male	1	0	0
12228.p1	83	108	TRUE	male	1	0	0
12228.s1	54		TRUE	female	0	0	0
12233.p1	51	106	FALSE	male	0	0	0
12233.s1	36		FALSE	female	0	0	0
12235.p1	76	79	TRUE	male	0	0	0
12235.s1	36		TRUE	male	0	0	0
12241.p1	90	72	TRUE	female	0	0	0
12241.s1	41		TRUE	male	0	0	0
12243.p1	83	97	TRUE	male	0	0	0
12243.s1	44		TRUE	male	0	0	0
12252.p1	74	87	FALSE	male	2	0	0
12252.s1	41		FALSE	male	3	0	0
12285.p1	75	75	FALSE	female	2	0	0
12285.s1	47		FALSE	male	2	0	0

12295.p1	90	104	FALSE	male		1	0	0	0
12295.s1			FALSE	male		1	0	0	0
12297.p1	89	97	TRUE	male		1	0	0	0
12297.s1	45		TRUE	male		2	0	0	0
12301.p1		119	FALSE	male		0	0	0	0
12301.s1	59		FALSE	female		0	0	0	0
12303.p1	90	79	TRUE	male		2	0	0	0
12303.s1	41		TRUE	female		0	0	0	0
12304.p1	65	83	FALSE	male		3	0	0	0
12304.s1	39		FALSE	male		2	0	0	0
12308.p1	90	105	TRUE	female		3	0	0	1
12308.s1	39		TRUE	male		0	0	0	0
12313.p1	90	115	TRUE	female		1	0	0	0
12313.s1	38		TRUE	female		1	0	0	0
12317.p1	53	91	FALSE	male		2	0	0	0
12317.s1	46		FALSE	female		3	0	0	0
12321.p1	80	96	TRUE	male		0	0	0	0
12321.s1	44		TRUE	female		0	0	0	0
12327.p1	82	97	TRUE	male		0	0	0	0
12327.s1	53		TRUE	female		0	0	0	0
12334.p1	61	84	FALSE	male		6	0	0	0
12334.s1	48		FALSE	male		3	2	0	0
12340.p1	90	26	TRUE	female		1	1	0	0
12340.s1	39		TRUE	female		0	1	0	0
12343.p1		30	FALSE	female		0	0	1	1
12343.s1	39		FALSE	female		0	0	0	0
12345.p1	82	91	TRUE	male		0	0	0	0
12345.s1	42		TRUE	female		0	0	0	0
12358.p1	90	36	TRUE	female		1	0	0	0
12358.s1	39		TRUE	male		0	0	0	0
12360.p1	90	104	TRUE	male		0	0	0	0
12360.s1	41		TRUE	female		0	0	0	0
12368.p1	81	47	TRUE	male		1	0	0	0
12368.s1	39		TRUE	male		1	0	0	0
12370.p1	80	130	TRUE	male		2	0	0	0
12370.s1	59		TRUE	female		1	0	0	0
12373.p1	90		TRUE	male		2	0	0	0
12373.s1	52		TRUE	female		1	0	0	0
12375.p1	89	94	TRUE	male		1	0	0	0
12375.s1	46		TRUE	female		1	0	0	0

12383.p1	80	101	FALSE	male		2	0	0
12383.s1	72		FALSE	male		3	0	0
12390.p1	90	86	TRUE	male		0	0	0
12390.s1	36		TRUE	male		1	0	0
12394.p1	77	88	TRUE	male		2	0	0
12394.s1	36		TRUE	male		2	0	0
12396.p1	90	99	TRUE	male		2	0	0
12396.s1	51		TRUE	female		1	0	0
12403.p1	90	104	TRUE	male		1	0	0
12403.s1	39		TRUE	female		0	0	0
12409.p1	77	107	TRUE	male		2	1	0
12409.s1	44		TRUE	female		3	0	0
12412.p1	90	98	TRUE	male		0	0	0
12412.s1	55		TRUE	male		0	0	0
12420.p1	48	131	FALSE	male		3	0	0
12420.s1	35		FALSE	female		3	0	0
12424.p1	63	69	FALSE	male		1	0	0
12424.s1	45		FALSE	female		0	0	0
12438.p1	87	75	TRUE	male		0	1	0
12438.s1	48		TRUE	male		0	0	0
12441.p1	90	28	TRUE	male		1	0	0
12441.s1	45		TRUE	female		0	0	0
12445.p1	81	104	TRUE	male		1	0	0
12445.s1	34		TRUE	male		1	0	0
12460.p1	83	68	TRUE	male		1	0	0
12460.s1	42		TRUE	male		0	0	0
12462.p1	76	112	TRUE	male		0	0	0
12462.s1	36		TRUE	female		1	1	0
12463.p1	83	84	TRUE	male		3	1	0
12463.s1	45		TRUE	female		0	0	0
12467.p1	70	102	FALSE	male		1	0	0
12467.s1	47		FALSE	female		1	0	0
12473.p1	88	49	TRUE	male		2	0	0
12473.s1	37		TRUE	female		0	0	0
12480.p1	90	86	TRUE	male		1	0	0
12480.s1	36		TRUE	male		0	0	1
12481.p1	63	42	FALSE	male		3	0	0
12481.s1	34		FALSE	male		2	0	0
12498.p1	90	65	TRUE	male		2	0	0
12498.s1	38		TRUE	female		2	0	0

12507.p1	89	80	FALSE	male		2	0	0
12507.s1			FALSE	female		2	0	0
12510.p1	90	52	TRUE	male		1	0	0
12510.s1	45		TRUE	male		1	0	0
12512.p1	75	85	FALSE	male		0	0	0
12512.s1	38		FALSE	female		2	0	0
12515.p1	80	104	TRUE	female		0	0	0
12515.s1	46		TRUE	male		0	0	0
12518.p1	62	88	FALSE	male		1	0	0
12518.s1	54		FALSE	female		3	0	0
12522.p1	81	68	TRUE	male		0	0	0
12522.s1	41		TRUE	male		0	0	0
12523.p1	49	91	FALSE	male		2	0	0
12523.s1	56		FALSE	female		1	0	0
12524.p1	80	146	TRUE	female		1	0	0
12524.s1	50		TRUE	female		0	0	0
12526.p1	82	103	TRUE	male		0	0	0
12526.s1	36		TRUE	female		1	0	0
12534.p1	90	81	TRUE	female		3	0	0
12534.s1	42		TRUE	female		0	0	0
12536.p1	87	85	FALSE	male		0	0	0
12536.s1	75		FALSE	female		0	0	0
12552.p1	90	104	TRUE	male		0	0	0
12552.s1	54		TRUE	male		1	0	0
12561.p1	73	102	FALSE	male		2	1	0
12561.s1	50		FALSE	female		1	0	0
12578.p1	79	81	TRUE	male		3	0	0
12578.s1	42		TRUE	female		1	0	0
12579.p1	90	33	TRUE	male		2	0	0
12579.s1	43		TRUE	female		2	0	0
12581.p1	90	34	TRUE	female		1	0	0
12581.s1	37		TRUE	male		1	0	0
12582.p1	90	57	TRUE	male		1	0	0
12582.s1	41		TRUE	male		0	0	0
12588.p1	85	106	TRUE	male		1	0	0
12588.s1	47		TRUE	male		2	1	0
12616.p1	63	111	FALSE	male		1	0	0
12616.s1	38		FALSE	male		1	0	0
12618.p1	90	106	TRUE	male		1	0	0
12618.s1	42		TRUE	male		1	0	0

12620.p1	72	83	FALSE	male		0	0	0
12620.s1	46		FALSE	female		0	0	0
12626.p1	87	92	TRUE	male		2	0	0
12626.s1	45		TRUE	female		2	0	0
12628.p1	90	118	TRUE	male		1	0	0
12628.s1	54		TRUE	female		1	0	0
12630.p1	90	129	TRUE	male		0	0	0
12630.s1	41		TRUE	female		1	0	0
12631.p1	78	80	FALSE	male		2	0	0
12631.s1			FALSE	male		2	0	0
12633.p1	64	80	FALSE	male		0	0	0
12633.s1	37		FALSE	female		0	0	0
12637.p1	72	105	FALSE	male		1	0	0
12637.s1	40		FALSE	male		1	0	0
12638.p1	59	73	FALSE	male		0	0	0
12638.s1	42		FALSE	female		0	0	0
12642.p1	88	99	TRUE	male		0	0	0
12642.s1	43		TRUE	male		0	0	0
12644.p1	90	106	TRUE	male		1	0	0
12644.s1	39		TRUE	male		1	0	0
12645.p1	90	86	TRUE	male		1	1	0
12645.s1	36		TRUE	male		0	0	0
12647.p1	88	72	TRUE	male		2	0	0
12647.s1	40		TRUE	male		1	0	0
12650.p1	79	104	FALSE	male		1	0	0
12650.s1	63		FALSE	female		2	0	0
12651.p1	85	24	TRUE	male		0	0	0
12651.s1	38		TRUE	male		1	0	0
12652.p1	90	73	TRUE	male		0	1	0
12652.s1	45		TRUE	female		1	0	0
12653.p1	85	77	TRUE	male		0	2	0
12653.s1	38		TRUE	female		0	0	0
12655.p1	76	48	TRUE	male		1	0	0
12655.s1	35		TRUE	male		1	0	0
12656.p1	83	40	TRUE	male		2	0	0
12656.s1	35		TRUE	male		1	0	0
12657.p1	66	80	FALSE	female		0	0	0
12657.s1	39		FALSE	female		0	0	0
12664.p1	89	93	TRUE	male		0	0	0
12664.s1	35		TRUE	male		0	0	0

12680.p1	77	47	TRUE	male		1	0	0	0
12680.s1	47		TRUE	male		0	0	0	0
12683.p1	66	89	FALSE	male		1	1	0	0
12683.s1	36		FALSE	male		1	0	0	0
12685.p1	87	100	TRUE	male		0	1	0	0
12685.s1	44		TRUE	female		0	0	0	0
12688.p1	71	116	FALSE	male		0	0	0	0
12688.s1	36		FALSE	male		0	0	0	0
12690.p1	90	110	TRUE	male		1	0	0	0
12690.s1	40		TRUE	male		0	0	0	0
12691.p1	73	34	FALSE	female		2	0	0	1
12691.s1			FALSE	male		1	0	0	0
12697.p1	80	85	TRUE	male		3	0	0	0
12697.s1	46		TRUE	female		0	0	0	0
12703.p1	72	59	FALSE	male		0	0	0	0
12703.s1	38		FALSE	female		0	0	0	0
12705.p1	75	99	FALSE	male		1	1	0	0
12705.s1	38		FALSE	male		1	0	0	0
12708.p1	90	79	TRUE	male		0	0	0	0
12708.s1	41		TRUE	female		0	0	0	0
12716.p1	88	112	TRUE	male		1	0	0	0
12716.s1	56		TRUE	female		1	0	0	0
12719.p1	90	46	TRUE	male		1	0	0	0
12719.s1	48		TRUE	female		1	0	0	0
12720.p1	59	69	FALSE	male		0	0	0	0
12720.s1	49		FALSE	male		0	0	0	0
12723.p1	82	48	TRUE	male		1	0	0	0
12723.s1	57		TRUE	female		1	0	0	0
12724.p1	87	83	TRUE	female		0	0	0	0
12724.s1	49		TRUE	male		0	0	0	0
12727.p1	78	90	TRUE	male		1	0	0	0
12727.s1	58		TRUE	male		2	1	0	0
12729.p1	90	74	TRUE	female		2	0	0	0
12729.s1	37		TRUE	male		2	0	0	0
12733.p1	66	90	FALSE	male		1	0	0	0
12733.s1	45		FALSE	female		2	0	0	0
12735.p1	90	55	TRUE	male		2	0	0	1
12735.s1	39		TRUE	male		0	0	0	0
12736.p1	85	101	TRUE	male		1	0	0	1
12736.s1	53		TRUE	female		1	0	0	0

12739.p1	82	98	TRUE	male		0	0	0	0
12739.s1	41		TRUE	male		0	0	0	0
12741.p1	86	74	TRUE	male		2	0	0	0
12741.s1	38		TRUE	male		1	0	0	0
12743.p1	78	116	FALSE	male		1	0	0	0
12743.s1	62		FALSE	male		0	0	0	0
12748.p1	90	92	TRUE	male		0	0	0	0
12748.s1	43		TRUE	male		0	0	0	0
12758.p1	80	74	TRUE	male		1	0	0	0
12758.s1	40		TRUE	female		0	0	0	0
12759.p1	62	105	FALSE	male		0	0	0	0
12759.s1	39		FALSE	female		0	0	0	0
12763.p1	90	37	TRUE	male		1	0	0	0
12763.s1	42		TRUE	male		0	0	0	0
12764.p1	90	94	TRUE	female		0	1	0	0
12764.s1	40		TRUE	female		0	0	0	0
12770.p1	80	122	TRUE	male		1	0	0	0
12770.s1	47		TRUE	male		1	0	0	0
12780.p1	79	115	TRUE	male		1	0	0	0
12780.s1	49		TRUE	male		2	0	0	0
12790.p1	84	31	TRUE	male		1	0	0	0
12790.s1	41		TRUE	female		0	0	0	0
12802.p1	90		TRUE	male		1	0	0	0
12802.s1	35		TRUE	male		1	0	0	0
12810.p1	90	63	TRUE	male		4	0	0	0
12810.s1	38		TRUE	female		1	0	0	0
12826.p1	90	66	TRUE	female		3	1	0	0
12826.s1	42		TRUE	male		2	0	0	0
12829.p1	72	133	FALSE	male		1	0	0	0
12829.s1	70		FALSE	male		2	0	0	0
12833.p1	76	65	TRUE	male		1	0	0	0
12833.s1	49		TRUE	female		1	0	0	0
12836.p1	65	127	FALSE	male		2	0	0	0
12836.s1	56		FALSE	male		2	0	0	0
12837.p1	86	89	TRUE	male		4	0	0	0
12837.s1	57		TRUE	female		2	0	0	0
12838.p1	80	78	TRUE	male		0	0	0	0
12838.s1	41		TRUE	female		1	1	0	0
12840.p1	90	54	TRUE	male		2	2	0	0
12840.s1	39		TRUE	female		2	0	0	0

12843.p1	72	93	FALSE	male		1	0	0
12843.s1	45		FALSE	female		0	0	0
12851.p1	90	37	TRUE	male		2	0	0
12851.s1	41		TRUE	male		5	0	0
12852.p1	78	77	TRUE	male		0	0	0
12852.s1	44		TRUE	male		0	0	0
12869.p1	90	31	TRUE	female		3	0	0
12869.s1	36		TRUE	female		1	0	0
12905.p1	90	66	TRUE	male		0	0	0
12905.s1	41		TRUE	male		0	0	0
12906.p1	90	61	TRUE	male		1	0	0
12906.s1	40		TRUE	female		1	0	0
12937.p1	90	35	TRUE	male		0	0	0
12937.s1	46		TRUE	male		2	0	0
12958.p1	57	88	FALSE	male		0	0	0
12958.s1	43		FALSE	male		0	0	0
12962.p1	79	79	TRUE	male		0	0	0
12962.s1	37		TRUE	female		1	0	0
12975.p1	68	89	FALSE	male		0	0	1
12975.s1	41		FALSE	female		0	0	0
12984.p1	82	100	TRUE	male		0	0	0
12984.s1	46		TRUE	male		0	0	0
12997.p1	81	97	TRUE	male		4	0	0
12997.s1	40		TRUE	female		2	0	0
13000.p1	79	38	TRUE	male		0	0	0
13000.s1	37		TRUE	female		1	0	0
13016.p1	52	98	FALSE	male		0	0	0
13016.s1	40		FALSE	female		0	0	0
13018.p1	76	78	TRUE	male		2	1	1
13018.s1	41		TRUE	male		1	0	0
13048.p1	90	35	TRUE	female		1	0	0
13048.s1	40		TRUE	female		1	0	0
13063.p1	74	49	FALSE	male		1	0	0
13063.s1	41		FALSE	female		1	0	0
13073.p1	90	43	TRUE	male		0	0	0
13073.s1	34		TRUE	male		0	0	0
13094.p1	67	71	FALSE	male		2	1	0
13094.s1	45		FALSE	male		1	0	0
13096.p1	85	107	TRUE	male		0	1	0
13096.s1	39		TRUE	male		1	0	0

13097.p1	73	34	FALSE	male		1	0	0
13097.s1	40		FALSE	female		3	0	0
13099.p1	86	92	TRUE	male		1	0	0
13099.s1	38		TRUE	male		1	0	0
13101.p1	90	94	TRUE	female		0	0	0
13101.s1	51		TRUE	female		1	0	0
13104.p1	87	97	TRUE	male		0	0	0
13104.s1	44		TRUE	female		0	0	0
13116.p1	90	35	TRUE	female		3	0	0
13116.s1	42		TRUE	female		3	0	0
13120.p1	90	113	TRUE	male		0	0	0
13120.s1	38		TRUE	male		0	0	0
13125.p1	86	97	TRUE	male		2	1	0
13125.s1	58		TRUE	female		2	0	0
13129.p1	78	31	TRUE	male		1	0	0
13129.s1	35		TRUE	female		2	0	0
13131.p1	80	75	TRUE	male		0	0	0
13131.s1	45		TRUE	male		0	1	0
13139.p1	73	127	FALSE	male		2	0	0
13139.s1	39		FALSE	female		1	0	0
13144.p1	77	112	TRUE	male		1	0	0
13144.s1	53		TRUE	male		1	1	0
13146.p1	86	106	TRUE	male		0	0	0
13146.s1	37		TRUE	female		0	0	0
13148.p1	78	52	TRUE	male		2	0	0
13148.s1	51		TRUE	female		2	0	0
13152.p1	90	77	TRUE	male		0	0	0
13152.s1	41		TRUE	female		0	0	0
13153.p1	76	75	TRUE	male		2	0	0
13153.s1	34		TRUE	male		1	0	0
13154.p1	80	40	TRUE	male		0	0	0
13154.s1	36		TRUE	male		0	0	0
13159.p1	75	79	FALSE	male		0	0	0
13159.s1	48		FALSE	male		0	0	0
13162.p1	90	74	TRUE	male		3	1	0
13162.s1	36		TRUE	female		3	0	0
13165.p1	90	51	TRUE	male		0	0	0
13165.s1	49		TRUE	male		1	0	0
13166.p1	90	46	TRUE	male		0	0	0
13166.s1	39		TRUE	male		1	0	0

13168.p1	70	104	FALSE	female		0	1	0
13168.s1	45		FALSE	female		0	0	0
13169.p1	49	45	FALSE	male		0	0	0
13169.s1	40		FALSE	male		3	0	0
13171.p1	90	51	TRUE	female		1	0	0
13171.s1	40		TRUE	female		0	0	0
13174.p1	86	106	TRUE	male		0	0	0
13174.s1	38		TRUE	male		0	0	0
13176.p1	90	137	TRUE	female		1	1	0
13176.s1	42		TRUE	female		2	0	0
13183.p1	90	60	TRUE	male		1	1	0
13183.s1	42		TRUE	female		1	0	0
13187.p1	75	84	FALSE	male		0	0	0
13187.s1	39		FALSE	male		0	0	0
13188.p1	82	78	FALSE	male		0	0	0
13188.s1			FALSE	female		0	0	0
13193.p1	90	81	TRUE	male		0	0	0
13193.s1	39		TRUE	male		0	0	0
13195.p1	89	76	TRUE	male		1	0	0
13195.s1	52		TRUE	male		1	0	0
13196.p1	90	92	TRUE	male		1	0	0
13196.s1	55		TRUE	male		1	0	0
13197.p1	79	65	TRUE	male		0	1	0
13197.s1	42		TRUE	female		1	0	0
13215.p1	90	74	TRUE	male		2	0	0
13215.s1	51		TRUE	female		2	0	0
13216.p1	90	105	TRUE	male		1	0	0
13216.s1	50		TRUE	male		2	0	0
13218.p1	66	98	FALSE	male		0	0	0
13218.s1	37		FALSE	male		0	0	0
13227.p1	86	104	TRUE	male		0	0	0
13227.s1	39		TRUE	female		1	0	0
13239.p1	64	83	FALSE	female		0	0	0
13239.s1	41		FALSE	male		1	0	0
13258.p1	90	28	TRUE	male		0	0	0
13258.s1	50		TRUE	female		0	0	0
13263.p1	73	81	FALSE	male		0	0	0
13263.s1	37		FALSE	female		1	0	0
13266.p1	90	87	FALSE	male		0	0	0
13266.s1	68		FALSE	female		0	0	0

13269.p1	69	97	FALSE	male	0	0	0
13269.s1	40		FALSE	male	1	0	0
13271.p1	90	60	TRUE	male	1	0	0
13271.s1	47		TRUE	female	0	0	0
13293.p1	90	83	TRUE	male	2	0	0
13293.s1	41		TRUE	female	2	0	0
13296.p1	87	30	TRUE	male	4	0	0
13296.s1	40		TRUE	female	4	0	0
13307.p1	90	30	TRUE	male	0	0	0
13307.s1	45		TRUE	male	1	0	0
13309.p1	89	79	TRUE	male	0	0	0
13309.s1	38		TRUE	male	0	0	0
13312.p1	75	119	FALSE	male	0	0	0
13312.s1	45		FALSE	female	1	0	0
13315.p1	90	118	TRUE	male	0	0	0
13315.s1	37		TRUE	male	0	0	0
13322.p1	77	59	TRUE	male	3	0	0
13322.s1	36		TRUE	female	1	0	0
13327.p1	90	103	TRUE	male	3	0	0
13327.s1	45		TRUE	female	2	0	0
13328.p1	61	100	FALSE	male	1	0	0
13328.s1	45		FALSE	male	0	0	0
13330.p1	60	101	FALSE	male	1	0	0
13330.s1	58		FALSE	female	0	0	0
13335.p1		18	FALSE	female	2	0	0
13335.s1	36		FALSE	male	5	0	0
13338.p1	90	105	TRUE	male	2	0	0
13338.s1	51		TRUE	female	2	0	0
13346.p1	84	59	FALSE	female	1	1	3
13346.s1	60		FALSE	female	2	0	0
13349.p1	76	87	TRUE	male	0	1	0
13349.s1	42		TRUE	female	0	0	0
13355.p1	90	30	TRUE	male	2	0	1
13355.s1	39		TRUE	male	1	0	0
13366.p1	74	124	FALSE	male	1	0	0
13366.s1	56		FALSE	female	2	0	0
13374.p1	90	19	TRUE	male	0	0	0
13374.s1	39		TRUE	female	0	0	0
13385.p1	90	16	TRUE	male	1	0	0
13385.s1	52		TRUE	female	0	0	0

13387.p1	90	95	TRUE	male		0	0	0	0
13387.s1	42		TRUE	male		2	0	0	0
13393.p1	59	77	FALSE	female		2	0	0	0
13393.s1	47		FALSE	female		2	0	0	0
13396.p1	83	103	TRUE	male		3	0	0	0
13396.s1	43		TRUE	female		2	0	0	0
13398.p1	90	75	TRUE	male		1	1	0	0
13398.s1	42		TRUE	female		1	0	0	0
13412.p1	90	33	TRUE	male		2	0	0	0
13412.s1	36		TRUE	female		0	0	0	0
13418.p1	61	135	FALSE	male		4	0	0	0
13418.s1	46		FALSE	female		1	0	0	0
13424.p1	74	114	FALSE	male		0	0	0	0
13424.s1	42		FALSE	female		0	0	0	0
13439.p1	87	82	TRUE	male		0	1	0	0
13439.s1	42		TRUE	female		0	0	0	0
13443.p1	76	102	TRUE	male		1	0	0	0
13443.s1	45		TRUE	female		0	0	0	0
13444.p1	85	76	TRUE	male		0	0	0	0
13444.s1	51		TRUE	female		0	0	0	0
13447.p1	90	44	FALSE	female		0	1	0	0
13447.s1			FALSE	female		1	0	1	1
13462.p1	87	70	TRUE	male		0	0	0	0
13462.s1	41		TRUE	male		0	0	0	0
13465.p1	90	121	TRUE	male		1	0	0	0
13465.s1	46		TRUE	female		1	0	0	0
13486.p1	83	93	TRUE	male		0	0	0	0
13486.s1	42		TRUE	male		0	0	0	0
13487.p1	56	96	FALSE	male		2	0	0	0
13487.s1	38		FALSE	female		2	0	0	0
13493.p1	77	104	TRUE	male		2	0	0	0
13493.s1	57		TRUE	male		1	0	0	0
13496.p1	77	107	TRUE	male		0	0	0	0
13496.s1	42		TRUE	male		0	0	1	1
13502.p1	90	55	TRUE	male		1	0	0	0
13502.s1	49		TRUE	female		1	0	0	0
13504.p1	77	64	TRUE	male		4	0	0	0
13504.s1	43		TRUE	female		3	0	0	0
13505.p1	60	112	FALSE	male		0	0	0	0
13505.s1	41		FALSE	female		0	0	0	0

13507.p1	46	101	FALSE	male		1	0	0
13507.s1	37		FALSE	male		2	1	0
13508.p1	63	96	FALSE	male		1	0	0
13508.s1	41		FALSE	female		2	0	0
13509.p1	89	70	TRUE	female		2	0	0
13509.s1	44		TRUE	female		1	0	0
13512.p1	83	110	TRUE	male		4	0	0
13512.s1	38		TRUE	female		1	0	0
13513.p1	67	88	FALSE	male		1	1	0
13513.s1	36		FALSE	female		1	0	0
13533.p1	71	47	FALSE	male		2	0	0
13533.s1	38		FALSE	male		2	0	0
13543.p1	82	42	TRUE	male		1	0	0
13543.s1	40		TRUE	female		0	0	0
13589.p1	78	56	TRUE	male		2	0	0
13589.s1	38		TRUE	male		3	0	0
13590.p1	90	86	TRUE	male		3	2	0
13590.s1	36		TRUE	male		2	0	0
13593.p1	84	39	FALSE	male		1	0	0
13593.s1			FALSE	male		2	0	0
13599.p1	79	65	TRUE	male		2	0	0
13599.s1	39		TRUE	male		2	0	0
13601.p1	85	78	TRUE	female		3	0	0
13601.s1	40		TRUE	male		4	0	1
13606.p1	90	54	TRUE	male		0	1	0
13606.s1	49		TRUE	female		1	0	0
13608.p1	90	42	FALSE	female		2	1	0
13608.s1			FALSE	male		1	0	0
13618.p1	90	44	FALSE	female		0	0	0
13618.s1			FALSE	female		0	0	0
13621.p1	90	56	TRUE	female		0	0	0
13621.s1	40		TRUE	male		1	0	0
13625.p1	85	94	TRUE	male		2	0	0
13625.s1	42		TRUE	male		1	0	0
13629.p1	82	52	TRUE	male		1	0	0
13629.s1	44		TRUE	female		2	0	1
13660.p1	80	53	TRUE	male		0	0	0
13660.s1	46		TRUE	male		1	0	0
13684.p1	76	106	TRUE	male		1	0	0
13684.s1	49		TRUE	female		0	0	0

13689.p1	79	97	TRUE	male		0	0	0
13689.s1	50		TRUE	female		1	1	0
13695.p1	88	54	TRUE	male		0	0	0
13695.s1	37		TRUE	female		0	0	0
13698.p1	89	98	FALSE	male		1	0	0
13698.s1	64		FALSE	male		1	0	0
13726.p1	80	61	TRUE	male		1	0	1
13726.s1	45		TRUE	male		1	0	0
13730.p1	68	94	FALSE	female		2	0	0
13730.s1	41		FALSE	male		0	0	0
13739.p1	90	25	TRUE	female		1	0	0
13739.s1	37		TRUE	male		2	0	0
13752.p1	76	98	TRUE	female		0	0	0
13752.s1	47		TRUE	female		0	0	0
13774.p1	90	33	TRUE	female		1	0	0
13774.s1	44		TRUE	female		0	0	0
13793.p1	87	52	TRUE	male		1	0	0
13793.s1	45		TRUE	female		2	0	0
13795.p1	90	34	TRUE	female		2	0	0
13795.s1	38		TRUE	female		1	0	0
13798.p1	68	103	FALSE	female		2	0	0
13798.s1	37		FALSE	female		0	0	0
13808.p1	79	40	TRUE	male		2	0	0
13808.s1	44		TRUE	male		2	0	0
13809.p1	77	61	TRUE	male		2	0	0
13809.s1	51		TRUE	female		0	0	0
13815.p1	82	51	TRUE	male		3	0	1
13815.s1	44		TRUE	female		2	0	0
13821.p1	67	92	FALSE	female		0	0	0
13821.s1	43		FALSE	female		0	0	0
13825.p1	90	58	FALSE	female		3	0	0
13825.s1	71		FALSE	female		4	0	0
13832.p1	90	25	TRUE	male		0	0	0
13832.s1	38		TRUE	female		1	0	0
13835.p1	90	119	TRUE	female		1	0	0
13835.s1	52		TRUE	female		1	0	0
13840.p1								0
13840.s1								0
13843.p1	90	66	TRUE	female		3	0	0
13843.s1	42		TRUE	female		1	0	0

13876.p1	82	40	FALSE	male		1	0	2
13876.s1			FALSE	female		1	0	0
13887.p1	67	81	FALSE	female		0	0	0
13887.s1	43		FALSE	male		0	0	0
13890.p1	82	37	TRUE	female		2	1	0
13890.s1	47		TRUE	female		2	0	0
13912.p1	83	95	TRUE	female		0	0	0
13912.s1	44		TRUE	female		1	0	0
13922.p1	90	100	TRUE	female		1	0	0
13922.s1	41		TRUE	male		2	0	0
13926.p1	79	62	TRUE	female		1	0	0
13926.s1	41		TRUE	female		3	0	0
13992.p1	80	106	TRUE	female		0	0	0
13992.s1	45		TRUE	female		1	0	0
14009.p1	85	98	TRUE	female		0	0	0
14009.s1	35		TRUE	female		0	0	0
14011.p1	90	53	TRUE	female		1	0	0
14011.s1	36		TRUE	female		1	0	0
14110.p1	69	77	FALSE	male		3	0	0
14110.s1	39		FALSE	female		3	0	0
14167.p1	76	132	TRUE	male		0	0	0
14167.s1	42		TRUE	male		0	0	0
14201.p1	90	43	TRUE	male		2	0	0
14201.s1	45		TRUE	female		2	0	0

Table S8

Comparison Name	Group	# of quads	Count	Enrichment	Binomial	Paired t-test	95% CI (bootstrap)
Overall rare CNVs	Probands	411	458	1.15	0.04	0.004	1.09 - 1.29
	Siblings		397				
Overall genes in rare CNVs	Probands	411	921	1.27	< 0.00001	0.029	1.10 - 1.52
	Siblings		726				
Category	# of Quads	Probands Count	Siblings Count	enrichment	Binomial	Paired t-test	
# of CNVs							
High FSIQ: Proband IQ score ≥ 70	267	273	236	1.16	0.11	0.018	
Low FSIQ: Proband IQ score < 70	141	157	126	1.25	0.074	0.015	
SRS discordant proband-sibling pairs	276	316	251	1.26	0.0071	0.00015	
SRS concordant proband-sibling pairs	115	117	113	1.04	0.84	0.7	
Discordant SRS, Low IQ	109	138	104	1.33	0.034	0.0038	
Concordant SRS, Low IQ	22	19	22	0.86	0.76	0.53	
Discordant SRS, High IQ	167	175	145	1.21	0.1	0.016	
Concordant SRS, High IQ	93	98	91	1.08	0.66	0.46	
Genes Affected							
High IQ	267	537	472	1.14	0.044	0.14	
Low IQ	141	384	252	1.52	1.90E-07	0.084	
Discordant SRS	276	707	510	1.39	1.80E-08	0.02	
Concordant SRS	115	222	219	1.01	0.92	0.9	
Discordant SRS, Low IQ	109	348	207	1.68	2.30E-09	0.061	
Concordant SRS, Low IQ	22	36	45	0.8	0.37	0.52	
Discordant SRS, High IQ	167	351	298	1.18	0.041	0.18	
Concordant SRS, High IQ	93	186	174	1.07	0.56	0.55	
By CNV frequency							
Private CNVs only	411	271	245	1.11	0.27	0.1	
All Rare CNVs	411	453	394	1.15	0.046	0.0044	
By expression profile							
Genes with brain expression (average)	411	19/317	6/224	2.24	NT	NT	
... and in discordant SRS quads only	276	15/256	2/170	#VALUE!	NT	NT	
Genes with brain expression (any region)	411	73	43	1.7	NT	NT	
By previous pathogenic gene association							
Genes with previous association	411	83	59	1.41	NT	0.049	
... and in discordant SRS quads only	276	66	35	1.89	NT	0.006	
... and in concordant SRS quads only	115	17	24	0.71	NT	0.069	
By sex, family size and birth order							

CNVs	# of Quads	Probands Count	Siblings Count	Comparison	enrichment	Binomial	Paired t-test
Male Pro	335	358	313	0.93	1.14	0.089	0.012
Female Pro	76	95	81		1.17	0.33	0.19
Male Sib	191	204	186	0.53	1.1	0.39	0.2
Female Sib	220	249	208		1.2	0.061	0.007
Sex Concord.	211	217	199	0.49	1.09	0.4	0.22
Sex Discord.	200	236	195		1.21	0.054	0.0049
Both Male	163	163	152	0.82	1.07	0.57	0.37
Both Female	48	54	47		1.15	0.55	0.39
Female Pro, Male Sib	28	41	34	1	1.21	0.49	0.32
Male Pro, Female Sib	172	195	161		1.21	0.08	0.0083
Proband Older	224	232	209	0.52	1.11	0.29	0.12
Sibling Older	171	203	166		1.22	0.061	0.0068
1 sibling	247	264	242	0.53 (chi-sq)	1.09	0.35	0.15
2 siblings	117	131	110		1.19	0.2	0.092
3+ sibling	47	58	42		1.38	0.13	0.017

Table S9

Tissue Name	Ratio	Brain?	Proband Frac	Proband genes (top 5% in each tissue)	Sibling Fract	Sibling Count	Sibling genes (top 5% in each tissue)
temporal lobe	8.6328125	TRUE	0.05078125	DOCA2, PGBD5, MAPK3, XKR3, CRLF1, UCHL1, DNAJC6, GAN, NIP1A, C16orf45, NDRG2, SEZBL2, CNDP1	0.00588235	1	TMEFF2
pons	4.99046875	TRUE	0.05859375	FAM57B, KCNE2, PGBD5, KIF1A, GSTT1, CA2, DNAJC6, SLFN12, UCHL1, TRIM58, C16orf45, NDRG2, AGT, SEZBL2, CNDP1	0.01176471	2	CORIN, SLC22A1
cerebellum	4.6484375	TRUE	0.0546875	AQP4, PGBD5, MTSS1L, UCHL1, DNAJC6, OR2W3, NFAM1, CPLX1, IQSEC1, C16orf53, NDRG2, AGT, SEZBL2, CNDP1	0.01176471	2	MAPKBP1, PDZK1
parietal lobe	4.6484375	TRUE	0.0546875	RNASE1, AQP4, PGBD5, MTSS1L, CA2, DNAJC6, UCHL1, ZNF517, IQSEC1, C16orf45, NDRG2, AGT, SEZBL2, CNDP1	0.01176471	2	TMEFF2, PPM1J
amygdala	4.31640625	TRUE	0.05078125	DOCA2, KCTD13, AQP4, PGBD5, CA2, DNAJC6, UCHL1, SGIP1, CPLX1, C16orf45, AGT, SEZBL2, CNDP1	0.01176471	2	TMEFF2, MAPKBP1
hypothalamus	3.76302083	TRUE	0.06640625	KCTD13, AQP4, PGBD5, GSTT2, GSTT1, RNASE6, CHODL, DNAJC6, ORC3, UCHL1, C3, CA2, C16orf45, NDRG2, AGT, SEZBL2, CNDP1	0.01764706	3	TMEFF2, MAPKBP1, TANC1
subthalamic n	3.76302083	TRUE	0.06640625	AQP4, PGBD5, KIF1A, CRLF1, MTSS1L, UCHL1, DNAJC6, GAN, FGFR3, ATP8B3, C16orf45, CPLX1, AKR1C4, PDE4B, ZNF17, SEZBL2, CNDP1	0.01764706	3	OR4F15, TMEFF2, ZNF613
BM-CD34+	3.48632813	FALSE	0.08203125	XPO1, TACC3, RNASE3, RNASE2, SNRPD3, CHD1L, RPL8, GLIPR1, PPP4C, RNASE6, VPRES3, CLC, RAD51AP1, CORO1A, TRIM58, IGLL1, MYC, ZNF84, ATF1, PRC1, UCHL3	0.02352941	4	SLC20A1, KARS, MAVS, RUVBL2
medulla oblon	3.3203125	TRUE	0.078125	KCTD13, ADORA2A, AQP4, PGBD5, GSTT2, CRLF1, CA2, DNAJC6, SLFN12, HCRTR1, CPLX1, RNASE1, UCHL1, C22orf43, C16orf45, ANXA10, NDRG2, SEZBL2, CNDP1, KIF1A	0.02352941	4	TMEFF2, SLC22A1, KRT19, ZNF608
caudate nucle	3.15429688	TRUE	0.07421875	KCTD13, ADORA2A, AQP4, PGBD5, CA2, DNAJC6, DDHD2, HMX3, UCHL1, CPLX1, PTCO3, CEP72, RNASE1, ALDH5A1, NDRG2, AGT, SEZBL2, CNDP1, BCR	0.02352941	4	TMEFF2, ZNF608, C20orf27, STARD7
prefrontal cort	3.09958533	TRUE	0.0546875	KCTD13, AQP4, PGBD5, GSTT2, GSTT1, CRLF1, UCHL1, DNAJC6, IQSEC1, C16orf45, NDRG2, AGT, SEZBL2, CNDP1	0.01764706	3	MYOM2, TMEFF2, MAPKBP1
occipital lobe	2.87760417	TRUE	0.05078125	AQP4, PGBD5, CA2, DNAJC6, UCHL1, FGFR3, IQSEC1, SGIP1, C16orf45, NDRG2, AGT, SEZBL2, CNDP1	0.01764706	3	TMEFF2, MAPKBP1, ZNF608
PB-CD19+ Bc	2.87760417	FALSE	0.05078125	TACC3, SNRPD3, RPL8, GLIPR1, PPP4C, RNASE6, C22orf13, CORO1A, RSL24D1, C1orf131, MYC, ATF1, VPRES3	0.01764706	3	TXNIP, ANKRD39, ARID5A
adrenal gland	2.82226563	FALSE	0.06640625	RNASE1, METTL9, ZNF219, PGBD5, SIK1, GSTT2, GSTT1, QPRT, NQO2, MOCOS, ALDH2, TXN2, OR2W3, C3, ZNF70, VPRES3, ARHGEF40	0.02352941	4	RAB20, ERCC2, HGSNAT, STARD7
whole brain	2.5234375	TRUE	0.07421875	DOCA2, KCTD13, AQP4, ASPHD1, SEZBL2, GSTT2, CRLF1, CA2, DNAJC6, UCHL1, PGBD5, IQSEC1, CORO1A, SGIP1, C16orf45, NDRG2, AGT, COX6A1, CNDP1	0.02941176	5	MYOM2, TMEFF2, MAPKBP1, RUVBL2, DUSP2
adipocyte	2.49023438	FALSE	0.05859375	HSD17B12, FAM89A, GSTT1, CRLF1, THBS2, PPARG, GLIPR1, NQO2, ETFDH, DDT, C3, GGT5, ALDOA, MYC, UCHL1	0.02352941	4	TNFAIP8L3, CLIC3, PTGIS, LXL1
lymph node	2.21354167	FALSE	0.0390625	ZNF140, TACC3, GLIPR1, RNASE6, CORO1A, IL32, GGT5, C3, MEGF6, VPRES3	0.01764706	3	SLC20A1, ARID5A, DUSP2
fetal brain	2.21354167	TRUE	0.0390625	KCTD13, PGBD5, UCHL1, DNAJC6, ORC3, SGIP1, CORO1A, TP53BP1, C16orf45, SEZBL2	0.01764706	3	PNPLA3, TMEFF2, MAPKBP1
cerebellum pe	1.9921875	TRUE	0.046875	AQP4, PGBD5, KCNG2, UCHL1, DNAJC6, FGFR3, C16orf45, C16orf53, NDRG2, AGT, SEZBL2, CNDP1	0.02352941	4	MAPKBP1, TANC1, CNINM4, PDZK1
pancreatic isle	1.9921875	FALSE	0.05859375	RAB27A, GSTT2, GSTT1, UCHL1, KCNG2, RPL8, CHST9, HCRTR1, IL32, ACADSB, C3, ANXA10, MAPKAPK5, SHPK, SEZBL2	0.02941176	5	CBLC, KARS, KRT15, KRT19, CCNC
cingulate corti	1.9921875	TRUE	0.046875	KCTD13, PGBD5, KIF1A, CA2, DNAJC6, UCHL1, SGIP1, IQSEC1, C16orf45, AGT, SEZBL2, CNDP1	0.02352941	4	MYOM2, ANKRD36, TMEFF2, SLC22A1
leukemia lymph	1.89732143	FALSE	0.078125	QPRT, MYC, SNRPD3, CHD1L, BUB3, PPP4C, MAZ, CCT4, RAD51AP1, CORO1A, IL32, HNRNPAB21, CEP72, RSL24D1, IGLL1, TACC3, MAPKAPK5, METTL17, PRC1, HIRIP3	0.04117647	7	KARS, CDIE, RUVBL2, SMYD3, NCAPH, ADSL, NNL
testis seminife	1.859375	FALSE	0.0546875	KCTD13, RNASE1, XKR3, UPB1, CHODL, TACC3, ORC3, ARL6, SLC26A8, CAPS2, IGLL1, PRC1, C9orf93, TRIP12	0.02941176	5	TEX101, RUVBL2, PPM1J, GEMIN4, NCAPH
bone marrow	1.80245536	FALSE	0.07421875	ADORA2A, RNASE3, RNASE2, METTL9, CA2, PPP4C, MYH9, RNASE6, CLC, OR2W3, C22orf13, CORO1A, TACC3, TRIM58, RAB27A, IGLL1, MYC, PRC1, VPRES3	0.04117647	7	OR4F15, LAT2, HK3, FPR2, ANKRD35, NCAPH, TXNIP
lymphoma Btu	1.77083333	FALSE	0.0625	ADORA2A, XKR3, RPL8, QPRT, TACC3, DDT, TNIP2, CORO1A, IL32, RSL24D1, AKR1C4, ALDH5A1, MYC, UCHL1, PRC1, VPRES3	0.03529412	6	LAT2, RUVBL2, ADRA2B, CCNC, PLAG1, DUSP2
liver	1.77083333	FALSE	0.0625	QPRT, SHPK, UPB1, CA2, PPP4C, MOCOS, ALDH2, DDT, C3, IL32, PQLC1, AKR1C4, ANXA10, AGT, CNDP1, NQO2	0.03529412	6	RUVBL2, HFE2, PNPLA3, CES2, SLC22A1, ABCO6
spinal cord	1.74316406	TRUE	0.08203125	HSD17B12, RNASE1, AQP4, ASPHD1, RNASE6, MTSS1L, CA2, DNAJC6, DDHD2, ORC3, UCHL1, PGBD5, C3, C1orf198, AKR1C4, ANXA10, NDRG2, AGT, SEZBL2, CNDP1, PACS2	0.04705882	8	ARHGEF10, TNFAIP8L3, FNTA, GABARAPL2, TMEFF2, TANC1, SCCPDH, SEMA4C
lymphoma Btu	1.74316406	FALSE	0.08203125	XPO1, TACC3, PGBD5, SNRPD3, RPL8, BUB3, RAD51AP1, KONG2, RBFA, CCT4, DDT, IGLL1, CORO1A, HNRNPAB21, DGLOK, AKR1C4, ALDH5A1, MYC, RSL24D1, PRC1, VPRES3	0.04705882	8	KARS, SLC20A1, RUVBL2, CCNC, ANKRD36, NCAPH, ADSL, DUSP2
ciliary ganglior	1.7265625	FALSE	0.05078125	LCN6, RNASE8, PRPH, ZNF300, CHST9, OR4D10, PTCO3, CAPS2, ANXA10, C1orf131, ADAMTS18, ZNF70, UCHL1	0.02941176	5	ALG10, PLAX1, LUX1, USP50, OR10A6
globus pallidu	1.7265625	TRUE	0.05078125	AQP4, LCN6, MTSS1L, UCHL1, DNAJC6, LCN8, OR4E2, PGBD5, C16orf45, NDRG2, ZNF17, SEZBL2, CNDP1	0.02941176	5	C3orf53, PNPLA3, SLC22A1, ST6GALNAC1, USP50
leukemia prorr	1.66015625	FALSE	0.078125	RPL8, ADORA2A, TACC3, SNRPD3, EIF3M, BUB3, RBFA, ORC3, CCT4, RAD51AP1, CORO1A, SHPK, HNRNPAB21, RSL24D1, 20 ALDH5A1, MYC, UCHL1, PRC1, UCHL3, VPRES3	0.04705882	8	KARS, RUVBL2, TANC1, EIF2A, SAMM50, NCAPH, ADSL, DUSP2
testis germ ce	1.66015625	FALSE	0.05859375	KCTD13, RNASE1, LCN6, XKR3, ACTG2, CHODL, TACC3, ORC3, CHST9, UPB1, PGBD5, SLC26A8, PRC1, EDM8B, TRIP12	0.03529412	6	RUVBL2, MAPKBP1, GEMIN4, TEX101, NCAPH, ABCC1
721 B lympho	1.61272321	FALSE	0.06640625	QPRT, SNRPD3, BUB3, RNASE6, UCHL1, PPP4C, TACC3, DDT, RAD51AP1, CORO1A, IL32, RSL24D1, MYC, MXI, ATF1, PRC1, HIRIP3	0.04117647	7	KARS, ARID5A, RUVBL2, C15orf41, NCAPH, ANKRD38, DUSP2
BM-CD105+ e	1.59375	FALSE	0.09375	PPRAC, RAD51AP1, C1orf131, MYC, SHPK, ARV1, UCHL3, SNRPD3, GSTT1, CA2, DNAJC6, TACC3, ATP8B3, CEP72, IGLL1, 24 ATF1	0.05882353	10	KARS, SLC20A1, C20orf27, RUVBL2, SAMM50, NCAPH, TTC27, PLAG1, C15orf41, DUSP2
thalamus	1.4609375	TRUE	0.04296875	RNASE1, PGBD5, GSTT1, CA2, DNAJC6, UCHL1, C3, C16orf45, AGT, SEZBL2, CNDP1	0.02941176	5	TMEFF2, TANC1, PPM1J, ST6GALNAC1, ZNF608
dorsal root gae	1.43880208	FALSE	0.05078125	FAM57B, MTRN1A, OR2G3, DNAJC6, OR4E2, PRPH, CLC, ATP8B3, DGKI, TBX8, ARL6, UCHL1, NIP1A	0.03529412	6	ZNF613, OR4F6, TMEFF2, ZNF649, SPINK5, STARD7
kidney	1.328125	FALSE	0.0625	QPRT, GSTT1, RTDR1, CA2, SLC13A3, NQO2, ALDH2, DDT, NFAM1, GGNL1, CEP72, GGT5, UPB1, IL32, CDH3, AGT	0.04705882	8	PNPLA3, TEK, ATP9VA04, CES2, ST6GALNAC1, ABCC6, KRT19, PDZK1
cardiac myocy	1.328125	FALSE	0.0703125	TACC3, RNASE3, MYH9, GSTT2, CRLF1, PPP4C, GLIPR1, THBS2, VPRES3, TUBGCP5, BMPER, AKR1C4, ANXA10, MYC, F11, UCHL1, PRC1, GALNT2	0.05294118	9	TMEM127, SLC20A1, CLIC3, PTGIS, KRT15, TEK, NCAPH, RHOC, C15orf41
testis interstiti	1.328125	FALSE	0.0390625	KCTD13, XKR3, CHODL, TACC3, ORC3, SLC26A8, IGLL1, PRC1, C9orf93, TRIP12	0.02941176	5	TEX101, RUVBL2, PPM1J, GEMIN4, NCAPH
olfactory bulb	1.328125	FALSE	0.0546875	QPRT, SLC13A3, SIK1, RNASE6, CA2, EHPB1, UCHL1, OR2W3, C1orf198, KANK1, C3, AGT, ATF1, CNDP1	0.04117647	7	ARHGEF10, ARID5A, TANC1, SAG, PIAS3, SEMA4C, STARD7
whole blood	1.23325893	FALSE	0.05078125	RAB27A, RNASE2, YPEL3, GLIPR1, PPP4C, MYH9, RNASE6, CLC, TACC3, CORO1A, IL32, RAF1, ARHGEF40	0.04117647	7	TMEM127, ARID5A, LAT2, HK3, FPR2, PPP1R12A, TXNIP
BM-CD71+ ea	1.1953125	FALSE	0.0703125	TMNTL1, TACC3, MAZ, SNRPD3, CA3, CA2, PPP4C, KIF22, DDT, RAD51AP1, C22orf13, RFPL2, TRIM58, OR2W3, PRC1, ADOCK1, 18 ATF1, PQLC1	0.05882353	10	RIOK3, SLC20A1, C20orf27, GABARAPL2, RUVBL2, PRR5, ANKRD35, TERF2IP, NCAPH, ANKRD39
adrenal cortex	1.1953125	FALSE	0.03515625	QPRT, SIK1, RNASE6, NQO2, HCRTR1, ACADSB, GGT5, C3, ARHGEF40	0.02941176	5	PNPLA3, RAB20, KRT15, SLC22A1, USP50
BM-CD33+ my	1.16210938	FALSE	0.0546875	METTL9, RNASE3, RNASE2, SIK1, NLRP3, GLIPR1, PPP4C, RNASE6, ATP8B3, CORO1A, TACC3, RAB27A, ATF1, MON1B	0.04705882	8	TMEM127, C20orf47, LAT2, HK3, FPR2, CAR52, HGSNAT, DUSP2
trigeminal gan	1.16210938	FALSE	0.0546875	FAM57B, RNASE3, OR2G3, NIP1A, PRPH, ARL6, ATP8B3, RAD51AP1, ZFP37, ACADSB, ANXA10, EDM8B, UCHL1, VPRES3	0.04705882	8	OR4F15, CORIN, OR10A6, USP50, C3orf33, SLC22A1, NCAPH, ALG10
appendix	1.13839286	FALSE	0.046875	TACC3, ACTG2, CRLF1, CA3, RNASE6, SLFN12, CHST9, SLC39A9, RFPL2, C3, ANXA10, SLC39A2	0.04117647	7	CORIN, SDPR, SNRNP200, FPR2, ATP9VA04, CENPF, MCTP2
testis Leydig c	1.10677083	FALSE	0.05859375	KCTD13, LCN6, ACTG2, CHODL, TACC3, LCN8, ORC3, SLC26A8, UPB1, HCRTR1, PRC1, EDM8B, C9orf93, TRIP12, HIRIP3	0.05294118	9	OR4F15, RUVBL2, PPM1J, GEMIN4, SCCPDH, CORIN, NCAPH, TEX101, KRT19
leukemia chro	1.10677083	FALSE	0.05859375	QPRT, EIF3M, MAPKAPK5, RAB27A, TACC3, NQO2, ORC3, CCT4, RAD51AP1, XKR3, CEP72, RSL24D1, MYC, ARV1, PRC1	0.05294118	9	KARS, SLC20A1, RUVBL2, TFB2M, SMYD3, ST6GALNAC1, NCAPH, ADSL, KRT19
smooth muscl	1.10677083	FALSE	0.05859375	SNRPD3, RPL8, GSTT1, GLIPR1, MYH9, THBS2, C3, IL32, DGUOK, 15 ALDOA, ANXA10, MYC, UCHL1, PRC1, GALNT2	0.05294118	9	SLC20A1, LTBP1, RUVBL2, ERCC2, PTGIS, GYS1, ANKRD23, RHOC, IL1A
PB-CD4+ Tcel	1.04352679	FALSE	0.04296875	RAB27A, TACC3, RPL8, BUB3, USP34, PPP4C, MYH9, CORO1A, IL32, MYC, ATF1	0.04117647	7	SLC20A1, CLIC3, ARID5A, C9orf142, TXNIP, PLAG1, DUSP2
pituitary gland	0.99609375	FALSE	0.03515625	RAB27A, PGBD5, UCHL1, DNAJC6, ORC3, TP53BP1, C16orf53, ZNF10, SEZBL2	0.03529412	6	CBLC, USP8, CLIC3, LMAN2L, ERCC2, PLAG1
trachea	0.99609375	FALSE	0.046875	RNASE1, HMX3, SIK1, ACTG2, CA2, ZNF70, CHST9, C3, NFAM1, ATP2C2, CDH3, MYC	0.04705882	8	TANC1, TXNIP, ATP12A, KRT19, KRT15, HGSNAT, ALG10, ST6GALNAC1
tonsil	0.99609375	FALSE	0.046875	LCN6, RNASE6, TACC3, MX1, ZNF517, CORO1A, IL32, VPRES3, C3, CDH3, MYC, PRC1	0.04705882	8	CBLC, CLIC3, LAT2, MAPKBP1, KRT15, NCAPH, SPINK5, KRT19
placenta	0.99609375	FALSE	0.046875	RNASE1, QPRT, MYH9, MMP11, ACTG2, LGALS14, PPP4C, CRLF1, PPARG, TRIM58, FAM89A, BCR	0.04705882	8	SLC20A1, CLIC3, LTBP1, CYP19A1, TEK, SEMA4C, BCAR1, KRT19
lung	0.94075521	FALSE	0.06640625	RNASE1, QPRT, CYP1P1, GSTT1, ACTG2, CRLF1, RPL8, PPP4C, MYH9, ALDH2, DDT, PPARG, CORO1A, IL32, RNASE6, C3, SIK1	0.07058824	12	TMEM127, CLIC3, ARID5A, RUVBL2, HK3, TXNIP, ST6GALNAC1, TRPM4, NDN, BCAR1, KRT19, DUSP2
PB-BDCA4+ c	0.91308594	FALSE	0.04296875	TACC3, SIK1, CHD1L, NLRP3, GLIPR1, PPP4C, RNASE6, MX1, CORO1A, ATF1, UCHL3	0.04705882	8	TMEM127, SLC20A1, CLIC3, ARID5A, HK3, PPM1J, C9orf142, DUSP2
ovary	0.91308594	FALSE	0.04296875	XKR3, ACTG2, PPARG, OR2G3, OR2G2, KIAA1958, C3, OR4D10, CPLX1, AKR1C4, ZNF70	0.04705882	8	OR10A3, OR4F6, ZNF615, PTGIS, SYT10, XRN1, ALG10, ASTL

fetal liver	0.91308594	FALSE	0.04296875	QPR1, TACC3, CLC, RAD51AP1, TRIM58, ACAD5B, C3, ANXA10, 11 PRC1, AGT, VPREB3	0.04705882
atrioventricula	0.88541667	FALSE	0.03125	8 SPN, XKR3, OR4E2, SLFN12, OR4D10, DGK, PTCHD3, CAPS2	0.03529412
heart	0.88541667	FALSE	0.046875	RNASE1, QPR1, GSTT1, ACTG2, RPL8, DDT, TXN2, CORO1A, IL32, 12 CEP72, ALDOA, AGT	0.05294118
salivary gland	0.88541667	FALSE	0.046875	SK1, RNASE6, CA2, ZNF84, ATP8B3, HCRTR1, OR4D10, GGT5, 12 C3, CDH3, EDDM3B, VPREB3	0.05294118
testis	0.81163194	FALSE	0.04296875	KCTD13, RNASE1, CAPS2, RTDR1, CHODL, PPP4C, TACC3, 11 CEP72, SLC26A8, PRC1, HIRIP3	0.05294118
PB-CD8+ Tcel	0.81163194	FALSE	0.04296875	RAB27A, BUB3, USP34, PPP4C, TACC3, ZNF84, CORO1A, IL32, 11 MAPKAPK5, ATF1, MYC	0.05294118
bronchial epit	0.77473958	FALSE	0.0546875	SNRPD3, CYFIP1, RPL8, PPP4C, MYH9, CCT4, TMEM40, PPARG, 14 UCHL3, RSL24D1, SK1, CDH3, MYC, PRC1	0.07058824
skin	0.77473958	FALSE	0.0546875	SK1, SPN, XKR3, OR2G3, THBS2, ARL8, GAN, OR4D10, SLC5A4, 14 CAPS2, ADAMTS18, ZNF70, FGFR3, ZNF396	0.07058824
superior cervi	0.77473958	FALSE	0.02734375	7 SLFN12, RAD51AP1, PTCHD3, TRIM58, ANXA10, EDDM3B, UCHL1	0.03529412
thymus	0.76622596	FALSE	0.05859375	ADORA2A, TACC3, SNRPD3, RNASE6, PPP4C, MYH9, MX1, IGLL1, 15 CORO1A, IL32, C3, CDH3, MYC, PRC1, HIRIP3	0.07647059
prostate	0.76622596	FALSE	0.05859375	RAB27A, CYFIP1, GSTT1, ACTG2, CA3, SEZ6L2, PPP4C, RPL8, 15 CLC, ATP2C2, SK1, CDH3, DOT, MEGF6, GSTT2	0.07647059
pancreas	0.74707031	FALSE	0.03515625	9 RNASE1, HMX3, RPL8, PPARG, GAN, OR4D10, IL32, C3, MYC	0.04705882
fetal lung	0.71940104	FALSE	0.05078125	RNASE1, SK1, ACTG2, CRLF1, RNASE6, MYH9, IL32, C3, 13 SNRPD3, CDH3, AGT, MEGF6, PRC1	0.07058824
uterus corpus	0.6640625	FALSE	0.0390625	CHST9, KCNE1, ACTG2, CA3, TEC, THBS2, ATP8B3, GGT5, 10 C16orf45, ZNF268	0.05882353
PB-CD14+ mc	0.6640625	FALSE	0.03125	RAB27A, RNASE2, NLRP3, GLIPR1, PPP4C, RNASE6, CORO1A, 8 TACC3	0.04705882
fetal thyroid	0.6640625	FALSE	0.0390625	GSTT1, CRLF1, CA3, PPARG, ATP8B3, C3, ZNF34, GGT5, AKR1C4, 10 ARHGAP40	0.05882353
colorectal ade	0.60369318	FALSE	0.0390625	METTL8, BUB3, GSTT1, TACC3, RAD51AP1, CEP72, CDH3, ATF1, 10 MYC, PRC1	0.06470588
tongue	0.59765625	FALSE	0.03515625	CRLF1, CA3, ZNF10, GAN, TRIM58, CAPS2, SLC39A2, MEGF6, 9 FGFR3	0.05882353
skeletal muscul	0.59027778	FALSE	0.03125	8 IGLL1, XKR3, CA3, SLC2A11, HCRTR1, IL32, ALDOA, ZNF268	0.05294118
thyroid	0.58105469	FALSE	0.02734375	7 GSTT1, CA3, PPARG, TRIM58, C16orf53, MTNR1A, ATF1	0.04705882
uterus	0.48697917	FALSE	0.04296875	RNASE3, CYFIP1, GSTT2, ACTG2, MEGF6, THBS2, OR2G2, 11 KANK1, AKR1C4, MTNR1A, ATF1	0.08823529
PB-CD56+ NK	0.44270833	FALSE	0.03125	8 RAB27A, GLIPR1, PPP4C, TACC3, CORO1A, IL32, PPP2R5E, ATF1	0.07058824

8 MUT, CYP19A1, HFE2, PNPLA3, NCAPH, ABCG6, C15orf41, PDZK1
6 CLIC3, BAIAP2L2, STARD7, PNPLA3, TRPM7, C15orf41
9 MYOM2, TMEM127, CNNM4, OR10A6, RUVBL2, FAM96B, GYS1, RHOC, PPP1R13L
9 CORIN, BAIAP2L2, PPM1J, ATP6V0A4, TRPM4, KRT15, XRN1, ALG10, KRT19
9 CNNM4, RUVBL2, ERCC2, MAPKBP1, PPM1J, GEMIN4, TEX101, KRT15, NCAPH
9 SLC20A1, CLIC3, ARID5A, CD160, ANKRD36, C3orf142, TXNIP, PLAG1, DUSP2
12 CBLC, KARS, SLC20A1, CLIC3, RUVBL2, TANC1, KRT15, RHOC, IL1A, NXN, PLAG1, KRT19
12 C3orf53, FAM57A, PNPLA3, USP50, ATP6V0A4, VP553, KRT15, ZNF649, CORIN, SPINK5, PPP1R13L, KRT19
6 ZNF613, OR4F6, BAIAP2L2, FPR2, PTGIS, SLC22A1
13 CD1E, CLIC3, ARID5A, RUVBL2, KRT15, ANKRD36, ANKRD23, NCAPH, TXNIP, LMAN2L, NINL, KRT19, DUSP2
13 CBLC, CD1E, CLIC3, RUVBL2, TMEFF2, TANC1, CARS2, KRT15, ZNF432, TRPM4, NXN, KRT19, DUSP2
8 CBLC, OR10A3, OR10A6, DZIP1L, ZNF613, KRT15, STARD7, KRT19
12 OR4F15, SDRP, CLIC3, RUVBL2, PLAG1, HFE2, TANC1, KRT15, TXNIP, LMAN2L, KRT19, PDZK1
10 SNRNP200, TNFAIP8L3, SAG, PTGIS, ATP12A, ZNF557, SLC22A1, VP553, PLAG1, KRT19
8 TMEM127, LAT2, EMR1, FPR2, CARS2, HK3, TXNIP, DUSP2
10 RAB20, CLIC3, LMAN2L, TANC1, PNPLA3, HGSNAT, TXNIP, IL1A, PLAG1, KRT19
11 RAB20, SLC20A1, RUVBL2, ARHGDI1, TANC1, GEMIN4, KRT15, NCAPH, NXN, KRT19, DUSP2
10 MYOM2, CLIC3, HFE2, GYS1, KRT15, CES2, IL1A, SPINK5, PPP1R13L, KRT19
9 MYOM2, HFE2, CD160, CYP19A1, GYS1, KRT15, ANKRD23, SLC22A1, NXN
8 MYOM2, CNNM4, RUVBL2, MAPKBP1, TANC1, CCNC, KRT19, CLIC3
15 SLC37A3, SEMA4C, TNFAIP8L3, FAM18A3, ARID5A, RUVBL2, TANC1, LIX1L, PTGIS, CORIN, PIAS3, PPP1R12A, ZNF649, NXN, KRT19
12 MYOM2, TMEM127, SLC20A1, CLIC3, ARID5A, CD160, POLR3GL, PPP1R12A, HK3, TXNIP, ANKRD39, DUSP2