

## Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Weiss LA, Shen Y, Korn JM, et al. Association between microdeletion and microduplication at 16p11.2 and autism. *N Engl J Med* 2008;358:667-75. DOI: [10.1056/NEJMoa075974](https://doi.org/10.1056/NEJMoa075974).

## SUPPLEMENTARY MATERIALS

### Supplementary Materials and Methods

#### AGRE

##### Samples

The Autism Genetic Resource Exchange (AGRE) has a collection of DNA from multiplex families with autism spectrum disorder (ASD) available for genetic research <sup>1</sup>. We genotyped 751 families, selecting for those with at least one child diagnosed with autism by the Autism Diagnostic Interview-Revised (ADI-R) <sup>2</sup>, while the second affected child had an AGRE classification of autism, broad spectrum (patterns of impairment along the spectrum of pervasive developmental disorders, including PDD-NOS and Asperger's syndrome) or Not Quite Autism (NQA, individuals who are no more than one point away from meeting autism criteria on any or all of the social, communication, and/or behavior domains and meet criteria for "age of onset"; or, individuals who meet criteria on all domains, but do not meet criteria for the "age of onset"). The self-reported race/ethnicity of these samples is 69% white, 12% Hispanic/Latino, 10% unknown, 5% mixed, 2.5% each Asian and African American, less than 1% Native Hawaiian/Pacific Islander and American Indian/Native Alaskan.

We excluded families with known chromosomal abnormalities (where karyotyping was available), and those with inconsistencies in genetic data (generating excess Mendelian segregation errors or showing genotyping failure on a test panel of 24 SNPs used to check gender and sample identity with the full array data). Our final dataset included 1,441 affected individuals (1,252 autism, 123 broad spectrum, 66 NQA), 1,420 parents, and 132 unaffected/unknown siblings (Table 1). Some of these same samples were analyzed on Affymetrix 500K and 5.0 arrays at Johns Hopkins, and analyzed for copy number using Affymetrix CNAT 4.0 software.

##### Controls

Additional samples from other ongoing studies used as controls for this study were 1087 cases with bipolar disorder from STEP-BD and 1727 controls obtained from the NIMH Genetics Repository genotyped on the Affymetrix 500K platform <sup>3</sup>, and analyzed for CNVs with COPPER (see below).

##### Genotyping

The samples were genotyped on Affymetrix 5.0 chips at the Genetic Analysis Platform of the Broad Institute. The 5.0 chip was designed to genotype nearly 500,000 SNPs across the genome in order to enable genome-wide association studies, as well as to improve the assessment of dosage changes by adding 500,000 sequence invariant probes (100,000 of which were targeted towards 1,900 regions of known common copy number polymorphism). In this study, the raw intensity data from these chips is

analyzed with two novel algorithms, COPPER and Birdseye, described below. In addition to standard quality control of scans, intensity data was examined for excess variance after normalization. Genotype calling (used here for allele normalization) and quantile normalization of arrays was performed within Birdseed or BRLMM. Both the genotype data and raw intensity files have been released to AGRE, and are available to the research community under AGRE guidelines.

## Analysis

### COPPER

COPY number Polymorphism Evaluation Routine (COPPER, developed by LAW, MARF with technical assistance from DMR) is a method that utilizes allele-specific intensity data from SNP probes to estimate copy number across the genome. It leverages large sample sizes in normalization and can detect small recurrent events in the population (including those that encompass only two local SNP probes). Additionally, it summarizes information from neighboring SNPs per individual to detect larger events (at least six SNP probes) present in single or few samples. Furthermore, COPPER can output normalized intensity data for verification of events by visual examination.

Briefly, it involves 4 steps: 1) normalization of intensity data, 2) copy number calling for each SNP probe, 3) consensus calling for pairs of nearby SNPs, and 4) summarizing data into copy number events. 1) Allele-specific intensity data is normalized and scaled to reflect copy number by assuming that the median intensity for allele A in individuals genotyped as AA corresponds to two copies, the median intensity for allele A in individuals genotyped as AB corresponds to one copy, and the median intensity for allele A in individuals genotyped as BB corresponds to zero copies; likewise for allele B. 2) After adjusting each individual to a genomewide average of copy number two, the allele-specific quantitative estimates are converted into integer calls of zero to four for each allele and summed across alleles for a copy number call at each locus. 3) Then 'consensus markers' are created by comparing copy number calls for each SNP with the most correlated flanking SNP (considering three on each side), and creating one marker from that pair by setting conflicting calls to missing. Each SNP cannot be used more than once to create a consensus marker. 4) Next, each individual is scanned in windows of five consensus SNPs (stepping by two), and a copy number variant is called if three of the markers in the window are consistent with deletion (copy number  $<2$ ) or duplication (copy number  $>2$ ). Therefore, a minimum of six SNPs (3 consensus markers) must reflect copy number less than or greater than two to make a CNV call, reducing false positive calls. Finally, these window calls are summarized into events in each individual with estimated boundaries, and overlapping events are summarized into regions of copy number variation across the sample, with output indicating the number of cases and controls with CNVs in each region, odds ratio, number of positive markers, etc. This method will be described in more detail in a subsequent publication.

### Birdseye

Birdseye (developed by JMK) uses both the SNP and copy number probes available on the Affymetrix 5.0 array to detect copy-number variants (CNVs). On an individual by individual basis, it considers the intensity deviation of each probe using models built from copy number differences on the X chromosome. Birdseye then identifies regions of dosage abnormality in individual samples by combining probe information using a Hidden Markov Model with five states (dosage=0,1,2,3,4) and employing the standard Viterbi algorithm<sup>4</sup>, and generates LOD scores expressing the likelihood of deviation from copy number two in any region. Scans for *de novo* deletion and duplication events were performed by requiring LOD>6 (i.e., million to one odds) in favor of deletion/duplication in the child but no deletion/duplication in either parent (also with LOD>6).

Briefly, each locus on each plate is modelled independently using empirical data. The response characteristics of each probe are maximum likelihood estimates of mixtures of Gaussians, assuming most samples have normal copy number at that locus. The response of copy number probes is a single Gaussian that best explains the distribution of sample intensities observed, while for SNPs the response is a mixture of 3 Gaussians which are learned using Birdseed (the SNP genotyping algorithm). These Gaussians define the copy-normal clusters. Copy-variable clusters (locus responses) are imputed from these, using a linear regression model built from chromosome X probes (which show differential response for haploid males versus diploid females). Each sample is then independently analyzed using a 5-state HMM—one state per copy number (where copy number at each locus is assumed to be 0, 1, 2, 3, or 4). The models for individual probes serve as point-estimates for the probability of being in each potential copy number state. The transitional probabilities are set low (0.0002 to transition from 2 to variable copy number, 0.05 to transition back to normal copy number, and 0.0001 to transition between two different copy variable states) to reflect the low expected number of CNVs in each genome. The Viterbi best path of copy number states along the genome is then computed for each sample, and for each discovered event a LOD score is reported that reflects the relative probability of the event being true versus false given the observed data.

To assess performance, we performed an in-silico gender mixing experiment. For each probe on the X chromosome, mixture models were built using only female data from a single plate. The order of probes was randomly permuted in order to remove any true CNVs present. We then simulated a deletion by taking the intensity of a female sample for 200 consecutive probes, followed by the intensity of a male sample for N probes, and then another sequence of 200 probes for which female intensity was used (where N is 0, 3, 5, 10, or 20). Each deletion size was simulated 1000 times.

3 probe deletions:

30% found; 28% if limit to correct breakpoints within 1 probe  
median LOD 1.13 (mean 1.32)  
55% of those found have LOD>1

5 probe deletions:

75% found; 71% if limit to correct breakpoints within 1 probe

median LOD 2.42 (mean 2.70)  
80% of those found have LOD>1

10 probe deletions:

98% found; 93% if limit to correct breakpoint within 2 probes  
median LOD 7.80 (mean 7.98)  
99% of those found have LOD>1

20 probe deletions:

99% found; 93% if limit to correct breakpoint within 2 probes  
median LOD 19.95 (mean 20.00)  
99% of those found have LOD>1

Two (false positive) duplications with LODs of 0.19 and 1.16 were found in the simulation, which covered 2,000,000 non-deleted probes (approximately 1 full genome's worth of data). No false deletions were found.

Children's Hospital Boston

## Samples

Array comparative genomic hybridization (CGH) was performed on 997 consecutively submitted clinical samples, after referral by specialists in Clinical Genetics, Neurology, and Developmental Medicine at Children's Hospital Boston. These fell into two groups: 512 had a primary diagnosis of developmental delay (DD), mental retardation (MR), and/or autism spectrum disorder (ASD) without noted dysmorphic features and 485 were referred for a primary diagnosis other than DD, MR or ASD. In the first group, 129 had ASD as primary indication by the referring clinician, 373 had DD as primary indication, and most have not been formally assessed for ASDs. Among the second group of 485, major referral categories included 141 with a primary diagnosis of multiple congenital anomalies, 106 with dysmorphic features, 43 with seizures, and 41 with congenital heart disease. Of these 485, 51 had a secondary diagnosis of DD or MR and were therefore excluded from analysis, leaving 434 non ASD/DD/MR controls (Table 1). More detailed clinical information about cases with aberrations at 16p11 was obtained by medical records review after approval by Children's Hospital Boston IRB.

## Array CGH and Analysis

Agilent 244K human genome oligonucleotide CGH microarrays (G4411B, Agilent Technologies, Palo Alto, CA) were used for array CGH analysis at Children's Hospital Boston<sup>5</sup>. Test samples labeled with Cy5 were compared to a reference sample labeled with Cy3 for standard two-color array CGH or vice-versa in reverse dye labeling setting. Images were captured by an Agilent scanner and quantified using Feature Extraction software v9.0. CGH analytic software v3.4 was subsequently used for data normalization, quality evaluation and data visualization. Copy number aberrations were indicated using the ADM-2 (Aberration Detection Method 2) algorithm.

## Multiplex Ligation-dependent Probe Amplification (MLPA) and Fluorescent *In Situ* Hybridization (FISH) Design and Analysis

To independently confirm the deletion/duplication of 16p11.2 region, four pairs of MLPA target-probes were designed based on the unique sequences of four genes within this interval: *SPN*, *MAZ*, *TAOK* and *TBX6*. Additionally, four pairs of MLPA control probes were included, two pairs corresponding to unique sequences of *SEPT1* and *LAT* genes located on chromosome 16 outside the deletion interval, and two pairs corresponding to unique sequences on other chromosomes. All probes were synthetic oligonucleotides. MLPA reagents were commercially available (MRC-Holland, Amsterdam, Netherlands), and reactions were performed according to the manufacturer's instructions. Final PCR products were analyzed on an ABI3730XL for peak identification and quantification. Copy number alterations were visually inspected by superimposing the peak profile of a test sample with the profile of a normal sample using SoftGenetics GeneMarker software (SoftGenetics, LLC. State College, PA) as shown in Figure 1C. For actual copy number quantification, the peak areas were exported to a Microsoft Excel worksheet. Peak area for each probe was normalized to the mean value for all control probes. The relative ratio of each peak was calculated by comparing between test sample and normal sample. Deletion was identified as relative ratio  $<0.75$  and duplication as relative ratio  $>1.25$ .

FISH confirmation was performed as described previously<sup>5</sup> using BAC clone RP11-504I2 as a probe specific to the 16p11.2 deletion interval.

### deCODE Iceland

#### Samples

Searching for sequence variants affecting the risk of autism was done through a population study of autistic individuals ascertained through the State Diagnostic Counseling Center and the Department of Child and Adolescent Psychiatry in Iceland. We have investigated 299 cases and for 90% of them the parents or caregivers of each affected individual were queried using the Autism Diagnostic Interview-Revised (ADI-R<sup>2</sup>) (Table 1). The rest of the group (10%) is comprised of older individuals diagnosed according to the ICD-9 (WHO 1978) criteria for autism. For this study, all ICD-9 diagnoses were converted by clinical specialists to ICD-10 diagnoses<sup>6</sup>. The diagnostic evaluation also included cognitive/developmental assessment, medical work-up and consultation with child and adolescent psychiatrist and/or developmental pediatrician.

All sample and data collection connected to the project, and relevant control samples, has been conducted by a licensed patient recruitment center (PRC). The PRC is a separate, not-for-profit trust company that serves as a link between study participants and researchers. This is where encryption of all research material takes place. The PRC is staffed with experienced nurses who work closely with the clinicians in the

recruitment process using the same approach. The DNA samples were isolated from whole blood by conventional methods.

All participants, cases and controls, returned signed informed consents prior to participation in the study. All personal identifiers associated with medical information, questionnaire results, and blood samples were encrypted according to the standards set by the Data Protection Committee of Iceland <sup>7</sup>. All procedures related to this study have been approved by the Data Protection Authority and National Bioethics Committee of Iceland.

### Genotyping and Analysis

A total of 24,258 Icelandic individuals were genotyped at deCODE Genetics for 317,503 SNPs using the HumanHap300 BeadChip (Illumina). Deletions and duplications at 16p11 were detected using a Hidden Markov Model applied to the SNPs normalized intensity data (similar to the approach described in <sup>8</sup>).

### Statistical Methods

The Fisher's Exact Test was used to compare carrier frequency between cases and controls.

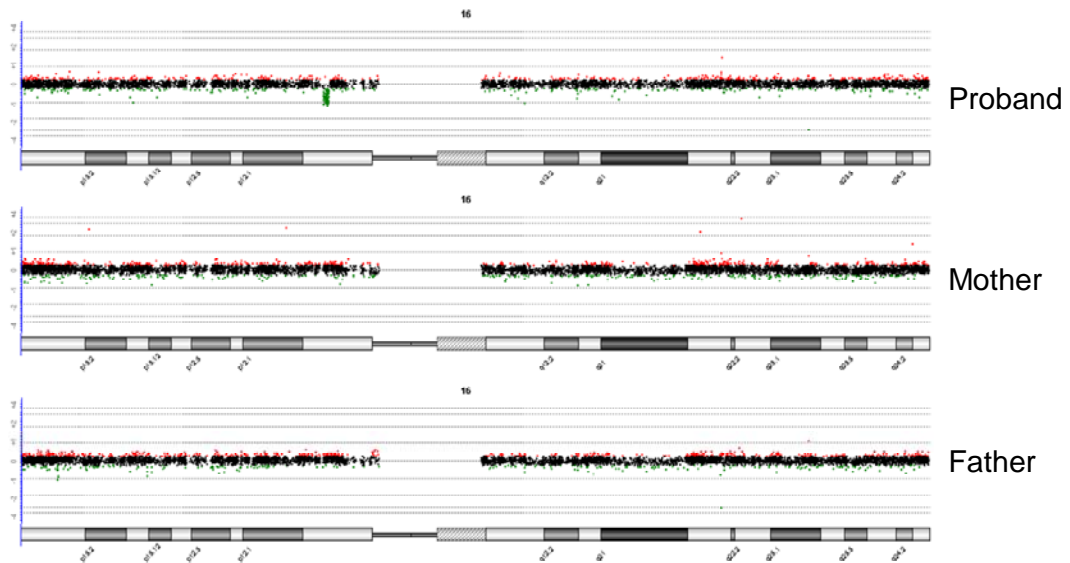
### Replication

Children's CGH sample was identified in advance as part of collaborative work supported by the Autism Consortium. deCODE Genetics researchers (RF, ES, HS, KS) contacted MJD after presentation of preliminary results at recent scientific meeting with offer to replicate this finding. We know of no additional replication attempts at this time.

Supplementary Figure 1. Oligo array CGH.

Data from one clinical sample with the same chromosome 16p11.2 deletion and parents screened by 244K whole genome array CGH platform at Children's Hospital Boston. On this scale, zero indicates that reference and test sample have equal copy number, below zero indicates that the Cy5 labeled sample has decreased copy number, and above zero indicates that the Cy5 labeled sample has increased copy number.

Supplementary Figure 1. Oligo Array CGH.



Supplementary Table 1. Phenotypic Data in AGRE families and Iceland probands with copy number variants at 16p11.2. Family structure and ADI-R and ADOS subscores, as well as performance measures are given for AGRE families and Iceland probands with copy number variants at 16p11.2

ID	Father ID	Mother ID	Gender	deletion/duplication	Scored Affected Status	ADOS_Diagnosis	Age at ADI-R exam	ADI-R social total	ADI-R communication - verbal total	ADI-R communication - non-verbal total	ADI-R behavior total	ADI-R development total	Age at ADOS exam	ADOS module	ADOS communication total	ADOS social total	ADOS communication + social total	ADOS play total	ADOS behavior total	age at Peabody Picture Vocabulary test	Peabody Picture Vocabulary test estimated age	age at Raven	Raven nonverbal IQ	age at Vineland	Vineland score	
AU002901	0	0	Female	duplication																						
AU002902	0	0	Male																							
AU002903	2	1	Male	duplication	Autism	Spectrum	13.07	28	16		6	4	15.46	4	3	4		7	1	2	15	>22yrs	15	94	16.2	6yrs 8months
AU002904	2	1	Female	duplication																						
AU002905	2	1	Female	duplication	Autism	Autism	9.84	27	17		8	3	11.94	3	6	12		18	2	0	11	10yrs 9mos	11	90	12.68	9yrs 5months
AU011001	9	8	Female																							
AU011002	0	0	Male																							
AU011003	2	1	Male																							
AU011004	2	1	Male	duplication	Autism	Autism	7.95	26	15		7	5	11.09	3	3	8		11	0	3			11		110	
AU011005	2	1	Male		Autism	Autism	5.76	27	17		4	5	8.9	3	6	12		18	2	2			8		136	
AU0154201	0	0	Male																							
AU0154202	0	0	Female																							
AU0154301	201	202	Female																							
AU0154302	201	202	Male	deletion	Autism	Autism	14.73	28		12	4	5	14.73	1	4	9		13	2	4	14		14		70	
AU0154303	201	202	Female	deletion	Autism	Autism	12.85	30		14	4	5	12.92	1	5	12		17	4	0	12		12			
AU029801	0	0	Female																							
AU029802	0	0	Male																							
AU029803	2	1	Male	deletion	Autism		4.79	16	14		3	3														
AU029804	2	1	Male		Autism		2.95	25	18		5	3														
AU032701	0	0	Female																							
AU032702	0	0	Male	duplication																						
AU032703	2	1	Female																							
AU032704	2	1	Male	duplication	Autism	Autism	9.99	28	18		4	5	14.82	3	4	7		11	0	1	14	14yrs 9mos	14		75	
AU032705	2	1	Male	duplication	Autism	Autism	8.36	28	23		8	5	13.19	2	5	8		13	1	5	13	7yrs 2mos	13		50	
AU032706	2	1	Male	duplication	Autism	Autism	6.86	27	18		4	5	11.68	3	3	7		10	0	0	11	14yrs 9mos	11		107	
AU032707	2	1	Male	duplication	Autism	Autism	10.55	26	24		8	5	10.55	3	6	13		19	1	1	10	9yrs 3mos	10		107	
AU041901	0	0	Female																							
AU041902	0	0	Male																							
AU041903	2	1	Female																							
AU041904	2	1	Male		Autism	Autism	7.56	28		14	6	5	9.65	1	6	8		14	1	5	9	4yrs 1mos	9		78	
AU041905	2	1	Male	deletion	Autism	Autism	5.92	15	15		6	5	7.96	3	5	10		15	1	4	7	5yrs 6mos	7		108	
AU0938201	0	0	Male																							
AU0938202	0	0	Female																							
AU0938301	201	202	Male	deletion	Autism	Autism	9.32	18	12		6	5	9.32	3	3	7		10	1	3	9	6yrs 5mos	9		80	
AU0938302	201	202	Male		Autism	Autism	7.43	16	13		4	5	7.43	3	5	9		14	1	3	7	4yrs 10mos	7		70	
Iceland Samples																				WISC-III Age	WISC-III VIQ	WISC-III PIQ	WISC-III FS			
Aut1	Adoptee 1a	Adoptee 1b	Female	deletion	Autism	Autism	18.33	10	10		4	1	18.42	4	3	8		11	0	0	17.75	65	61	61		
Aut2	ADHD 2a	2b	Male	deletion	Autism	Autism	10.58	14	9		7	3	10.58	3	7	11		18	3	2	7.1	75	75	73		
Aut3	3a	3b	Female	deletion	Autism	na	na	na	na		na	na	na	na	na	na		na	na	na	na	na	na	na		

Supplementary Table 2. Previously reported CNVs in autism. Events featured in the main tables or text of recent studies evaluating copy number in autism. For each reported event, we include events observed in our data with substantial overlap, and for inherited events, the number transmitted: not transmitted for affected offspring with data. Our sample has some overlap with previously reported samples, \* indicates a reported event in an overlapping sample which we also detect. \*\*, event not featured in the main text or tables of the paper, but listed in the supplementary table in an overlapping sample.

Event reported			Events observed			Inheritance, transmission (T:U)	Reference
Chr	Mb	event	Chr	Mb	event		
1	143.8-144.9	dup	1	144.8-145.1 143.5-145.0 143.6-145 143.5-145	del del dup dup	inherited, 1:0 <i>de novo</i> ** unknown, not shared by affected sib inherited 1:2	9
2	50.4-50.8	del	2	~50.4 ~51 (varied)	del	inherited, 1:1 (one add'l possible) inherited, 11:9	9
2	162.2-162.3	del	none			NA	10
2	236.4-249.4 238.2-242.7	del del	none			NA	10
3	60.7-60.8 61.1-61.4	del dup	3	60.9-61.1 60.8-61.2 60.8-60.9	del dup dup	parent, 0:4 parent, 0:2 parent, 0:1	10
6	14.0-15.3	del	none			NA	10
7	15.2-15.3*	del	no additional events			NA	10
10	50.6-61.5	dup	none			NA	10
13	44.2-46.1	del	none			NA	10

15	18.5-30.7	dup	15	21.2-26.3 18.8-26.4 23.3-24.7	dup dup dup	2 <i>de novo</i> inherited, 2:0; 1 unknown <i>de novo</i>	10
16	5.99-6.20	del	16	6.03-6.06 6.06-6.08 6.06-6.19	del del dup	parent, 0:2 inherited, 2:0 inherited, 2:0	10
16	29.6-31.1	del	16	29.5-30.0 29.5-30.0 29.6-29.7	del dup dup	<i>de novo</i> (5 cases in 4 families) inherited, 7:0, 1 <i>de novo</i> inherited, 1:1	10
17	14.3-15.2 14.3-15.2	del dup*	17	14.0-15.4	del	inherited, 2:0	9
20	0.08-0.37	del	20	0.20-0.22	del	parent, 0:1	10
20	2.8-3.9	del	20	3.72-3.74 2.9-3.0	del dup	inherited, 1:1 inherited, 1:0	10
22	17.5-19.8 17.5-20.3	del dup*	22	19.2-19.8 17.2-19.8 16.1-16.3 17.3-18.7 19.0-19.8	del dup dup dup dup	Unknown parents, 0:3 inherited, 2:1 unknown inherited, 1:1	9
22	45.1-49.5	del	22	45.60-45.65 48.17-48.23 48.14-48.18 47.7-48.3	del del del dup	parent, 0:2 inherited, 4:4 inherited, 2:0 inherited 3:1	10

Supplementary Table 3. Clinical information for Children’s Hospital deletion cases (Pt#) and Iceland deletion cases (Aut#). Clinical descriptive information is given for the 5 subjects in the Children’s Hospital Boston sample with deletions at 16p11.2, and the two subjects in the Iceland sample with deletions at 16p11.2 and clinical data available. The third Iceland subject has a history of seizures, but no other information available.

Supplementary Table 3  
Physical characteristics of individuals with del(16)(p11.2)

	Pt 1	Pt 2	Pt 3	Pt 4	Pt 5	Aut2	Aut1
<b>Gender</b>	Male	Male	Male	Male	Male	Male	Female
<b>History</b>							
Speech delay	+	+	+	+	+	+	+
Motor delay	+	+	+	+	+	-	+
Poor eye contact	+	NA	NA	+	+	+	+
Tactile aversion	+	NA	NA			-	-
Self stim behavior	+	NA	NA			+	-
Age at walking	18m	16m	17m	16m	16m	13m	
Regression	-	-	-	-	-	-	-
Other						ADHD, seizure	ADHD
<b>Physical Exam</b>							
Age at examination	6y6m	2y9m	17m	9y2m	9y2m	10y6m	5y2m
Height (%ile)	108.1cm (<3 <sup>rd</sup> )	NA	74.2 cm (<3 <sup>rd</sup> )	143.3cm (90 <sup>th</sup> )	150cm (>97 <sup>th</sup> )	152cm (95 <sup>th</sup> )	106cm (25 <sup>th</sup> )
Weight (%ile)	19kg (10 <sup>th</sup> )	NA	9kg (<3 <sup>rd</sup> )	65.4kg (>97 <sup>th</sup> )	71.9kg (>97 <sup>th</sup> )	69kg (>97 <sup>th</sup> )	18kg (50 <sup>th</sup> )

OFC (%ile)	51cm (50 <sup>th</sup> )	52cm (90 <sup>th</sup> )	48.5cm (75 <sup>th</sup> )	54.5cm (90 <sup>th</sup> )	56cm (>97 <sup>th</sup> )	54cm (60 <sup>th</sup> )	52.5cm (90 <sup>th</sup> )
Facial dysmorphism	-	-	-	-	-		
2-3 toe syndactyly	+	-	-	-	-		

Supplementary Table 4. Clinical information for Children's Hospital duplication cases. Clinical descriptive information is given for the 4 subjects in the Children's Hospital Boston sample with duplication at 16p11.2.

Supplementary Table 4  
Physical characteristics of individuals with dup(16)(p11.2)

	Pt 1	Pt 2	Pt 3	Pt 4
<b>History</b>				
Gender	Male	Female	Male	Female
Speech delay	unknown	+ (mild)	+	+
Motor delay	+	+	-	
Poor eye contact	NA	-	-	
Tactile aversion	NA	-	-	
Self stim behaviors	NA	-	-	
Age at walking	Not crawling or walking yet	NA	16m	14m
Regression	-	-	-	-
Other	Agenesis of corpus callosum			Seizures beginning age 6m; MR
<b>Physical Exam</b>				
Age at examination	14m	3y3m	2y6m	9y9m
Height (%ile)	NA	97.2 cm (50-75 <sup>th</sup> )	92.4cm (50 <sup>th</sup> )	
Weight (%ile)	11kg (50 <sup>th</sup> )	14.7 kg (50-75 <sup>th</sup> )	15.2kg (90 <sup>th</sup> )	33.5kg (50 <sup>th</sup> )
OFC (%ile)	80th	46.5 cm (< 3 <sup>rd</sup> )	51.25cm (95 <sup>th</sup> )	51cm (25 <sup>th</sup> )
Facial dysmorphism	-	-	-	-
Hypotonia	+ (mild)	-	-	-
Other		Mild spasticity; Fifth finger clinodactyly		
<b>Laboratory testing</b>				
Karyotype	NA	NA	46,XY,dup(16)(q11.2q12.1)	NA
Fragile X	NA	NA	NA	NA

NA = not assessed

Supplementary Table 5. Iceland phenotype information. Clinical diagnosis by ICD-10 category and transmission information is listed for the Icelandic autism samples with deletions at 16p11.2.

Alias ICE	ICD-10	Sex	Age	Transmission
Aut1	autism, mild MR	F	22	no info on parents
Aut2	Atypical Autism	M	12	transmitted from father
Aut3	Asperger's, unspecified MR, epilepsy	F	37	<i>de novo</i>

## REFERENCES

1. Geschwind DH, Sowinski J, Lord C, et al. The Autism Genetic Resource Exchange: A Resource for the Study of Autism and Related Neuropsychiatric Conditions. *Am J Hum Genet* 2001; 69:463-6.
2. Lord C, Storoschuk S, Pickles A, Rutter M. Using the ADI-R to diagnose autism in preschool children. *Infant Mental Health* 1993;14:234-52.
3. Sklar P, Smoller JW, Fan J, et al. Whole-genome Association Study of Bipolar Disorder. *Mol Psych.* *In Press.*
4. Viterbi A. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Trans Inform Theory* 1967;13:260-9.
5. Shen Y, Miller DT, Cheung SW, et al. Oligonucleotide-Array Comparative Genomic Hybridization Chip for Clinical Diagnosis of Genomic Imbalance. *Clin Chem* 2007;53(12):[Epub].
6. World Health Organization. The ICD-10 Classification of Mental and Behavioural Disorders. Diagnostic criteria for research. Geneva 1993.

7. Gulcher JR, Kristjansson K, Gudbjartsson H, Stefansson K. Protection of privacy by third-party encryption in genetic research in Iceland. *Eur J Hum Genet* 2000;8(10):739-42.
8. Colella S, Yau C, Taylor JM, et al. QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. *Nucleic Acids Res* 2007;35(6):2013-25.
9. Szatmari P, Paterson AD, Zwaigenbaum L, et al. Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nat Genet* 2007;39(3):319-28.
10. Sebat J, Lakshmi B, Malhotra D, et al. Strong association of de novo copy number mutations with autism. *Science* 2007;316(5823):445-9.