Title
Mapping autism risk loci using genetic linkage and chromosomal rearrangements.

Journal
Nature genetics, 39(3)

ISSN
1061-4036

Authors
Autism Genome Project Consortium
Szatmari, Peter
Paterson, Andrew D
et al.

Publication Date
2007-03-01

DOI
10.1038/ng1985

nature genetics

Mapping autism risk loci using genetic linkage and chromosomal rearrangements

The Autism Genome Project Consortium[1]

Autism spectrum disorders (ASDs) are common, heritable neurodevelopmental conditions. The genetic architecture of ASDs is complex, requiring large samples to overcome heterogeneity. Here we broaden coverage and sample size relative to other studies of ASDs by using Affymetrix 10K SNP arrays and 1,181 families with at least two affected individuals, performing the largest linkage scan to date while also analyzing copy number variation in these families. Linkage and copy number variation analyses implicate chromosome 11p12–p13 and neurexins, respectively, among other candidate loci. Neurexins team with previously implicated neuroligins for glutamatergic synaptogenesis, highlighting glutamate-related genes as promising candidates for contributing to ASDs.

Autism is a neurodevelopmental disorder characterized by impairments in reciprocal social interaction, communication deficits and repetitive and restricted patterns of behavior and interests. Autistic disorder is the prototypical pervasive developmental disorder (PDD or, equivalently, ASD), which form a group that also includes Asperger disorder, PDD not otherwise specified and Rett disorder[1]. Population prevalence of autism is approximately 15–20 in 10,000, and all ASDs affect about 60 in 10,000 children. Worldwide, males are affected four times as often as females[2]. Autism is associated with a recognized cause in only about 10% of individuals, most commonly with fragile X syndrome, tuberous sclerosis and chromosomal abnormalities[3,4].

Twin studies show a concordance of 60%–92% for monozygotic twins and 0%–10% for dizygotic pairs, depending on phenotypic definitions[5]. Milder phenotypes are similarly elevated in relatives of singleton probands, consistent with a spectrum of severity[6]. The estimated prevalence of autism in siblings is 5%–10% (refs. 7,8). The ratio of sibling recurrence risk to population prevalence varies from 67 to 25, both larger than for most multifactorial diseases. Although familial clustering in autism could reflect shared environmental factors, twin studies[5,9] and the distribution of milder phenotypes in families favors a model involving multiple interacting loci[10,11]. We hypothesize that liability to autism is due, in large part, to oligogenic inheritance in which combinations of susceptibility alleles contribute. Variation in phenotypic severity of sibling pairs and family members ascertained through an autistic proband are both consistent with this hypothesis. Based on numerous observations of karyotypic abnormalities in autism, we also hypothesize that submicroscopic alterations are involved.

Genome-wide linkage scans[12] (also reviewed in ref. 13) for autism susceptibility loci have identified chromosomal regions 2q, 7q and 17q, with 7q yielding the most consistently positive results, including support from meta-analysis. Moreover, substantial evidence suggests that chromosomal abnormalities contribute to autism risk, but the exact prevalence is unclear because literature surveys span different diagnostic and cytogenetic approaches and sample sizes. Recent surveys[3,4] show a mean rate of gross mutations and chromosomal abnormalities between 4.3% (78/1,826) and 7.4% (129/1,749), but many studies find rates of detected abnormalities in 5%–10% of affected individuals[3]. Among the most frequent findings are fra(X)(q27) (3.1%; 28/899) and anomalies involving proximal 15q (0.97%; 17/1,749), specifically the Prader-Willi and Angelman region[3,4]. Duplications of 15q11–q13, typically of maternal origin, are observed in 1%–3% of cases, either as interstitial duplications or supernumerary isodicentric marker chromosomes containing one or two extra copies of this region[3]. Linkage, association, and/or chromosome rearrangement studies have identified several ASD candidates, including genes encoding neuroligins and their binding partners, as having disease-associated mutations[14–17].

In our model for autism, combinations of multiple loci that possibly interact and microscopic or submicroscopic chromosomal abnormalities contribute to risk, complicating the detection of individual loci. Increasing the likelihood of detecting loci requires analyzing a large sample of multiplex families (i.e., families with two or more affected individuals), thereby enhancing the power of linkage analysis and controlling sources of etiologic heterogeneity (herein, the term families implies multiplex families).

We have assembled a sample of over 1,400 ASD families, a resource sufficiently large to implement multiple strategies for localizing susceptibility loci (see 'Power' in **Supplementary Methods** online). Although some linkage studies have attempted to control for heterogeneity attributable to chromosomal abnormalities by excluding the small number of affected families, none has attempted to merge

[1]A full list of authors is provided at the end of the article. Correspondence should be addressed to B.D. (devlinbj@upmc.edu) or S.W.S. (swscherer@sickkids.ca).

Received 16 October 2006; accepted 19 January 2007; published online 18 February 2007; corrected after print 26 September 2007; doi:10.1038/ng1985

Table 1 Families at each stage of quality control

		Narrow	Broad	hASD	Total
Diagnostic data	≥ 2 affected	675	942	554	1,496
Genotypic data	≥ 2 genotyped	675	942	549	1,491
Diagnostic and genotypic data	≥ 2 affected and genotyped	597	829	488	1,317
Edit 1	≥ 2 affected and genotyped	564	787	461	1,248
Edit 2	≥ 2 affected and genotyped	554	772	450	1,222
Edit 3	≥ 2 affected and genotyped	528	739	442	1,181

Quality control of samples was based on genotypes determined by the Affymetrix 10K array, which contains 10,112 SNPs. See **Supplementary Methods** for origin of sample by research center. Edit 1: eliminated pedigrees and samples incompatible with linkage assumptions (monozygotic twins and inconsistent nominal or genetic relationships). Edit 2: dropped loci missing ≥10% of their genotype calls and then dropped individuals missing ≥20% of their genotype calls (rates determined empirically based on the frequency distribution of missing genotypes). Edit 3: removed duplicate individuals and families who contributed DNA to more than one center. Most families were nuclear; after Edit 3, the number of marriages per family was 1.26, 1.24 and 1.16 for the 'narrow', 'broad' and 'hASD' categories.

linkage analysis with studies of fine-level chromosomal variation. We have developed an approach, using comparative analysis of hybridization intensities, to identify submicroscopic copy number variations (CNVs) as putative risk loci and as a tool to stratify the samples to reduce genetic heterogeneity for linkage analyses.

RESULTS

Individuals with ASD

The Autism Genome Project (AGP) Consortium, comprising scientists from 50 centers in North America and Europe, collected 1,496 ASD families (7,917 family members) for this study. Diagnosis was based on the Autism Diagnostic Interview-Revised (ADI-R) and the Autism Diagnostic Observation Schedule (ADOS) or clinical evaluation (see ref. 18 and Methods). Origins of the individuals screened are shown in **Supplementary Methods**; most were karyotyped (~71%) and screened for fragile X mutations (~94%), and families were excluded if either was abnormal in at least one affected individual. Most cell lines or DNA samples arising from the project are available at the National Institute for Mental Health (NIMH) Center for Collaborative Genetic Studies, the European Collection of Cell Cultures and the Autism Genetics Research Exchange. We genotyped genomic or in some cases whole genome–amplified (WGA) DNA.

Linkage analysis by diagnostic group

We successfully generated genotypes from 1,491 of 1,496 nominal families (6,709 samples; **Table 1**) using the Affymetrix 10K v2 SNP array. WGA had no discernible impact on genotyping accuracy: for 12 duplicate samples assessed, concordance of genotypes for WGA versus blood DNA was >99.6% with no significant difference in completion rates (both ~94%). From the 10,112 SNPs initially genotyped, quality control procedures resulted in marker exclusion for the following reasons: minor allele frequency <0.05 (removed 749 SNPs), high rate of missing genotypes (removed 1,112 SNPs), selection of tag SNPs (removed 1,734 SNPs) and deviations from Hardy-Weinberg Equilibrium (removed 391 SNPs). Following quality control, the discordant call rate per locus, based on 261 duplicate samples, was roughly 5/10,000.

Quality control on family data had a similar impact on reducing numbers (**Table 1**), yielding 1,181 families for linkage analysis. Of these families, we estimate that 64% were included in smaller, published linkage studies. The families were distributed across three diagnostic categories (narrow, broad and heterogeneous ASD (hASD)), which were defined according to the distribution of diagnosis (see Methods). Linkage analyses were performed for three nested diagnostic groupings (narrow, broad and all families); for analyses focusing on one diagnostic group, 'broad' is a reasonable choice because a substantial number of families fall in the group, and the sensitivity and specificity of this diagnostic method are reliable[18].

Before linkage analyses, we rebuilt the Affymetrix genetic map by linear interpolation from National Center for Biotechnology Information Build35 and markers of known genetic positions[19] to infer genetic locations for all SNPs. We then validated the new genetic map using the linkage data.

Linkage information, as reported by MERLIN, averaged ~95% over the genome (minima at telomeres, ≥71%) because of high coverage of markers across the genome and availability of parental genotypes (3% of families have no parental genotypes, whereas 79% have genotypes for both parents). Thus, the results are also insensitive to SNPs in linkage disequilibrium (see Methods). Only for all families does statistical evidence exceed the threshold for suggestive linkage[20], at 11p12–p13 (**Fig. 1**).

Characterization of copy number variation

We assessed our samples for CNV content using signal intensities obtained from the SNP arrays. Because the distribution of intensity is continuous, whereas copy numbers are discrete, an algorithm is required to infer copy number from signal intensity of a SNP genotype relative to intensity from other samples. Initially, to capture as many CNVs as possible, we used two approaches (termed 'batch' and 'plate-by-plate') for intensity comparisons, which yielded a total of 2,788 putative CNVs from 1,109 samples from 715 families (**Table 2**). To

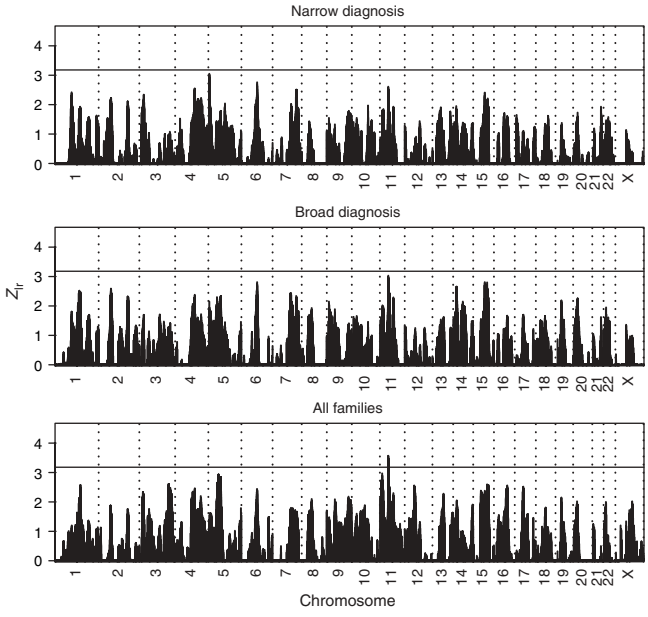

Figure 1 Linkage across the genome for all families and ancestries, based on levels of diagnostic certainty. Vertical reference lines separate chromosomes, which are ordered. The horizontal reference bar is given at a Z_{lr} of 3.18, the threshold for suggestive linkage according to the Lander/Kruglyak criterion[20], which is roughly accurate in this setting (4.1 is the threshold for significant linkage). The suggestive threshold would be expected to be crossed by chance once per genome scan. It is crossed once, and the peak falls within 11p12 (Z_{lr} = 3.57 at rs2421826).

Table 2 Characteristics of CNVs found in multiplex autism families using different stringencies of analysis

Features	Method[a]		
	Batch[b]	Plate-by-plate[c]	Filtered based on thresholds[d]
Total samples			
Number of CNVs	1,967	1,286	624
Number of samples	918	685	490
Number of CNVs/sample	2.14	1.88	1.28
Number of families	590	476	350
Mean/median size[e]	4.6/1.8 Mb	3.4/1.2 Mb	3.9/0.6 Mb
Gain/loss of CNV	1,749/228	1,064/232	402/222
Unaffected individuals			
Number of CNVs	1,186	802	370
Number of samples	538	425	292
Number of CNVs/sample	2.20	1.86	1.25
Number of families	419	329	235
Mean/median size[e]	4.7/1.9 Mb	3.6/1.3 Mb	4.3/0.69 Mb
Gain/loss of CNV	1,054/133	658/143	235/136
Affected individuals			
Number of CNVs	781	495	254
Number of samples	380	260	196
Number of CNVs/sample	2.06	1.90	1.29
Number of families	322	230	173
Mean/median size[e]	4.6/1.8 Mb	3.2/1.2 Mb	3.4/0.66 Mb
Gain/loss of CNVs	685/96	406/89	167/86
Inherited CNVs (number of regions)[f]	78 (62)	59 (46)	49 (39)
Sibling CNVs (number of regions)[g]	68 (34)	36 (18)	34 (17)
De novo CNVs (siblings)[h]	33 (26)	16 (10)	10 (6)
Familial recurrent CNVs (number of regions)[i]	28 (14)	16 (8)	14 (7)
Recurrent CNVs (number of regions)[j]	209 (66)	135 (43)	47 (18)
Overlapping CNVs[k]	422	251	79
Nonoverlapping CNVs	139	109	128
CNVs overlapping with ACRD[l]	68	27	18
CNVs overlapping with DGV[m]	20	10	9
CNVs validated (in affected)[n]	326 (162)	230 (106)	193 (95)
Number of families removed for linkage analysis modeling[o]	267	200	150

[a,b,c,d]CNV analysis for the arrays was performed using three approaches: after trimming the data set with the first-pass cutoffs, arrays were renormalized and analyzed using a 'batch comparison' (~1,000 scans; footnote b) or by a 'plate-by-plate' comparison (96 scans; footnote c) to avoid potential plate-specific batch effects. In the most stringent analysis (footnote d), thresholds were set based on validation data (see footnote n below) to minimize potential for false positives. We note that although column 4 contains data from the most stringent analysis, there are some CNVs found exclusively in one or both of columns 2 and 3 that are indeed real. All data can be viewed at the Autism Chromosome Rearrangement Database (see URL in Methods). [e]Range of CNV size: batch (100 bp to 240 Mb), plate-by-plate (100 bp to 134 Mb), filtered (100 bp to 134 Mb); distributions are shown in **Supplementary Figure 1**. [f]CNVs in affected individuals that were inherited from either parent. [g]CNVs present in two or more affected siblings that are either *de novo* or inherited. [h]CNVs in affected individuals that are *de novo* in origin with the total number of *de novo* events occurring in siblings in brackets. [i]CNVs that are familial and recurrent. [j]CNV gains or CNV losses with the same coordinates found in two or more unrelated families (a recurrent CNV gain and CNV loss at the same site is only counted once). [k]Two or more CNV gains or CNV losses with overlapping genomic coordinates (a CNV gain and CNV loss combination is not counted). [l]CNVs that overlap with mapped chromosome rearrangement breakpoints annotated in the Autism Chromosome Rearrangement Database. [m]CNVs that overlap with other CNVs not known to be associated with disease as catalogued in the Database of Genomic Variants (see URL in Methods). We note that some CNVs found in the DGV could be predisposing or disease related. [n]Total number of validated CNVs. A CNV was considered validated if it overlapped with mendelian genotype errors or if it was confirmed using an independent set of experiments (for example, karyotyping, quantitative PCR, array-CGH or Affymetrix 500K Mapping arrays) (**Supplementary Table 4**). [o]Number of families belonging to the group of 1,168 families passing data cleaning that were removed from linkage analysis. Families were only removed if at least one affected individual contained a CNV not found in the Database of Genomic Variants.

define a more stringent set of CNV calls, we examined the raw intensity data from 42 CNV calls that were presumed real based on overlap with non-mendelian genotype errors or by laboratory experimentation. We also assessed samples having identical CNVs within the same family, as these could also be considered to be validated calls. These analyses showed that our plate-by-plate signal intensity comparison had less background and was likely to contain fewer false-positive data compared with the batch approach (see Methods and ref. 21). Therefore, we then scrutinized these intensity files to guide threshold settings to define a highly stringent data set, called 'filtered,' containing 624 CNVs from 350 different families (**Supplementary Table 1** online and **Fig. 2**).

Caveats about these data are as follows: (i) there will be *bona fide* CNVs in the batch comparison data that fail to meet cutoffs in the filtered analysis; (ii) some CNVs could be somatic artifacts, such as cell culture–induced rearrangements and aneuploidy; (iii) the mapping resolution of CNV boundaries is dependent on local SNP density and is therefore nonuniform; (iv) smaller CNVs will be more likely to be missed (**Supplementary Fig. 1** online) and (v) balanced rearrangements will not be detected.

Considering solely our highest confidence data, we identified 254 CNVs in 196 ASD cases from 173 families (**Table 2**, **Fig. 2** and **Supplementary Tables 2** and **3** online). The average and median sizes were 3.4 Mb and 0.66 Mb, respectively, and the majority (66%) were CNV gains, probably owing to a greater tolerance in the genome for large gains versus deletions. The observations most relevant to ASD disease risk (**Supplementary Table 1**) included (i) the identification of ten families with apparent *de novo* CNVs (in three such families, the CNV was found in both ASD sibs); (ii) 18 CNVs in unrelated affected individuals having genomic locations coincident with published ASD chromosome rearrangements and (iii) 126 CNVs with recurrent (47) or overlapping (79) boundaries, suggesting they could be nonrandom events (**Supplementary Tables 3** and **4** online). We also detected seven samples from three families with ASD-associated chromosome 15q gains, all of which were maternally inherited, as would be expected (including at least two that escaped earlier karyotypic detection).

We highlight four CNV discoveries to demonstrate the utility and complexity of this data and also to serve as a prototype of how this new type of genetic information can be used in mapping studies. First, in family AS049, two female sibs with ASD had apparently identical 300-kb CNV losses of chromosome 2p16 not detected in either parent. Quantitative PCR analysis confirmed the microdeletion: microsatellite analysis showed the identical maternal chromosomal segment but no paternal DNA in the sibs, providing a likely explanation of paternal gonadal mosaicism. This hemizygous deletion eliminates coding exons from the neurexin 1 gene (*NRXN1*), which represents a functional

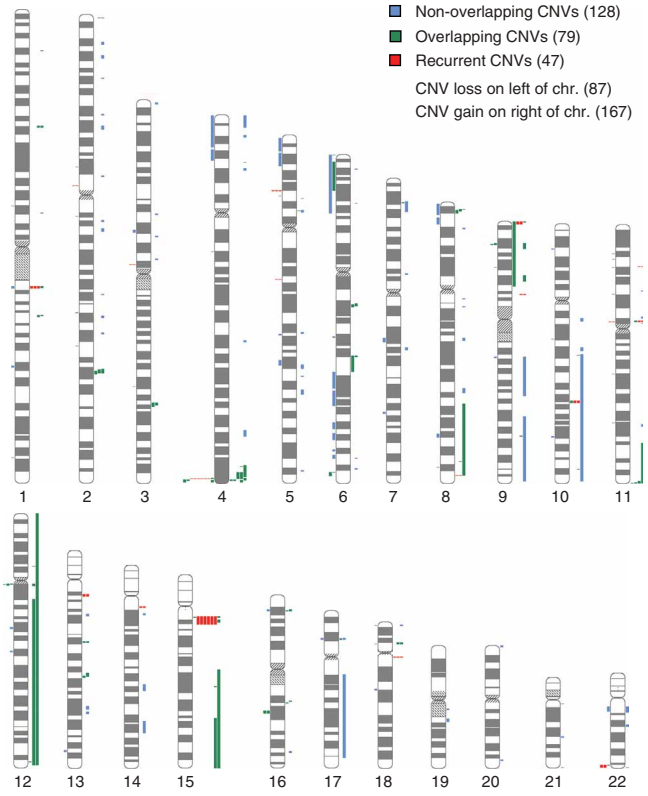

Figure 2 Chromosome ideogram depicting 253 inferred CNVs found in 196 individuals with ASDs. These CNVs are derived from the highest stringency 'filtered' data set, but many other true CNVs will also be found in the other analyses and should be examined further. Characteristics of the complete data set are described in **Table 2**. Some of the larger changes could represent somatic artifacts or missed karyotypic anomalies. All data are also downloadable or can be viewed in Genome Browser format at the Autism Chromosome Rearrangement Database (see URL in Methods). As additional analyses and validations are performed, the data will be posted at the same site.

candidate for ASD based on the role of *NRXN1* in synaptogenesis and its interaction with neuroligins. Rare *NRXN1* mutations apparently generate risk for ASD and mental retardation[16,22,23]. Both girls presented with typical autism, including characteristic developmental delays. Although we could not be certain of their verbal status because of their young age, one appeared nonverbal, whereas her sister had mild language regression. Neither parent had clinically important features.

Second, we found a recurring 1.1-Mb CNV gain at chromosome 1q21 in three families: AS048, with one affected male; AS039, one affected female; and AS007, two affected male sibs and their father of unknown affection status. It overlaps the same region implicated in mental retardation and other anomalies[21,24,25].

Third, we observed CNVs of ~933 kb at 17p12 as a *de novo* duplication in an affected male-female sib pair in one family (AS068), as a maternally inherited deletion in two affected male sibs (AS028) and as a paternally inherited deletion in an affected female (AS001). This interval, when duplicated, causes Charcot-Marie-Tooth 1A (CMT1A) and when deleted causes hereditary neuropathy with liability to pressure palsy[26]. This region also overlaps with microdeletions seen in some cases of Smith-Magenis syndrome, which itself has phenotypic overlap with ASDs[27]. Moreover, other microduplications of the same interval have been described in individuals with mental retardation, speech and language delay, autism and related phenotypes[28]. None of the implied disease-associated CNVs described above were observed in any known control sample database, which at the time of the study comprised ~500 samples from the general population[21,29].

Finally, we detected further complexity in two families with duplications of 22q11.2. In family AS063, the male proband diagnosed with autism inherited the duplication from his father, but a brother with PDD not otherwise specified does not have the duplication. In family AS019, the female sib diagnosed with autism carried the duplication, whereas an affected brother did not, and we did not observe the duplication in either parent. Genotyping confirmed the biological parents. FISH analysis confirmed duplications in both families and showed that it was *de novo* in the second family (data not shown).

Exploration of linkage by subsets of the data

Linkage analysis identifies regions harboring one or more genetic variants that account for a substantial portion of risk in families. Rare *de novo* or familial CNVs that confer risk for ASDs could be a source of noise or heterogeneity that decreases sensitivity in linkage analyses. Thus, in theory, linkage signals from major loci could be amplified if families with rare CNV risk alleles were removed. A sound strategy to evaluate subsets based on known and putative CNVs, however, is unclear. Consider our three levels of CNV discovery (namely, filtered, plate and batch), which are ordered by degree of stringency of evidence required to call a CNV. Moving from filtered to batch, it is reasonable to assume that the rate of false positives is increasing while the rate of false negatives is potentially decreasing. Selecting a single approach *a priori* favors a certain, unknown ratio of false positives to false negatives. This ratio might not be optimal, depending on how much of the attributable risk for ASDs accrues to CNVs: if CNVs account for a large fraction of the risk, the false negative rate is critical; conversely, if CNVs account for a small fraction of the risk, then the false positive rate is of greater importance. For this reason, we chose to explore the effects of all three levels of CNV discovery. Within each level, we removed families in which at least one individual was diagnosed with ASD and also carried at least one putative CNV (**Table 3**). Using the broad diagnostic group, we recomputed the linkage traces (**Supplementary Fig. 2** online).

Table 3 Number of families by diagnostic group, gender of ASD individuals and CNV-calling method

Diagnostic group[a]	Families[a]	Total	Filtered	Plate	Batch
Narrow	All	528	462	431	406
	MO	334	296	274	259
	FC	194	166	157	147
Broad	All	739	641	603	567
	MO	464	408	382	359
	FC	275	233	221	208
All	All	1,181	1,031	981	914
	MO	741	653	623	579
	FC	440	378	358	335

FC: families contained ASD-diagnosed females (female-containing); MO: families contained ASD-diagnosed males.
[a]A larger fraction of families tends to be removed from the more stringently diagnosed groups (narrow and broad) and from female-containing families, but neither trend approaches significance when the data are fitted to a log-linear model.

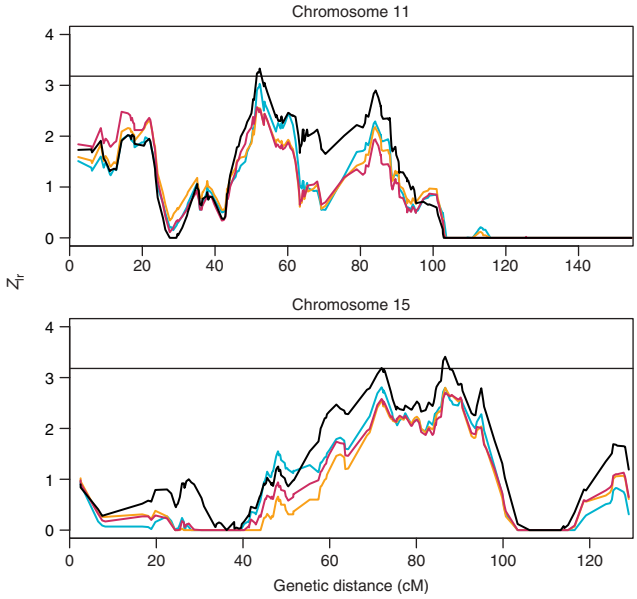

Figure 3 Highlighted linkage results due to removing families in which affected individuals carry putative CNVs. Results from all families (ignoring CNVs) noted by the cyan line; results from the filtered set, the orange line; results from the plate set, red line; and results from the batch set, black line. Families all fall in the broad diagnostic category. Complete results are in **Supplementary Figure 2**. For chromosome 11, the maximum occurs in 11p13 ($Z_{lr} = 3.33$ at rs2421826). For chromosome 15, there are two up-crossings: the smaller peak occurs at 15q23 ($Z_{lr} = 3.19$ at rs1372828) and the larger at 15q25.3 ($Z_{lr} = 3.41$ near rs1433452). For families removed versus those retained, heterogeneity of estimated identity-by-descent was tested in the ±5-cM linkage region surrounding each peak and reported as regional minimum heterogeneity *P* value m-P (11p12–p13, m-P = 0.074; 15q23, m-P = 0.044; 15q25.3, m-P = 0.004).

After removing 'CNV families', the data becomes suggestive for linkage in two regions, 11p12–p13 and 15q23–25.3 (**Fig. 3**), contrary to results from all broad families. The most noteworthy impact occurs for the batch method, which removes the greatest fraction of families. When families removed versus retained by the batch method are contrasted for identity-by-descent (IBD), heterogeneity is modest except for the 15q25.3 region (**Fig. 3**).

The 4:1 ratio of affected males to females, higher reported recurrence risk for siblings of female versus male probands[8] and published literature[30,31] suggest that a useful partition of the ASD families would be whether they contained affected females ('female-containing') or only affected males ('male-only'). Thus, we partitioned the families according to these criteria and recomputed linkage traces for each of the three nested groupings of diagnosis (**Fig. 4**). Consistent with theory[32], the female-containing families seem to be more informative for linkage. For instance, for the narrow diagnostic scheme, linkage traces cross the suggestive threshold three times, at 5p15.33, 9p24.1 and 11p13–12, whereas the traces do not approach this threshold for male-only families. Linkage traces for male-only families cross the suggestive threshold only for the most inclusive diagnostic level (5q12.3 and 9q33.3; **Fig. 4**) but not at the same locations as female-containing families. Although the differences between female-containing and male-only families in terms of linkage could be due to chance, tests of heterogeneity of IBD uncover substantial heterogeneity in 9p, 9q and 11p and modest heterogeneity in 5q (**Fig. 4**).

Dividing the data into subsets according to presence of CNVs and according to sex of affected individuals seems to generate more informative linkage signals. To test whether combining both approaches to form subsets would also be useful, we used the broad diagnostic grouping. Linkage using these six subsets (**Table 3**), female-containing versus male-only by three levels of CNV discovery, provides even more support for a risk locus in the vicinity of 11p12–p13 in female-containing families (**Fig. 5**). Removing families based on the batch method of CNV discovery nominates a 15q23 locus in

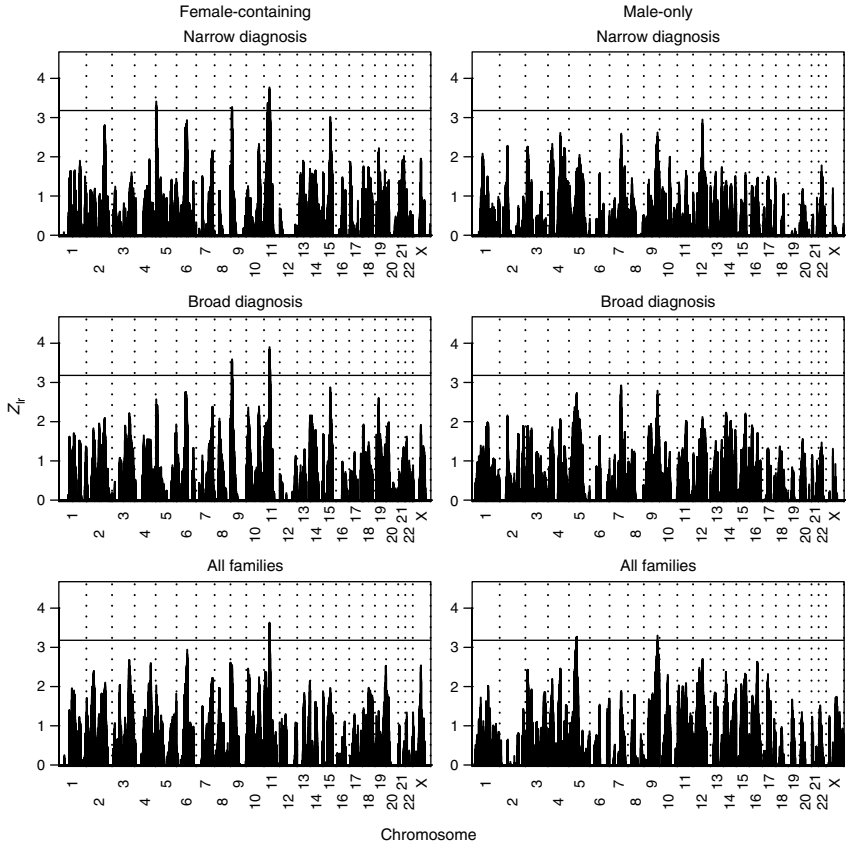

Figure 4 Linkage peaks by male-only versus female-containing families, based on levels of diagnostic certainty. For female-containing families and narrow diagnosis, peaks localize to 5p14.33 ($Z_{lr} = 3.41$ at rs1968011; m-P = 0.141), 9p24.1 ($Z_{lr} = 3.21$ at rs1340513; m-P = 0.0007) and 11p13 ($Z_{lr} = 3.77$ at rs1358054; m-P = 0.008); for female-containing families and broad diagnosis, to 9p24.1 ($Z_{lr} = 3.59$ at rs722628; m-P = 0.006) and 11p13 ($Z_{lr} = 3.90$ at rs1358054; m-P = 0.015) and for all female-containing families, to 11p12 ($Z_{lr} = 3.63$ at rs1039205; m-P = 0.078). For all male-only families, peaks localize to 5q12 ($Z_{lr} = 3.26$ at rs673743; m-P = 0.019) and 9q33.3 ($Z_{lr} = 3.30$ at rs536861; m-P = 0.0005).

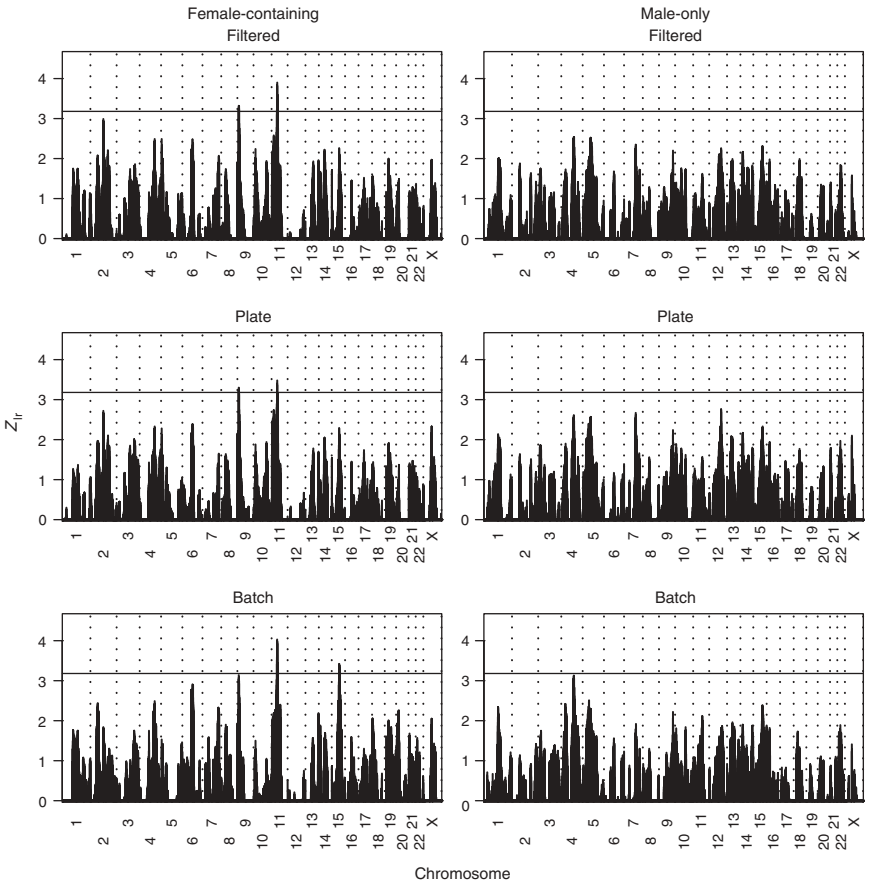

Figure 5 The effect on linkage of splitting families into female-containing and male-only families while also removing families in which affected individuals putatively carry CNVs. For female-containing families and the filtered subset, peaks localize to 9p24.1 (Z_{lr} = 3.32 at rs1575284; m-P = 0.105) and 11p12 (Z_{lr} = 3.90 at rs1039205; m-P = 0.411); for female-containing and the plate-by-plate subset, to 9p24.1 (Z_{lr} = 3.28 at rs1821892; m-P = 0.295) and 11p12 (Z_{lr} = 3.48 at rs1039205; m-P = 0.111); and for female-containing families and the batch subset, to 11p13 (Z_{lr} = 4.03 at rs1358054; m-P = 0.014) and 15q23 (Z_{lr} = 3.30 at rs1433452; m-P = 0.044).

female-containing families (**Fig. 5**), and the maximum at 11p12–p13 approaches genome-wide significance (4.03 versus 4.1). Affected individuals in female-containing families have a slightly elevated rate of CNV detection (2%–3%) relative to male-only families, regardless of CNV-calling method, and the broad diagnostic group of families has a similarly elevated rate of CNV detection relative to hASD; the 'paired' rates are not significantly different, nor is the interaction of these variables a significant predictor of the presence or absence of a detected CNV (data not shown).

Families participating in the AGP studies vary in their ancestry. To evaluate whether linkage would be strengthened by analyzing families of relatively homogeneous ancestry, we restricted the sample to 'European ancestry', as inferred by principal components[33] of SNP genotypes. We inferred that all founders in 995/1,168 (85.2%) families were of European origin (**Supplementary Fig. 3** online); inferred ancestry agreed with available self-reports (99.6%). Using this sample, we repeated all linkage analyses in **Figures 1** and **3–5**.

Exclusion of non-European families generally had only a modest impact on the results (see **Supplementary Fig. 4** online for complete results). Linkage traces at chromosome 11p remained prominent, especially for female-containing families, although the linkage region for 11p was much broader. By using the batch results to remove 'CNV families', a new region of suggestive linkage emerged: namely, 4q31.22, which also shows substantial heterogeneity of IBD ($P < 0.00006$) between retained and removed families. For all male-only families, the normalized likelihood ratio statistic for linkage (Z_{lr}) = 3.85 at 9q33.3. In addition, previously reported locations for linkage, especially 2q and 7q, gained more support from this 'European' subset of the data. For example, linkage traces approached genome-wide suggestive linkage at 2q31.1 (female-containing, narrow) and 7q22.2 (male-only, broad).

DISCUSSION

The results obtained from scanning the genomes of the largest cohort of ASD families yet assembled delineate a new understanding of the genetic basis for this complex disorder. It is widely understood that risk for ASD arises in small part from chromosomal copy number abnormalities (CNAs), such as duplications of chromosome 15q11–q13. Yet, for nuclear families containing two or more affected individuals and prescreened for microscopic CNAs, linkage analyses have implicitly assumed other CNAs would have little if any role in the heritable component of ASD. Our CNV results lie in stark contrast with this view. Instead, we find an appreciable number of families that could be assessed (68/590 or 11.5%, 36/476 or 7.6% and 34/350 or 9.7% in the batch, plate or filter analyses, respectively; **Table 2**), in which all affected individuals share possibly detrimental abnormalities (that is, possible CNAs). Owing to the relatively wide and uneven spacing of SNPs and our conservative approach of calling CNVs, we have missed many other events of this kind. By contrast, we also find a number of families in which only one of the affected relatives has a detected CNA. One possible implication of this finding is that, in these cases, relevant CNVs might be risk factors and not the only causal event. It is also possible that closely related individuals are etiologically heterogeneous.

With the goal of minimizing heterogeneity that might confound mapping of major loci conferring risk for ASDs, we invested substantial effort to standardize all phases of this multicenter project, including phenotypic assessment, sample ascertainment, genotyping and analysis. Linkage analyses based on a three-level diagnostic scheme produce suggestive evidence for linkage in the vicinity of 11p12–p13 (**Fig. 1**) for one level: all families. Relative to appropriate baseline, evidence for linkage at 11p12–p13 is amplified in select subsets of the data: (i) subsets obtained by removing families containing one or more affected individuals inferred to carry CNV (**Fig. 3**); (ii) the set of families containing affected females (**Fig. 4**) and (iii) subsets of families based on both sex of affected individuals and presence of CNVs (**Fig. 5**). For the subset of female-containing families without affected individuals who carry CNVs, the maximum Z_{lr} increases to 4.03. We believe these explorations motivate thorough fine-mapping of the 11p12–p13 region. Modest peaks for linkage have

been observed previously for this region, but 11p12–p13 has not been a major focus for discovery of autism risk loci.

Several regions have been featured prominently in previous linkage analyses: namely 2q, 7q and 17q. Of these regions, 2q and 7q garner modest linkage support from families of European ancestry (**Supplementary Fig. 4**), whereas the 17q linkage region does not. The largest linkage signal on 2q, $Z_{lr} = 3.1$, occurs for female-containing families in 2q31.1; the largest signal on 7q, $Z_{lr} = 3.1$, occurs for male-only families in 7q22.3. Several explanations are plausible for these results. First, the previous linkages could be false positives. Second, for linkage studies of complex disorders, statistics for identified linkage regions tend to be biased upward relative to that expected from the linked risk loci. Because of this bias, combining samples with mixed evidence for linkage and adding new families, as done here, will often diminish previously identified linkage signals. Third, studies could have collected samples that differ in heritable features tied to risk loci, exaggerating the heterogeneity already inherent in ASD. Stochastic variation combined with this heterogeneity could overwhelm the linkage signal.

None of our linkage results can be interpreted as 'statistically significant' because we have performed numerous analyses of the data. In fact, we performed 18 linkage analyses on the full sample and 18 using families of European ancestry. Because many of these analyses were performed on overlapping subsets of the data, we effectively performed the equivalent of four to five independent genome scans[34].

Our CNV analyses detect a hemizygous deletion of coding exons from *NRXN1* for a pair of affected siblings. Without other information, this finding might not be especially meaningful, in particular because of the prevalence of CNVs in the genome[21]. However, the alteration is a *de novo* event, and others[23] have reported rare missense variants in *NRXN1* in individuals with ASD that are not found in over 500 controls. Moreover, *NRXN1* interacts with neuroligins, for which rare mutations apparently generate risk for ASDs and mental retardation[14,15]. Therefore, we have evaluated transmissions in our families for four *NRXN1* SNPs (rs1363036, rs930752, rs1377238, rs2018909). Using the FBAT[35] empirical statistic, we have tested transmissions under additive and dominant models in all families and in the broad subset (dominant and recessive models are indistinguishable in this analysis). For all families, biased transmission is significant at two loci under the dominant model: the minor allele of rs1363036 ($P = 0.0091$) and the major allele of rs930752 ($P = 0.025$). (Only rs930752 showed significantly biased transmission under the additive model ($P = 0.014$).) These SNPs are in modest LD ($r^2 = 0.048$). In the broad subset, results are stronger for rs1363036 (dominant, $P = 0.0041$) but are weaker for rs930752 (dominant, $P = 0.076$; additive, $P = 0.072$). These two SNPs, separated by 88 kb, are intronic and are unlikely to convey risk directly.

Accumulating evidence thus implicates a role for neurexins and neuroligins in ASDs. For communication of signal between neurons, postsynaptic receptors must oppose neurotransmitter release sites on presynaptic axons. Neurexins have been shown to induce postsynaptic differentiation in contacting dendrites, while neuroligins induce presynaptic differentiation in glutamatergic axons[36]. The neurexin-neuroligin link is fundamentally important for glutamatergic synaptogenesis (and, apparently, GABAergic synaptogenesis[36,37]). Moreover, aberrant glutamate function is often cited as an important element of risk for ASDs[38,39], a hypothesis compatible with its role as the major excitatory neurotransmitter and critical factor in brain development[40]. Autism-like behaviors and diagnoses of autism are common for individuals with either Fragile-X syndrome or tuberous sclerosis, both of which are associated with dysregulated glutamate signaling[41,42].

Is oligogenic or 'major gene' variation associated with other glutamate-related genes? In addition to our results, intriguing evidence for association has been found for the mitochondrial aspartate/glutamate carrier *SLC25A12* (on 2q31) and *GRIK2* (ref. 13). Still, the protein product of *SLC25A12* is a mitochondrial aspartate-glutamate carrier not known to affect glutamatergic synaptic function. Knockout of *SLC25A12* in mice impairs myelination of neuronal cells resulting from limitations to aspartate delivery, not glutamate[43]. The protein product of *GRIK2*, GluR6, is an ionotropic kainate receptor that affects neuronal development. Based on its mapping to 6q16.3, it is not a positional candidate according to our results. For female-containing families, however, linkage results are modestly positive ($Z_{lr} = 1.9$ for all families and families of European ancestry; $Z_{lr} = \sim 2.40$ for families retained using the batch method of CNV calling).

The UCSC Genome Browser lists 168 genes with descriptions containing the keyword 'glutamate'. Many fall in linkage regions, including 11p13–12 (*SLC1A2* and *PRRG4*), 2q31 (*SLC25A12*), 4q28.3 (*SLC7A11*), 7q21.3 (*SLC25A13*), 9p24.2 (*SLC1A1*), 9q34.11 (*FGPS*) and 15q25.2 (*HOMER2*). Of ten glutamate solute carriers, half fall in the cited linkage regions (keywords 'glutamate' + 'solute' + 'carrier'), but not all are related to glutamatergic synaptic function. *SLC1A1* and *SLC1A2* fall close to linkage peaks, and their protein products affect glutamate synapse function and brain development; thus, they are excellent targets for positional candidate gene analyses.

METHODS

Linkage screening set. A SNP-based genome scan was conducted using the Affymetrix 10K v2 SNP array. Genotyping was contracted to the Translational Genomics Research Institute (TGEN). TGEN genotyped DNA samples falling into 1,496 nominal families, of which 1,168 could be used for linkage. Written permission was received from all adult participants and for all younger participants; procedures were approved by institutional review boards of all participating institutions.

Strategy for linkage analysis. For linkage analyses, we grouped families into three diagnostic classes: narrow, broad and heterogeneous ASD (hASD). To qualify for the narrow class, two or more affected individuals had to meet criteria for autism on both the ADI-R[18] and the ADOS[18]. For the broad category, at least one individual had to meet ADI-R criteria for autism and ADOS criteria for autism or ASD. At least one other family member had to meet criteria for impairment on the social or communication domains of the ADI-R and meet criteria for at least ASD on the ADOS. The hASD families were completely independent of the broad and narrow categories but were combined with the broad set to analyze linkage in all families. The hASD families consisted largely of either families that met ADI-R criteria for ASD or autism[18] but that did not undergo ADOS evaluation or families with multiple individuals per family who met ASD criteria by ADOS and demonstrated impairment on the social or communication domains of the ADI-R but did not meet full criteria for autism on the ADI-R. In addition to diagnostic categories, families were divided into subsets based on male-only or female-containing status.

Genetic quality control, tag SNP selection and ancestry. We evaluated three features of data quality: namely, degree of missing genotypes (missingness), mendelian errors and Hardy-Weinberg equilibrium (HWE). Individuals (20%) and loci (10%) with substantial missing data were not considered for linkage analyses because these features usually indicate poor DNA quality and problems with genotype calls, respectively. Likewise, loci with minor allele frequency (MAF) < 0.05 were discarded.

Mendelian errors were evaluated using PEDCHECK[44]. Loci showing multiple mendelian errors—for correct family structure—were discarded for linkage testing. To overcome possible problems arising owing to ancestry, we first selected tag SNPs and then evaluated HWE with the large sample inferred to be of homogeneous European ancestry. We analyzed linkage disequilibrium (LD)

using HCLUST[45], selecting tag SNPs to represent clusters of others in substantial LD ($r^2 > 0.8$). We chose those that were highly correlated with the other SNPs in the cluster; we estimated ancestry by using principal component analyses[33] and evaluated HWE using parental data. Loci were not used for analyses if HWE was rejected at a *P*-value < 0.005. Finally, we used MERLIN[46] to infer likely genotyping errors on the basis of apparent genetic recombination. When genotypes were likely to be errors ($P < 0.01$), they were set to 'missing'.

Linkage analysis. We used the BLUE method[47] to estimate allele frequencies. Linkage was estimated from the entire set of SNPs using MERLIN[46] and the exponential S-all statistic. Linkage was also estimated by using MERLIN[46] and ALLEGRO[48] from tag SNPs. We found virtually no difference in results using tag SNPs or all SNPs, with or without using the options in MERLIN to handle LD. We analyzed heterogeneity of linkage between strata using the methods of ref. 49, which test for significant differences in shared IBD among affected siblings in families. We computed heterogeneity statistics in each region or setting in which a linkage trace crossed the threshold for suggestive linkage.

CNV assessment. CNVs were inferred from Affymetrix 10K array scans using dChip 2006 software (DNA Chip Analyzer)[50]. We have also used other algorithms, and the data will be posted at the Autism Chromosome Rearrangement Database as it is validated. Initially, 7,610 scans were available (this number exceeds the 6,709 samples genotyped for linkage since the CNV experiments continued after the initial data freeze). We excluded those samples with a genotype call rate $<92\%$ and/or an array percentage outlier of $>5\%$, leaving 5,997 experiments suitable for CNV analyses. For the 'batch' analysis, we grouped the arrays into six cohorts of 1,000 samples each. The median probe intensities for the arrays varied greatly (<100 to $>1{,}000$), indicating the need for normalization to compare signals. Arrays were normalized at the probe intensity level using invariant set normalization to a baseline array within each group of 1,000 experiments[50]. A signal value was then calculated for each SNP using a model-based method and averaged across all samples for each SNP to obtain the mean signal of a diploid genome. The observed raw copy number was then defined and copy number inferred for each individual or SNP using a Hidden Markov Model[50]. As samples were submitted in 96-well plate formats and arrays were processed in the same manner, the 'batch' analysis contained plate-specific noise, apparently leading to many false positive CNV calls. In an attempt to increase the signal-to-noise ratio, we analyzed arrays in a 96-well plate-specific manner. For example, we excluded 12 plates having fewer than 40 samples after the initial filtering, leaving 5,823 scans for the plate analysis. Arrays were normalized within each set and copy number calculated in the same manner as was for the 'batch' analysis. We also excluded those samples with more than ten CNVs per sample from all analyses to avoid calling a high number of false positives. Because we had family data, certain CNV could be tentatively confirmed by using the family structure and mendelian errors (although the original CNV calls were blind to family status). Using these data as a benchmark, the 'plate' analysis produced a cleaner data set than the 'batch' analysis and was therefore parsed further using a combination of more stringent thresholds (fewer than five CNVs per sample) and manual curation of the raw data to give a 'filtered' data set. The inferred CNVs for all three data sets were interpreted on several levels. We also completed similar analysis for the X chromosome but did not include the results, as only 263 SNPs covered this segment of the genome (40 CNVs were found). Called CNVs were also examined for overlap with genomic features including mapped chromosome rearrangement breakpoints annotated in the Autism Chromosome Rearrangement Database and polymorphic CNVs in the Database of Genomic Variants. For all three analyses, affected individuals and families with CNVs that did not have complete overlap with the DGV were removed from linkage analysis.

Accession codes. Gene Omnibus Expression (GEO): raw data from the Affymetrix 10K experiments, GSE6754.

URLs. Autism Chromosome Rearrangement Database: http://projects.tcag.ca/autism/; Database of Genomic Variants: http://projects.tcag.ca/variation/; GEO: http://www.ncbi.nlm.nih.gov/geo/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

The authors are indebted to the participating families for their contribution of time and effort in support of this study. We gratefully acknowledge Autism Speaks, formerly the National Alliance for Autism Research, for financial support for data pooling, SNP genotyping and data analysis.

The Autism Genetics Cooperative thanks Assistance Publique-Hôpitaux de Paris, Canadian Institutes for Health Research (CIHR grant 11350 to P.S.), Catherine and Maxwell Meighan Foundation, Fondation de France, Fondation France Télécom, Fondation pour la Recherche Médicale, Genome Canada/Ontario Genomics Institute, The Hospital for Sick Children Foundation, Howard Hughes Medical Institute, INSERM, McLaughlin Centre for Molecular Medicine, National Institute of Child Health and Human Development, National Institute of Mental Health (MH066673 to J.D.B.; MH55135 to S.E.F.; MH52708 to N. Risch (University of California, San Francisco); MH061009 to J.S.S.), National Institute of Neurological Disorders and Stroke (NS042165 to J.H.; NS026630 and NS036738 to M.A.P.-V.; NS049261 to J.S.S.; NS043550 to T.H.W.), Swedish Science Council, Seaver Autism Research Foundation and The Centre for Applied Genomics (Toronto). S.W.S. is an Investigator of the CIHR and an HHMI International Scholar.

The Autism Genetic Resource Exchange Consortium gratefully acknowledges the resources provided by the participating families. The Autism Genetic Resource Exchange is a program of Cure Autism Now and is supported, in part, by the National Institute of Mental Health (MH64547 to D.H.G.).

The Collaborative Programs of Excellence thank the National Center for Research Resources (M01-RR00064), National Institute of Child Health and Human Development (U19HD34565 G.D. and G.S.), NIMH (MH057881), NINDS (5 U19 HD035476 to W.M.McM.) and the Utah Autism Foundation.

The International Molecular Genetic Study of Autism Consortium thanks the UK Medical Research Council, Wellcome Trust, BIOMED 2 (CT-97-2759), EC Fifth Framework (QLG2-CT-1999-0094), Telethon-Italy (GGP030227), Janus Korczak Foundation, Deutsche Forschungsgemeinschaft, Fondation France Telecom, Conseil Regional Midi-Pyrenees, Danish Medical Research Council, Sofiefonden, Beatrice Surovell Haskells Fond for Child Mental Health Research of Copenhagen, Danish Natural Science Research Council (9802210) and the US National Institutes of Health (U19 HD35482, MO1 RR06022, K05 MH01196, K02 MH01389). A.J.B. is the Cheryl and Reece Scott Professor of Psychiatry. A.P.M. is a Wellcome Trust Principal Research Fellow.

Requests for data or methods should be addressed to B.D. (devlinbj@upmc.edu) or S.W.S. (steve@genet.sickkids.on.ca).

AUTHOR CONTRIBUTIONS

P.S., A.D.P., S.W.S., V.J.V., M.A.P.-V., C.B., J.D.B., J.H., J.S.S., J.H., J.L.H., J.P., T.H.W., D.H.G., R.C., S.N., G.D., G.D.S., B.D., W.M.M., E.M.W., A.J.B., A.P.M. and E.H.C. were lead AGP investigators and contributed equally to this project.

The Autism Genome Project (AGP). The AGP comprises four existing consortia of partners or countries (listed alphabetically below).

Autism Genetics Cooperative. Canagen: P.S., A.D.P., L.Z., W.R., J.B., X.-Q.L., J.B.V., J.L.S., A.P.T., L.S., L.F., C.Q., S.E.B., M.B.J., C.R.M. and S.W.S. Iowa Data Coordinating Center: V.J.V., C.B., L.V.M., R.G. and A.S. University of Miami: M.A.P.-V., M.L.C. and J.R.G. University of South Carolina: H.H.W., R.K.A. Paris Autism Research: International Sibpair Study: C.B., T.B., C.G. and M.L. Seaver Autism Research Center: J.D.B., K.L.D., E.H. and J.M.S. Stanford University: J.H. and L.L. Vanderbilt University: J.S.S., J.L.H. and S.E.F. University of North Carolina/University of Iowa: J.P., T.H.W., K.J.M. and V.S.

The Autism Genetic Resource Exchange Consortium. D.H.G., M.B., W.T.B., R.M.C., J.N.C., T.C.G., M.H., C.LaJ., D.H.L., C.L.-M., J.M., S.N. C.A.S.-S., S.S., M.S. and R.E.T.

The Collaborative Programs of Excellence in Autism. H.C., G.D., B.D., A.E., P.F., L.K., W.M.McM., N.M., J.M., E.K., P.M.R., G.D.S., M.S., M.A.S., C.S., P.G.T., E.M.W. and C.-E.Y.

The International Molecular Genetic Study of Autism Consortium. France: B.R., C.M., K.W. and M.T. Germany: A.P., B.F., S.M.K., C.S., F.P., S.B., S.F.-M., E.H. and G.S. Greece: J.T., K.P. Italy: E.M., E.B., F.B., S.C. and C.T. The Netherlands: H.V.E., M. de J., C.K., F.K., M.L., C.H. and W.G.S. UK: G.B., P.F.B., M.L.R., E.W., J.G., C.A., J.-A.W., A.P., A.LeC., T.B., H.McC., A.J.B., K.F., G.H., A.H., J.R.P., S.W., A.P.M., G.B., K.K., J.A.L., I.S. and N.S. USA: E.H.C., S.J.G., B.L.L., J.S., C.L., C.C., V.H., D.E.W. and F.V. Canada: E.F.

Scientific management. A.S.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Published online at http://www.nature.com/naturegenetics
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. American Psychological Association. *Diagnostic and Statistical Manual of Mental Disorders* (American Psychological Association, Washington, D.C., 1994).
2. Chakrabarti, S. & Fombonne, E. Pervasive developmental disorders in preschool children: confirmation of high prevalence. *Am. J. Psychiatry* **162**, 1133–1141 (2005).
3. Veenstra-Vanderweele, J., Christian, S.L. & Cook, E.H., Jr. Autism as a paradigmatic complex genetic disorder. *Annu. Rev. Genomics Hum. Genet.* **5**, 379–405 (2004).
4. Xu, J., Zwaigenbaum, L., Szatmari, P. & Scherer, S.W. Molecular cytogenetics of autism. *Curr. Genomics* **5**, 347–364 (2004).
5. Bailey, A. *et al.* Autism as a strongly genetic disorder: evidence from a British twin study. *Psychol. Med.* **25**, 63–77 (1995).
6. Piven, J. The broad autism phenotype: a complementary strategy for molecular genetic studies of autism. *Am. J. Med. Genet.* **105**, 34–35 (2001).
7. Jones, M.B. & Szatmari, P. Stoppage rules and genetic studies of autism. *J. Autism Dev. Disord.* **18**, 31–40 (1988).
8. Ritvo, E.R. *et al.* The UCLA-University of Utah epidemiologic survey of autism: prevalence. *Am. J. Psychiatry* **146**, 194–199 (1989).
9. Hallmayer, J. *et al.* On the twin risk in autism. *Am. J. Hum. Genet.* **71**, 941–946 (2002).
10. Pickles, A. *et al.* Latent-class analysis of recurrence risks for complex phenotypes with selection and measurement error: a twin and family history study of autism. *Am. J. Hum. Genet.* **57**, 717–726 (1995).
11. Risch, N. *et al.* A genomic screen of autism: evidence for a multilocus etiology. *Am. J. Hum. Genet.* **65**, 493–507 (1999).
12. Schellenberg, G.D. *et al.* Evidence for genetic linkage of autism to chromosomes 7 and 4. *Mol. Psychiatry* **11**, 979 (2006).
13. Freitag, C.M. The genetics of autistic disorders and its clinical relevance: a review of the literature. *Mol. Psychiatry* **12**, 2–22 (2007).
14. Jamain, S. *et al.* Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nat. Genet.* **34**, 27–29 (2003).
15. Laumonnier, F. *et al.* X-linked mental retardation and autism are associated with a mutation in the *NLGN4* gene, a member of the neuroligin family. *Am. J. Hum. Genet.* **74**, 552–557 (2004).
16. Chubykin, A.A. *et al.* Dissection of synapse induction by neuroligins: effect of a neuroligin mutation associated with autism. *J. Biol. Chem.* **280**, 22365–22374 (2005).
17. Durand, C.M. *et al.* Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. *Nat. Genet.* **39**, 25–27 (2007).
18. Risi, S. *et al.* Combining information from multiple sources in the diagnosis of autism spectrum disorders. *J. Am. Acad. Child Adolesc. Psychiatry* **45**, 1094–1103 (2006).
19. Kong, X. *et al.* A combined linkage-physical map of the human genome. *Am. J. Hum. Genet.* **75**, 1143–1148 (2004).
20. Lander, E. & Kruglyak, L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat. Genet.* **11**, 241–247 (1995).
21. Redon, R. *et al.* Global variation in copy number in the human genome. *Nature* **444**, 445–454 (2006).
22. Persico, A.M. & Bourgeron, T. Searching for ways out of the autism maze: genetic, epigenetic and environmental clues. *Trends Neurosci.* **29**, 349–358 (2006).
23. Feng, J. *et al.* High frequency of neurexin 1β signal peptide structural variants in patients with autism. *Neurosci. Lett.* **409**, 10–13 (2006).
24. de Vries, B.B. *et al.* Diagnostic genome profiling in mental retardation. *Am. J. Hum. Genet.* **77**, 606–616 (2005).
25. Sharp, A.J. *et al.* Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. *Nat. Genet.* **38**, 1038–1042 (2006).
26. Houlden, H. & Reilly, M.M. Molecular genetics of autosomal-dominant demyelinating Charcot-Marie-Tooth disease. *Neuromolecular Med.* **8**, 43–62 (2006).
27. Potocki, L. *et al.* Molecular mechanism for duplication 17p11.2, the homologous recombination reciprocal of the Smith-Magenis microdeletion. *Nat. Genet.* **24**, 84–87 (2000).
28. Moog, U. *et al.* Hereditary motor and sensory neuropathy (HMSN) IA, developmental delay and autism related disorder in a boy with duplication (17)(p11.2p12). *Genet. Couns.* **15**, 73–80 (2004).
29. Iafrate, A.J. *et al.* Detection of large-scale variation in the human genome. *Nat. Genet.* **36**, 949–951 (2004).
30. Lamb, J.A. *et al.* Analysis of IMGSAC autism susceptibility loci: evidence for sex limited and parent of origin specific effects. *J. Med. Genet.* **42**, 132–137 (2005).
31. Stone, J.L. *et al.* Evidence for sex-specific risk alleles in autism spectrum disorder. *Am. J. Hum. Genet.* **75**, 1117–1123 (2004).
32. Falconer, D.S. *Introduction to Quantitative Genetics* (Longman, London, 1981).
33. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
34. Camp, N.J. & Farnham, J.M. Correcting for multiple analyses in genomewide linkage studies. *Ann. Hum. Genet.* **65**, 577–582 (2001).
35. Lange, C. & Laird, N.M. On a general class of conditional tests for family-based association studies in genetics: the asymptotic distribution, the conditional power, and optimality considerations. *Genet. Epidemiol.* **23**, 165–180 (2002).
36. Graf, E.R., Zhang, X., Jin, S.X., Linhoff, M.W. & Craig, A.M. Neurexins induce differentiation of GABA and glutamate postsynaptic specializations via neuroligins. *Cell* **119**, 1013–1026 (2004).
37. Varoqueaux, F. *et al.* Neuroligins determine synapse maturation and function. *Neuron* **51**, 741–754 (2006).
38. Purcell, A.E., Jeon, O.H., Zimmerman, A.W., Blue, M.E. & Pevsner, J. Postmortem brain abnormalities of the glutamate neurotransmitter system in autism. *Neurology* **57**, 1618–1628 (2001).
39. Shinohe, A. *et al.* Increased serum levels of glutamate in adult patients with autism. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **30**, 1472–1477 (2006).
40. Kugler, P. & Schleyer, V. Developmental expression of glutamate transporters and glutamate dehydrogenase in astrocytes of the postnatal rat hippocampus. *Hippocampus* **14**, 975–985 (2004).
41. Belmonte, M.K. & Bourgeron, T. Fragile X syndrome and autism at the intersection of genetic and neural networks. *Nat. Neurosci.* **9**, 1221–1225 (2006).
42. Tavazoie, S.F., Alvarez, V.A., Ridenour, D.A., Kwiatkowski, D.J. & Sabatini, B.L. Regulation of neuronal morphology and function by the tumor suppressors Tsc1 and Tsc2. *Nat. Neurosci.* **8**, 1727–1734 (2005).
43. Jalil, M.A. *et al.* Reduced N-acetylaspartate levels in mice lacking aralar, a brain- and muscle-type mitochondrial aspartate-glutamate carrier. *J. Biol. Chem.* **280**, 31333–31339 (2005).
44. O'Connell, J.R. & Weeks, D.E. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am. J. Hum. Genet.* **63**, 259–266 (1998).
45. Rinaldo, A. *et al.* Characterization of multilocus linkage disequilibrium. *Genet. Epidemiol.* **28**, 193–206 (2005).
46. Abecasis, G.R., Cherny, S.S., Cookson, W.O. & Cardon, L.R. Merlin–rapid analysis of dense genetic maps using sparse gene flow trees. *Nat. Genet.* **30**, 97–101 (2002).
47. McPeek, M.S., Wu, X. & Ober, C. Best linear unbiased allele-frequency estimation in complex pedigrees. *Biometrics* **60**, 359–367 (2004).
48. Gudbjartsson, D.F., Jonasson, K., Frigge, M.L. & Kong, A. Allegro, a new computer program for multipoint linkage analysis. *Nat. Genet.* **25**, 12–13 (2000).
49. McQueen, M.B., Blacker, D. & Laird, N.M. Variance calculations for identity-by-descent estimation. *Am. J. Hum. Genet.* **78**, 914–921 (2006).
50. Li, C. & Wong, W.H. DNA-chip analyzer (dChip). in *The Analysis of Gene Expression Data: Methods and Software* (eds. Parmigiani, G., Garrett, E.S., Irizarry, R. & Zeger, S.L.) (Springer, New York, 2001).

Peter Szatmari[1,68], Andrew D Paterson[2,68], Lonnie Zwaigenbaum[3], Wendy Roberts[4], Jessica Brian[4], Xiao-Qing Liu[2], John B Vincent[5], Jennifer L Skaug[2], Ann P Thompson[1], Lili Senman[4], Lars Feuk[2], Cheng Qian[2], Susan E Bryson[6], Marshall B Jones[7], Christian R Marshall[2], Stephen W Scherer[2,68], Veronica J Vieland[8,68], Christopher Bartlett[8], La Vonne Mangin[8], Rhinda Goedken[9], Alberto Segre[9], Margaret A Pericak-Vance[10,68], Michael L Cuccaro[10], John R Gilbert[10], Harry H Wright[11], Ruth K Abramson[11], Catalina Betancur[12,68], Thomas Bourgeron[13], Christopher Gillberg[14], Marion Leboyer[12,15], Joseph D Buxbaum[16,68], Kenneth L Davis[16], Eric Hollander[16], Jeremy M Silverman[16], Joachim Hallmayer[17,68], Linda Lotspeich[17], James S Sutcliffe[18,19,68], Jonathan L Haines[19,68], Susan E Folstein[20], Joseph Piven[21,68], Thomas H Wassink[22,68], Kacie J Meyer[22], Val Sheffield[22], Daniel H Geschwind[23,68], Maja Bucan[24], W Ted Brown[25], Rita M Cantor[26,68], John N Constantino[27], T Conrad Gilliam[28], Martha Herbert[29], Clara LaJonchere[30], David H Ledbetter[31], Christa Lese-Martin[31], Janet Miller[30], Stan Nelson[26,68], Carol A Samango-Sprouse[32], Sarah Spence[33], Matthew State[34], Rudolph E Tanzi[35], Hilary Coon[36],

Geraldine Dawson[37,38,68], Bernie Devlin[39,68], Annette Estes[38], Pamela Flodman[40], Lambertus Klei[39], William M McMahon[36,68], Nancy Minshew[39], Jeff Munson[37], Elena Korvatska[41,42], Patricia M Rodier[43], Gerard D Schellenberg[41,42,44,68], Moyra Smith[40], M Anne Spence[40], Chris Stodgell[43], Ping Guo Tepper, Ellen M Wijsman[45,46,68], Chang-En Yu[41,42], Bernadette Rogé[47], Carine Mantoulan[47], Kerstin Wittemeyer[47], Annemarie Poustka[48], Bärbel Felder[48], Sabine M Klauck[48], Claudia Schuster[48], Fritz Poustka[49], Sven Bölte[49], Sabine Feineis-Matthews[49], Evelyn Herbrecht[49], Gabi Schmötzer[49], John Tsiantis[50], Katerina Papanikolaou[50], Elena Maestrini[51], Elena Bacchelli[51], Francesca Blasi[51], Simona Carone[51], Claudio Toma[51], Herman Van Engeland[52], Maretha de Jonge[52], Chantal Kemner[52], Frederieke Koop[52], Marjolein Langemeijer[52], Channa Hijmans[52], Wouter G Staal[52], Gillian Baird[53], Patrick F Bolton[54], Michael L Rutter[55], Emma Weisblatt[56], Jonathan Green[57], Catherine Aldred[57], Julie-Anne Wilkinson[57], Andrew Pickles[58], Ann Le Couteur[59], Tom Berney[59], Helen McConachie[59], Anthony J Bailey[60,68], Kostas Francis[60], Gemma Honeyman[60], Aislinn Hutchinson[60], Jeremy R Parr[60], Simon Wallace[60], Anthony P Monaco[61,68], Gabrielle Barnby[61], Kazuhiro Kobayashi[61], Janine A Lamb[61], Ines Sousa[61], Nuala Sykes[61], Edwin H Cook[62,68], Stephen J Guter[62], Bennett L Leventhal[62], Jeff Salt[62], Catherine Lord[63], Christina Corsello[63], Vanessa Hus[63], Daniel E Weeks[64], Fred Volkmar[34], Maïté Tauber[65], Eric Fombonne[66] & Andy Shih[67]

[1]Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton, Ontario L8N 3Z5, Canada. [2]The Centre for Applied Genomics and Program in Genetics and Genome Biology, The Hospital for Sick Children and University of Toronto, Toronto, Ontario M5G 1X8, Canada. [3]Department of Pediatrics, McMaster University, Hamilton, Ontario L8N 3Z5 Canada. [4]Autism Research Unit, The Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada. [5]Centre for Addiction and Mental Health, Clarke Institute and Department of Psychiatry, University of Toronto, Toronto, Ontario M5G 1X8, Canada. [6]Departments of Pediatrics and Psychology, Izaak Walton Killam Health Centre – Dalhousie University, Halifax, Nova Scotia B3K 6R8, Canada. [7]Department of Neural and Behavioral Sciences, The Pennsylvania State University College of Medicine, Hershey, Pennsylvania 17033, USA. [8]Columbus Children's Research Institute, Center for Quantitative and Computational Biology, Columbus, Ohio 43205 USA. [9]University of Iowa, Department of Computer Science, Iowa City, Iowa 52242, USA. [10]Miami Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, Florida 33101, USA. [11]W.S. Hall Psychiatric Institute, University of South Carolina, Columbia, South Carolina 29208, USA. [12]INSERM U513, Université Paris XII, Créteil 94000, France. [13]Human Genetics and Cognitive Functions, Institut Pasteur, Paris 75015, France. [14]Department of Child and Adolescent Psychiatry, Goteborg University, Goteborg 41119, Sweden. [15]Department of Psychiatry, Groupe hospitalier Henri Mondor-Albert Chenevier, AP-HP, Créteil 94000, France. [16]Seaver Autism Research Center and the Greater New York Autism Research Center for Excellence, Department of Psychiatry, Mount Sinai School of Medicine, New York, New York 10029, USA. [17]Stanford University, Department of Psychiatry, Stanford, California 94304, USA. [18]Center for Molecular Neuroscience, Vanderbilt University, Nashville, Tennessee 37232, USA. [19]Center for Human Genetics Research, Vanderbilt University, Nashville 37232, Tennessee, USA. [20]Department of Psychiatry, Johns Hopkins University, Baltimore, Maryland 21287, USA. [21]University of North Carolina, Chapel Hill, North Carolina 27599, USA. [22]University of Iowa, Iowa City, Iowa 52242, USA. [23]Department of Neurology, University of California - Los Angeles (UCLA) School of Medicine, Los Angeles, California 90095, USA. [24]University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. [25]New York State Institute for Basic Research in Developmental Disabilities, Staten Island, New York 10314, USA. [26]Department of Human Genetics, UCLA School of Medicine, Los Angeles, California 90095, USA. [27]Washington University School of Medicine, St. Louis, Missouri, USA. [28]University of Chicago, Chicago 60637, Illinois, USA. [29]Harvard Medical School, Boston, Massachusetts 02115, USA. [30]Cure Autism Now, Los Angeles, California 90036, USA. [31]Emory University, Atlanta, Georgia 30322, USA. [32]George Washington University, Washington, D.C. 20052, USA. [33]UCLA, Los Angeles, California 90095, USA. [34]Yale University Child Study Center, New Haven, Connecticut 06520, USA. [35]Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [36]Department of Psychiatry, University of Utah, Salt Lake City 84132, USA. [37]Department of Psychology and the Center on Human Development and Disability and [38]Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, Washington 98195, USA. [39]Department of Psychiatry, University of Pittsburgh, Pittsburgh, Pennsylvania 15213, USA. [40]Department of Pediatrics, University of California, Irvine, California 92697, USA. [41]Geriatrics Research Education and Clinical Center, Puget Sound Veterans Affairs Medical Center, Seattle, Washington 98108, USA. [42]Gerontology and Geriatric Medicine, Department of Medicine, University of Washington, Seattle, Washington 98104, USA. [43]Department of OB/GYN, University of Rochester Medical Center, Rochester, New York 14642, USA. [44]Departments of Neurology and Pharmacology, University of Washington, Seattle, Washington 98195, USA. [45]Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, Washington 98195, USA. [46]Department of Biostatistics, University of Washington, Seattle, Washington 98195, USA. [47]Université de Toulouse Le Mirail, Centre d'Etudes et de Recherches en Psychopathologie (CERPP), 31058 Toulouse, France. [48]Deutsches Krebsforschungszentrum (DKFZ), Division of Molecular Genome Analysis, D-69120 Heidelberg, Germany. [49]J W Goethe Universität Frankfurt, Klinik für Psychiatrie und Psychotherapie des Kindes- und Jugendalters, D-60528 Frankfurt, Germany. [50]National and Kapodistrian University of Athens, Department of Child Psychiatry, "Agia Sophia" Children's Hospital, Athens 115 27, Greece. [51]Department of Biology, University of Bologna, 40126 Bologna, Italy. [52]University Medical Center, Utrecht, Department of Child and Adolescent Psychiatry, 3508 GA Utrecht, The Netherlands. [53]Guy's Hospital, Newcomen Centre, London SE1 9RT, UK. [54]Institute of Psychiatry, Department of Child and Adolescent Psychiatry, London, UK. [55]Institute of Psychiatry, Social, Genetic and Developmental Psychiatry Centre (SGDP), London SE5 8AF, UK. [56]University of Cambridge Clinical School, Cambridge CB2 2AH, UK. [57]University of Manchester Department of Child Psychiatry, Booth Hall Children's Hospital, Manchester M9 7AA, UK. [58]University of Manchester School of Epidemiology and Health Science, Manchester M13 9P, UK. [59]University of Newcastle, Child and Adolescent Mental Health, Sir James Spence Institute, Newcastle upon Tyne NE1 4LP, UK. [60]University of Oxford Department of Psychiatry, Oxford OX3 7LQ, UK. [61]Wellcome Trust Centre for Human Genetics, Oxford OX3 7BN, UK. [62]Institute for Juvenile Research, Department of Psychiatry, University of Illinois at Chicago, Chicago, Illinois, 60608, USA. [63]University of Michigan Autism and Communicative Disorders Center (UMACC), Ann Arbor, Michigan, 48109, USA. [64]University of Pittsburgh Department of Human Genetics, Pittsburgh, Pennsylvania 15213, USA. [65]Service de pédiatrie et génétique médicale, CHU de Toulouse, Toulouse, France. [66]Montreal Children's Hospital, McGill University, Montreal, Canada. [67]Autism Speaks, 2 Park Avenue, New York, New York 10016, USA. [68]These authors contributed equally to this work.

Corrigendum: Mapping autism risk loci using genetic linkage and chromosomal rearrangements

The Autism Genome Project Consortium
Nat. Genet. **39**, 319–328 (2007); **published online 18 February; corrected after print 26 September 2007**

In the version of this article originally published, Kacie J. Meyer (University of Iowa, Iowa City) was inadvertently omitted from the list of authors, and the names of three authors (Frederieke Koop, Marjolein Langemeijer and Channa Hijmans) were misspelled. Also, the third sentence of the abstract incorrectly stated that 1,168 families were analyzed. The correct number is 1,181 families. Finally, the last paragraph of the Discussion mistakenly identified one of the linkage regions as 11q13-12. This should read 11p13-12. These errors have been corrected in the HTML and PDF versions of the article.

Corrigendum: Nuclear reprogramming of cloned embryos and its implications for therapeutic cloning

Xiangzhong Yang, Sadie L Smith, X Cindy Tian, Harris A Lewin, Jean-Paul Renard & Teruhiko Wakayama
Nat. Genet. **39**, 295–302 (2007); **published online 26 February; corrected after print 26 September 2007**

In the version of this article initially published, the legend for Figure 1 is inaccurate. The original legend for Figure 1c failed to note that the cells derived from cattle embryos were embryonic stem cell (ESC)-like or trophectoderm cell colonies. The full legend for Figure 1c should read "Therapeutic cloning: percentage ESC line derivation (mice) or ESC-like or trophectoderm cell colony formation (cattle) from embryos derived from both IVF and nuclear transfer." This error has been corrected in the HTML and PDF versions of the article.

Corrigendum: Variation in *FTO* contributes to childhood obesity and severe adult obesity

Christian Dina, David Meyre, Sophie Gallina, Emmanuelle Durand, Antje Körner, Peter Jacobson, Lena M S Carlsson, Wieland Kiess, Vincent Vatin, Cecile Lecoeur, Jérome Delplanque, Emmanuel Vaillant, François Pattou, Juan Ruiz, Jacques Weill, Claire Levy-Marchal, Fritz Horber, Natascha Potoczna, Serge Hercberg, Catherine Le Stunff, Pierre Bougnères, Peter Kovacs, Michel Marre, Beverley Balkau, Stéphane Cauchi, Jean-Claude Chèvre & Philippe Froguel
Nat. Genet. **39**, 724–726 (2007); **published online 13 May; corrected after print 26 September 2007**

In the version of this article initially published, the authors failed to acknowledge that recruitment of obese cases was supported by both Assistance-Publique Hôpitaux de Paris and Centre National de la Recherche Scientifique. This error has been corrected in the PDF version of the article.