A GENOME-WIDE LINKAGE AND ASSOCIATION SCAN REVEALS NOVEL LOCI FOR AUTISM

Lauren A. Weiss[*], **Dan E. Arking**[*], and **The Gene Discovery Project of Johns Hopkins the Autism Consortium**[**]

Summary

Although autism is a highly heritable neurodevelopmental disorder, attempts to identify specific susceptibility genes have thus far met with limited success [1]. Genome-wide association studies (GWAS) using half a million or more markers, particularly those with very large sample sizes achieved through meta-analysis, have shown great success in mapping genes for other complex genetic traits (http://www.genome.gov/26525384). Consequently, we initiated a linkage and association mapping study using half a million genome-wide SNPs in a common set of 1,031 multiplex autism families (1,553 affected offspring). We identified regions of suggestive and significant linkage on chromosomes 6q27 and 20p13, respectively. Initial analysis did not yield genome-wide significant associations; however, genotyping of top hits in additional families revealed a SNP on chromosome 5p15 (between *SEMA5A* and *TAS2R1*) that was significantly associated with autism ($P = 2 \times 10^{-7}$). We also demonstrated that expression of *SEMA5A* is reduced in brains from autistic patients, further implicating *SEMA5A* as an autism susceptibility gene. The linkage regions reported here provide targets for rare variation screening while the discovery of a single novel association demonstrates the action of common variants.

For a high-resolution genetic study of autism, we selected families with multiple affected individuals (multiplex) from the widely studied Autism Genetic Resource Exchange (AGRE) and US National Institute for Mental Health (NIMH) repositories (Supplementary Methods, Supplementary Table 1). Although the phenotypic heterogeneity in autism spectrum disorders is extensive, in our primary screen we selected families in which at least one proband met ADI-R criteria for diagnosis of autism and included additional siblings in the same nuclear family

affected with any autism spectrum disorder. We previously reported an early copy number analysis that revealed a significant role for microdeletion and duplication of 16p11.2 in ASD causation [2]; here, we present extensive genome-wide linkage and association analyses performed with this high density of SNPs and identify independent and novel genome-wide significant results by both linkage and association analyses.

A PUBLIC AUTISM DATASET

We combined families and samples from two sources for the primary genetic association screen. The AGRE sample included nearly 3,000 individuals from over 780 multiplex autism families in the AGRE collection [3] genotyped at the Broad Institute on the Affymetrix 5.0 platform, which includes over 500,000 SNPs. The NIMH sample included a total of 1,233 individuals from 341 multiplex nuclear families (258 of which were independent of the AGRE sample) genotyped at the Johns Hopkins Center for Complex Disease Genomics on Affymetrix 5.0 and 500K platforms, including the same SNP markers as were genotyped in the AGRE sample.

Before merging, we carefully filtered each data set separately to ensure the highest possible genotype quality for analysis, since technical genotyping artifacts can create false positive findings. We therefore examined the distribution of χ^2 values for the highest quality data, and used a series of quality control (QC) filters designed to identify a robust set of SNPs, including data completeness for each SNP, Mendelian errors per SNP and per family, and a careful evaluation of inflation of association statistics as a function of allele frequency and missing data (see Methods). As 324 individuals were genotyped at both centers, we performed a concordance check to validate our approach. After excluding one sample mix-up, we obtained an overall genotype concordance between the two centers of 99.7% for samples typed on 500K at JHU and 5.0 at Broad and 99.9% for samples run on 5.0 arrays at both sites. The combined dataset, consisting of 1,031 nuclear families (856 with two parents) and a total of 1,553 affected offspring, was employed for genetic analyses (Supplementary Table 1). These data were publicly released in October, 2007 and are directly available from AGRE and NIMH.

For linkage analyses, the common AGRE/NIMH dataset was further merged with Illumina 550K genotype data generated at the Children's Hospital of Philadelphia (CHOP) and available from AGRE, adding ~300 nuclear families (1,499 samples). We used the extensive overlap of samples between the AGRE/NIMH and the CHOP datasets (2,282 samples) to select an extremely high quality set of SNPs for linkage analysis. Specifically, we only included SNPs genotyped in both datasets with >99.5% concordance and ≤1 Mendelian error.

LINKAGE ANALYSIS

Linkage analysis involving high densities of markers, where clusters of markers are in linkage disequilibrium (LD), can falsely inflate the evidence for genetic sharing among siblings when neither parent is genotyped [4]. To alleviate these concerns, we analyzed a pruned set of 16,311 highly polymorphic, high-quality autosomal SNPs which were filtered to remove any instances in which two nearby markers were correlated with $r^2 > 0.1$, providing a marker density of ~0.25cM (see Methods). In this analysis of 878 families, four genomic regions showed LOD scores in excess of 2.0 and one region, 20p13, exceeded the formal genome-wide significance threshold of 3.6 [5] (maximum LOD, 3.81; Figure 1a, Supplementary Table 2). Restricting analysis to only those families with both parents genotyped (784 families) showed that these results are not an artifact of missing parental data (Figure 1b). We further tested the stability of these results by varying the recombination map and halving the marker density by placing every other marker into two non-overlapping SNP sets (Methods Summary); all analyses showed consistent and strong linkage to the same regions (data not shown).

FAMILY-BASED ASSOCIATION ANALYSIS

We used the transmission disequilibrium test (TDT) across all SNPs passing quality control in the complete family dataset for association analyses since the TDT is not biased by population stratification. We estimated a threshold for genome-wide significance using both permutation ($P < 2.5 \times 10^{-7}$) and estimating the effective number of tests ($P < 3.4 \times 10^{-7}$) and use the more conservative, here (see Methods). No SNP met criteria for genome-wide significance at $P < 2.5 \times 10^{-7}$. However, we observed an excess of independent regions associated at $P < 10^{-5}$ (6 observed vs. 1 expected) and $P < 10^{-4}$ (30 observed vs. 15 expected) despite the lack of overall statistical inflation ($\lambda = 1.03$, Supplementary Figure 1), suggesting that common variants in autism exist, but that our initial scan did not have sufficient statistical power to identify them definitively (Supplementary Figure 2, Table 1).

For the TDT associations with $P < 10^{-4}$, we additionally utilized the cases that were excluded from the TDT due to missing parental data. We matched 90 independent and unrelated cases with 1,476 NIMH control samples genotyped on the Affymetrix 500K arrays [6], and performed case-control association analysis (Supplementary Table 3), combining these results with the TDT data. Promisingly, we now observed 8 SNPs (in 7 independent regions) with association at $P < 10^{-5}$ (Table 1). Of note, comparing Caucasian with non-Caucasian samples in the AGRE/ NIMH dataset, we did not observe significant heterogeneity for top results.

Our strongest associations were at chromosome 4q13 (rs17088254, $P = 8.5 \times 10^{-6}$) between *CENPC1*, a centromere autoantigen, and *EPHA5*, an ephrin receptor potentially involved in neurodevelopment; at 5p15 (rs10513025, $P = 1.7 \times 10^{-6}$) in the EST DB512398, located between *SEMA5A* and *TAS2R1*; at 6p23 (rs7766973, $P = 6.8 \times 10^{-7}$) in *JARID2,* an ortholog of the mouse jumonji gene, encoding a nuclear protein essential for embryogenesis, especially neural tube formation; at 9p24 (rs4742409, $P = 7.9 \times 10^{-6}$) between *PTPRD*, a protein tyrosine phosphatase involved in neurite outgrowth, and *JMJD2C*, a jumonji-domain containing protein involved in tri-methyl specific demethylation; at 9q21 (rs952834, $P = 7.8 \times 10^{-6}$) between *ZCCHC6*, a zinc finger and CCHC domain containing protein, and *GAS1*, growth arrest specific protein; at 10q21 (rs7923367, $P = 3.4 \times 10^{-6}$) in *CTNNA3*, alpha 3 catenin, which may be involved in the formation of stretch-resistant cell-cell adhesion complexes; and two SNPs on 11p14 (rs12293188, $P = 1.1 \times 10^{-6}$; rs16910194, $P = 3.7 \times 10^{-6}$) in *GAS2*, a caspase-3 substrate that plays a role in regulating microfilament and cell shape changes during apoptosis and can modulate cell susceptibility to p53-dependent apoptosis by inhibiting calpain activity (Table 1).

REPLICATION STUDIES

To confirm whether any of these top results might indicate true susceptibility loci, we attempted to replicate these signals, as well as others with $P < 10^{-4}$ in the initial TDT that met stringent genotyping quality criteria (Supplementary Table 3). We used several data sources to replicate the association results. First, we utilized additional autism family samples (318 trios collected by investigators of the Autism Consortium and in Montreal) with genome-wide Affymetrix 5.0/500K array data also genotyped at the Genetic Analysis Platform of the Broad Institute using the same conditions, QC, and analysis pipelines (Methods).

Second, independent Autism Genome Project (AGP) families, along with a set of Finnish families and a set of Iranian trios were used for replication of our top findings (n=1,755 trios). Two Sequenom replication pools were designed, attempting to include as many of the regions associated at $P < 10^{-4}$ as possible. The full set of SNPs considered and those successfully genotyped are shown in Supplementary Table 3, with linkage disequilibrium (r^2) noted for SNPs selected as proxies for Affymetrix markers. One of the eight SNPs with $P < 10^{-5}$ (rs10513025) that failed in this Sequenom assay was subsequently replaced in a subset of AGP

samples with a TaqMan assay. This assay showed 99.89% concordance with Affymetrix genotypes in the overlapping AGRE-NIMH samples (2,797/2,800 concordant genotypes), with manual review of the Affymetrix genotype calls also confirming the marker to be of extremely high quality (Supplementary Figure 4). In the independent replication effort, only rs10513025 was associated with $P < 0.01$ (Table 1).

Combining the scan and replication data, only rs10513025 met criteria for genome-wide significance defined by LD and permutation analyses ($P < 2.5 \times 10^{-7}$). To increase coverage of this region and fill in missing genotypes and SNPs that failed quality control, we performed imputation analysis. rs10513026 was highly (but not perfectly) correlated to the replicated chromosome 5 SNP (rs10513025) and showed even stronger association than originally observed with rs10513025 (Supplementary Figure 3). These and several other promising SNPs were directly genotyped in the original scan samples and, in fact, showed higher levels of significance (Table 2). Direct genotyping confirmed that rs10513026 showed stronger association than rs10513025 (P-value 4.5×10^{-6} vs. 9.8×10^{-6} in the re-genotyped scan trios), increasing the significance of this observation further. Several other promising results from this analysis were genotyped in a subset of scan samples, and, of note, the top SNP in imputation analysis (rs10874241, imputation $P = 9.8 \times 10^{-7}$, OR = 0.43) showed consistent results (OR = 0.4, $P = 4 \times 10^{-7}$) when directly genotyped (Supplementary Table 4).

rs10513025 and neighbors are on chromosome 5p15 in a region of LD containing several other ESTs and *TAS2R1*, a bitter taste receptor (Supplementary Figure 3). The SNPs are ~80 kb upstream of semaphorin 5A (*SEMA5A*), a gene implicated in axonal guidance and known to be down-regulated in lymphoblastoid cell lines of autism cases versus healthy controls [7]. An independent study at Children's Hospital Boston using whole blood (SWK, LK, ZK, manuscript in preparation) confirms this lower expression ($P = 0.0034$) of *SEMA5A* in autism cases versus controls. To more completely evaluate the role of this locus in autism pathogenesis, we evaluated the entirety of 5p15 for copy-number variation. Despite excellent probe coverage throughout the locus, no common or rare copy number variants were detected in the entire AGRE scan in the region of LD surrounding the associated SNPs and the entire *SEMA5A* locus including 250 kb up and downstream (see Methods).

SEMA5A EXPRESSION IN AUTISTIC BRAINS

To directly test *SEMA5A* expression in brains from autistic patients, tissue samples from 20 cases with a primary diagnosis of autism and 10 controls were obtained through the Autism Tissue Program and the Harvard Brain Bank. Samples were dissected from Brodmann area 19 of the occipital lobe cortex, a region demonstrating differences between autism cases and controls in functional imaging studies, and subjected to quantitative PCR [8]. *SEMA5A* expression, determined relative to *MAP2* (neuron specific), was significantly lower in autism brains than controls after adjustment for the age at brain acquisition, post-mortem interval, and sex ($P = 0.024$, Figure 2).

CANDIDATE AND LINKAGE REGION ASSOCIATION

We also analyzed our data for association signals at candidate genes or regions with prior evidence of involvement in autism. Although there are few well-replicated associations of biological candidate genes, there are many rare genetic variants, diseases, and syndromes associated with autism. Most of these loci have not been systematically assessed to see whether common variation in the gene or region might contribute to autism. We assessed four categories of candidate loci: 1) genes with previous evidence for association with common variation, 2) genes implicated by rare variants leading to autism, 3) genes causing Mendelian diseases associated with autism, and 4) regions where microdeletion or microduplication syndromes

are associated with autism. For each gene, we included all SNPs passing basic quality criteria within 2 kb of the transcript.

Overall, there were no compelling results in these sets (all $P > 10^{-4}$), considering the number of SNPs tested, and only two regions met criteria for region-wide (only SNPs in that gene/region considered) or set-wide (e.g. all candidate regions in the set of common variant genes considered) significance by permutation-testing (Supplementary Table 5). *MECP2* (Rett syndrome) met criteria for region-wide association ($P = 0.0071$, 5 SNPs, Supplementary Table 5). Moreover, the Williams syndrome region was borderline for set-wide significance ($P = 0.051$, Supplementary Table 5). One SNP in particular showed strong association (rs2267831, $P = 0.00012$, OR = 0.56) – as this was a rare SNP with undertransmission of the minor allele, we genotyped a subset of families and observed similar, slightly less significant distortion (OR=0.61). The SNP is located within *GTF2IRD1*, a transcription factor within the critical region for the Williams syndrome cognitive behavioral profile [9,10,11].

There appears to be little overlap between the regions of strongest linkage and association in this study. A more detailed assessment of SNP and haplotype association in the most significant linkage regions did not yield common variation that could explain the evidence for linkage (Supplementary Table 6). This is an expected outcome if linkage signals arise from rare, high penetrance variation (for which the genotyping arrays do not offer an adequate proxy) while association is sensitive to common variation with lower penetrance (that cannot be detected by linkage). For example, a 0.3% variant that increases risk by 10-fold would readily be picked up by this informative linkage scan, but would very likely not be assessed by the common SNPs on the Affymetrix 5.0 array; by contrast, the modest and protective impact of the 5% variant at the *SEMA5A* rs10513025 creates no detectable excess allele sharing among siblings but is strongly detected by association.

During review of this manuscript, another GWAS was published which identified significant association to SNPs on chromosome 5p14 [12]. While there was significant overlap between study samples, each of these scans contained a large set of unique families, so we sought to evaluate independent evidence of the top SNP (rs4307059) reported at 5p14. This SNP happens to be directly genotyped by both Affymetrix and Illumina platforms. We have a sizable number (n=796) of affected subjects with two parents genotyped (and of predominantly similar European background). However, we observed no support for association at this locus (T:U 354:335 in favor of the minor allele, a trend in the opposite direction as reported).

DISCUSSION

Autism genes have been difficult to identify, despite the high heritability of autism spectrum disorders. Up to 10% of autism cases may be due to rare sequence and gene dosage variants, for example, mutations in *NRXN1, NLGN3/4X, SHANK3,* and copy number variants at 15q11–q13 and 16p11.2. A number of diseases of known etiology, including Rett syndrome, fragile X syndrome, neurofibromatosis type I, tuberous sclerosis, Potocki-Lupski syndrome, and Smith-Lemli-Opitz syndrome are also associated with autism [1,13]. However, the remaining 90% of autism spectrum disorders, while highly familial, have unknown genetic etiology. A genome-wide linkage study using the Affymetrix 10K SNP array to genotype over 1,000 families found no genome-wide significant linkage signals, but documented suggestive linkage at 11p12–p13 and 15q23–q25 and reinforced a modest role for rare copy-number variants [14].

Many complex diseases have recently had great success with GWAS approaches, but most identified modest effects with odds ratios less than 1.3 (http://www.genome.gov/26525384). Our association analysis has excellent statistical power (>80%) to find effects of relatively common alleles (0.01–0.25 in frequency) explaining 1% of the variance in autism at the

genome-wide significant level. It is near-perfectly powered for alleles down to 1% at the replication cut-off $P < 10^{-4}$, assuming additive background genetic variance of 0.8 and shared environmental variance of 0.05 with prevalence of 0.006. One of the advantages of a family-based association test is that we avoid false positive results generated by population stratification, and in addition, we have performed careful quality control to reduce the chances of being misled by technical artifacts. However, the SNP coverage of the Affymetrix 5.0 chips is incomplete; in fact, a recent resequencing survey suggests that these arrays assay only 57% of variants with MAF > 5% at $r^2 = 0.8$ [15]. We therefore cannot exclude untested variation of large effect in autism. The linkage analysis, assuming a fully informative marker in 800 sibpairs, should detect sibling allele sharing of at least 55.125% [16].

Our linkage analysis revealed two novel regions of linkage, 6q27 (LOD = 2.94) and 20p13 (LOD = 3.81), with the latter formally exceeding the threshold for genome-wide significance. There is some overlap between the more modest signals (LOD > 2 on chr15 and chr17) and previously reported suggestive signals, but little overlap with the most promising regions of common SNP association. This suggests that the regions of the genome showing linkage may harbor rare variation, potentially with allelic heterogeneity across families, which would require re-sequencing to uncover, as has been demonstrated for the 7q35 region[17,18,19]. Interestingly, several of these regions overlap with rare syndromes or genetic events known to be strong risk factors for autism. For example, an autism case with a translocation disrupting 15q25 has been reported, while the 17p region overlaps the Smith Magenis and Potocki-Lupski Syndrome region.

The initial TDT analysis of this large multiplex autism dataset, did not reveal any associations meeting criteria for genome-wide significance, suggesting that there are not many common loci of moderate to large effect size even in a highly heritable disorder like autism. Nevertheless, replication data in our study identified a novel locus with genome-wide significant evidence for association to autism. In addition, several other SNPs in the region show similarly strong association (rs10513026, rs16883317). We ascertained a large replication sample from independent family studies with a replication at $P = 0.0061$ and meta-analysis showed this association ($P = 2.12 \times 10^{-7}$) to meet criteria for genome-wide association in our experiment. This region on chromosome 5 harbors the gene encoding the bitter taste receptor, *TAS2R1*, and several uncharacterized ESTs and is adjacent to *SEMA5A*, a member of the semaphorin axonal guidance protein family, which has shown down-regulated expression in transformed B lymphocytes from autism samples[15]. We have further extended this finding by directly demonstrating lowered *SEMA5A* gene expression in autism brain tissue. This is an attractive candidate gene given that its protein is a bifunctional guidance molecule, which is both attractive and inhibitory for developing neurons. Interestingly, the SEMA5A receptor is plexin B3, which also signals through the tyrosine kinase MET, a previously reported autism susceptibility gene[20,21].

Finally, we investigated whether different classes of genes or regions -- loci previously implicated by functional or positional candidate gene association studies, rare variants implicated in autism, Mendelian disorder genes with association to autism, or regions of copy number variation associated with autism -- showed association with common alleles included in our marker set. Although there were several nominally significant associations, only the Williams syndrome region (one SNP in *GTF2IRD1*) was borderline statistically significant ($P = 0.051$), after correcting for the microdeletion/duplication syndrome regions tested. In the category of Mendelian disorders associated with autism, *MECP2*, the gene for Rett syndrome, showed region-wise statistical significance. These results raise the possibility that Rett and Williams syndrome genes may contribute more generally to autism spectrum disorders. Although the genes in which common variation has been reported to be associated with autism do not show evidence for association, this cannot be interpreted as failure to replicate previous

results in all cases, because much of the variation reported as associated is not captured on the Affymetrix platform (e.g. length polymorphisms, microsatellites, untagged SNPs such as the promoter variant at *MET*[21]). Instead, despite a high density of markers, our results suggest that we did not identify additional common variation with evidence for association. Overall however, our results imply that these postulated candidate regions, mostly based on rare events known to cause autism, are not among the regions with common alleles having the strongest risk effects for autism.

Interestingly, both our linkage and association analyses, from the primary and replication analyses, suggest that low frequency (<0.05) minor alleles may be common in autism. Intriguingly, the linkage studies reveal low frequency susceptibility alleles whereas the association analyses have uncovered rare alleles with odds ratios less than 0.6 (the common alleles in the population associated with increased risk for autism). This can occur when the ancestral allele, that was previously neutral or beneficial, now has detrimental effects revealed by an evolutionarily recent environment, or when a pleiotropic function of the allele is selectively advantageous, or when this variation is hitchhiking on a shared haplotype with a distinct beneficial allele [22]. However, it is worth noting that our study design of ascertaining multiplex families is not well-powered to identify loci under this genetic model of common major alleles associated with autism susceptibility.

In summary, we report genome-wide significant linkage as well as an association of common genetic variation with autism. Our results will require follow-up to identify the functional variation in the linkage and association regions that we report here and to probe the functions of the relatively unstudied transcripts implicated. These results could provide completely novel insight into the biology and pathogenesis of a common neurodevelopmental disorder.

METHODS SUMMARY

SAMPLES AND GENOTYPING

Our primary samples are from the AGRE and NIMH Repositories. Replication with Affymetrix technology included NIMH controls, families collected by members of the Autism Consortium, and families ascertained from Montreal. Replication with Sequenom technology included the Autism Genome Project, Finnish, and Iranian subsets of Autism Consortium investigator-collected families. Details of the ascertainment for each sample collection, genotyping, and quality control processes can be found in Methods.

LINKAGE AND ASSOCIATION ANALYSIS

The linkage analysis was conducted with a pruned autosomal SNP set [see Methods for details of marker selection] and chromosome X set (670 SNPs) using the cluster option in MERLIN/MINX ($r^2 < 0.1$) [23], yielding 16,581 independent markers. We performed confirmatory analysis on non-overlapping datasets by selecting alternate SNPs.

Association analysis was performed in PLINK[24]. The basic association test was a transmission disequilibrium test (TDT), and the extra cases vs. controls analysis was performed by allelic association, after excluding cases that were not well-matched to the controls, based on multi-dimensional scaling ($\lambda < 1.1$). Combining the TDT and case-control tests was performed using expected and observed allele counts by the formula: $Z_{meta} = (\Sigma EXP - \Sigma OBS)/\sqrt{\Sigma VAR}$. Meta-analysis of AGRE/NIMH and replication data was performed using the statistic $(Z_{AGRE/NIMH} + Z_{replication})/\sqrt{2}$. Gene-set analysis was performed in PLINK using the set-based TDT. Imputation-based association was performed in PLINK with the proxy-tdt command, using the HapMap CEU parent samples as the reference panel and information score >0.8. Haplotype analysis in the linkage regions was performed using 5-SNP sliding windows, as

implemented in PLINK hap-tdt. See Methods section for details of determination of genome-wide significance thresholds.

METHODS

PRIMARY STUDY SAMPLES

All samples used in this study arose from investigations approved by the individual and respective Institutional Review Boards in the USA and at international sites where relevant. Informed consent was obtained for all adult study participants; for children under age 18, both the consent of the parents or guardians and the assent of the child were obtained.

AGRE samples—The Autism Genetic Resource Exchange (AGRE) curates a collection of DNA and phenotypic data from multiplex families with autism spectrum disorder (ASD) available for genetic research [3]. We genotyped individuals from 801 families, selecting those with at least one child meeting criteria for autism by the Autism Diagnostic Interview-Revised (ADI-R) [25], 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 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"). We excluded probands with widely discrepant classifications of affection status via the ADI-R and ADOS that could not be reconciled. We also 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). 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.

NIMH samples—The NIMH Autism Genetics Initiative maintains a collection of DNA from multiplex and simplex families with ASD. We genotyped individuals from 341 nuclear families, 258 of which were independent of the AGRE dataset, with at least one child meeting criteria for autism by the ADI-R, and a second child considered affected using the same criteria as described for the AGRE dataset above. Similar exclusion criteria were used, including known chromosomal abnormalities and excess non-Mendelian inheritance. The self-reported race/ethnicity of these samples is 83% white, 4% Hispanic, 2% unknown, 7% mixed, 3% Asian, and 1% African American.

Merged dataset for primary screening—We utilized the Birdseed algorithm for genotype calling at both genotyping centers[26,27]. As 324 individuals were genotyped at both centers, we performed a concordance check. One sample showed substantial differences between the two centers, but no excess of Mendelian errors, indicating that a sample mix-up occurred in which each center genotyped a different sibling that was identified as the same sample. Excluding this sample, overall genotype concordance between the two centers was 99.72%.

Before merging data, we examined the distribution of chi-square values and used a series of quality control (QC) filters designed to identify a robust set of SNPs. We discovered that filtering AGRE genotypes to 98% completeness and less than 10 MEs was sufficient to remove SNPs that artificially inflated the chi-square distribution for SNPs with MAF (minor allele frequency) > 0.05. For MAF < 0.05, we observed much greater inflation ($\lambda = 1.17$), due entirely

to a strong excess of SNPs with under-transmission of the minor allele (OR < 1). While the same filters yielded high-quality results for SNPs with over-transmission of the minor allele ($\lambda = 1.04$), we found that much stricter filtering was required for rarer SNPs with OR<1 (missing data < .005). This is not unexpected based on a well-documented bias in the TDT: if missing data are preferentially biased against heterozygotes or rare homozygotes, significant, artificial over-transmission of the common allele is expected [28,29]. To achieve comparable quality for the NIMH dataset, we filtered on 96% completeness and fewer than 4 MEs. Our final QQ plot for the combined dataset is shown in Supplementary Figure 1 and has a $\lambda \sim 1.03$, less than that observed in the Wellcome Trust Case Control Consortium paper for five of the seven phenotypes studied [30]. The combined data set, consisting of 1,031 families (856 with two parents) and a total of 1,553 affected offspring, was employed for association testing.

For linkage analyses, the combined AGRE/NIMH dataset was further merged with Illumina 550K genotype data generated at the Children's Hospital of Philadelphia (CHOP) and available from AGRE, adding ~300 nuclear families (1,499 samples). We used the extensive overlap of samples between the AGRE/NIMH and the CHOP datasets (2,282 samples) to select an extremely high quality set of SNPs for linkage analysis. Specifically, we required SNPs to be on both the Affymetrix 500K/5.0 and Illumina 550K platforms, with >99.5% concordance across platforms. We further restricted SNPs to MAF > 0.2, < 1% missing data, Hardy Weinberg P> 0.01, and no more than 1 ME. This left ~36,000 SNPs of outstanding quality. For autosomal SNPs, we further pruned using PLINK to remove SNPs with $r^2 > 0.1$, yielding 16,311 SNPs.

REPLICATION SAMPLES

NIMH control samples—Controls obtained from the NIMH Genetics Repository were genotyped on the Affymetrix 500K platform at the Broad Institute Genetic Analysis Platform for another study [6]. Of these, 1,494 matched well with our sample, and were used as controls to compare with the cases and parents in our study.

Montreal samples—Subjects diagnosed with autism spectrum disorders with both of their parents were recruited from clinics specializing in the diagnosis of Pervasive Developmental Disorders (PDD), readaptation centers, and specialized schools in the Montreal and Quebec City regions, Canada, as described [31]. Subjects with ASD were diagnosed by child psychiatrists and psychologists expert in the evaluation of ASD. Evaluation based on the Diagnostic and Statistical Manual of Mental Disorders (DSM) criteria included the use of the Autism Diagnostic Interview-Revised (ADI-R)[25] and the Autism Diagnostic Observation Schedule (ADOS)[32]. As an additional screening tool for the diagnosis of ASD, the Autism Screening Questionnaire, which is derived from the ADI-R, was completed [33]. Furthermore, all proband medical charts were reviewed by a child psychiatrist expert in PDD to confirm their diagnosis and exclude subjects with any co-morbid disorders. Exclusion criteria were: (1) an estimated mental age < 18 months, (2) a diagnosis of Rett syndrome or Childhood Disintegrative Disorder and (3) evidence of any psychiatric and neurological conditions including: birth anoxia, rubella during pregnancy, fragile X syndrome, encephalitis, phenylketonuria, tuberous sclerosis, Tourette and West syndromes. Subjects with these conditions were excluded based on parental interview and chart review. However, participants with a co-occurring diagnosis of semantic-pragmatic disorder (due to its large overlap with PDD), attention deficit hyperactivity disorder (seen in a large number of patients with ASD during development), and idiopathic epilepsy (related to the core syndrome of ASD) were eligible for the study.

Santangelo EDSP family samples—Families were ascertained for having one or more autistic children and at least one non-autistic child aged 16 or older for an extremely discordant sib-pair linkage study. Recruitment took place in Massachusetts and surrounding states through

contacts with parent support and patient advocacy groups, brochures, newsletters, and the study web site. Parents were interviewed about their children, and non-autistic children were interviewed about themselves. An informant/caregiver, usually the proband's mother, was interviewed using the Autism Diagnostic Instrument-Revised (ADI-R) to confirm the diagnosis of autism at age 4–5 years[25,34]. Families were included if the affected children met Diagnostic and Statistical Manual of Mental Disorders-IV (DSM-IV) criteria for autistic disorder and their non-autistic siblings (aged 16 and older) did not display any of the broader autism phenotype traits, which were assessed with the (M-PAS-R), the Pragmatic Language Scale (PLS), and the Friendship Interview [35,36]. Probands were excluded if they had medical conditions associated with autism such as fragile X syndrome or gross CNS injury, or if they were under four years of age, due to the possible uncertainty in diagnosis at younger ages. Twenty-nine families met eligibility criteria for the study and comprised the final sample for analysis.

High Functioning Autism family samples—Families were included if their affected child had been previously diagnosed with Autism or Asperger syndrome, had a level of intellectual functioning above the range of mental retardation (i.e., Full Scale, Verbal, and Performance IQ > 70), chronological age between 6 and 21 years, and an absence of significant medical or neurological disorders (including fragile X syndrome and tuberous sclerosis). Families were ascertained and recruited through the Acute Residential Treatment (ART) programs and outpatient child and adolescent services at McLean Hospital, as well as through associated hospitals and clinics. Brochures and a website were also utilized. Thirty-three families (133 participants) were enrolled in the study. Participation was voluntary.

MGH-Finnish collaborative samples—Altogether 58 individuals with a diagnosis of High Functioning Autism (HFA) or Asperger's Syndrome (AS) were recruited in Finland. Fifty-two children and adolescents aged 8 to 15 years were identified from patient-records at the Oulu University Hospital in 2003. These children and adolescents have been evaluated for HFA/AS at the Oulu University Hospital. In addition, six children (3 boys, 3 girls) 11 years of age were recruited from an epidemiological study conducted in 2001 [37].

All participants had full scale IQ scores greater than or equal to 80 measured with the Wechsler Intelligence Scale for Children—Third Revision [38]. Furthermore, none of the children subjects were diagnosed with other developmental disorders (e.g., dysphasia, fragile X syndrome). Clinical diagnoses of HFA/AS were confirmed by administering the Autism Diagnostic Interview-Revised [25] and the Autism Diagnostic Observation Schedule [32]. Of the 58 participants with HFA/AS, 35 met the diagnostic criteria for AS and 21 met the diagnostic criteria for HFA according to ICD-10 diagnostic criteria [39]. Two participants met diagnostic criteria for PDD-NOS; these participants were excluded due to their manifesting different and less severe symptoms than our sample of children with HFA or AS.

Children's Hospital Boston samples—Probands with a documented history of clinical diagnosis of ASD were recruited at Children's Hospital Boston. To participate, they had to be over 24 months of age and have at least one biological parent or an affected sibling available. Subjects were excluded if they had an underlying metabolic disorder or any chronic systemic disease, an acquired developmental disability (e.g. birth asphyxia, trauma-related injury, meningitis, etc.), or cerebral palsy. All participants provided informed consent and a phenotyping battery was performed including the Autism Diagnostic Observation Schedule (ADOS), the Autism Diagnostic Interview- Revised (ADI-R) and other measures to assess cognitive status. 75% of subjects with a clinical diagnosis met strict research criteria for ASD on both ADI-R and ADOS. In addition, a complete family and medical history was obtained.

Homozygosity Mapping Collaborative for Autism (HMCA) samples—Families with cousin marriages and children affected by autism spectrum disorder (ASD) with or without

mental retardation (MR) were recruited by multiple collaborators in the HMCA. The patients from Istanbul were evaluated by a child psychiatrist (Nahit M. Mukaddes) trained in the Autism Diagnostic Observation Schedule (ADOS) and Autism Diagnostic Interview - Revised (ADI-R), and who made diagnoses according to DSM-IV-TR criteria and the Childhood Autism Rating Scale (CARS). Patients from Kuwait were enrolled from the Kuwait Centre for Autism by Samira Al-Saad. In Jeddah, Saudi Arabia, patients were evaluated by both a developmental pediatrician (Soher Balkhy) and a pediatric neurologist (Generoso Gascon) and diagnoses were based on DSM-IV-TR criteria. In Lahore, Pakistan, a neurologist (Asif Hashmi) with training in the ADOS and ADI-R diagnosed patients using DSM-IV-TR criteria. In most settings, patients were enrolled from tertiary clinical centers and these patients had standard of care neuromedical assessments, including physical examination, medical and neurological history, fragile X testing, and other genetic and metabolic testing when indicated. MRI was obtained for patients in whom a brain malformation was suspected or seizures were present. In addition, IQ scores (usually from the Stanford-Binet) and adaptive behavior measures were obtained from the patients' existing medical records. Secondary assessments were conducted on the most informative pedigrees by the Boston clinical team in collaboration with local multi-disciplinary teams. Clinical members of the Boston team included: developmental psychologists (Janice Ware, Elaine LeClaire, Robert M. Joseph), pediatric neurologists (Ganesh H. Mochida, Anna Poduri), a clinical geneticist (Wen-Han Tan), and a neuropsychiatrist (Eric M. Morrow). The secondary assessment battery was designed to obtain a comprehensive description of current and historical autism symptomatology, cognitive and adaptive functioning, and neurological and physical morphological status in the patient and pedigree. The secondary assessment included: neurologic examination; genetic dysmorphology examination; the CARS; the Social Communication Questionnaire (SCQ) administered with probing on par with the ADI-R by ADI-R reliable examiners; the ADOS (usually Module 1); the Vineland Adaptive Behavior Scales, Second Edition (VABS-II); Kaufman Brief Intelligence Test, Second Edition (KBIT-II). ADOS assessments were videotaped and dysmorphology findings were photographed for archival purposes.

AGP samples—Individuals typically received at least two of three evaluations for autism symptoms: ADI-R, ADOS and clinical evaluation. Of the 1,679 affected individuals from 1,443 families, 966 met criterion for autism on the ADI-R and ADOS and most of these also had a clinical evaluation of autism; 160 affected individuals met criteria for autism on one of the two diagnostic instruments (ADI-R, ADOS) but were missing information on the other instrument; and, 553 individuals met criteria for spectrum disorder on one or both instruments. Affected individuals were recruited from both simplex and multiplex families, 71% of this sample being from multiplex families. The majority of the families were of European ancestry (83%).

Finnish autism family samples—Families were recruited through university and central hospitals. Detailed clinical and medical examinations were performed by experienced child neurologists as described elsewhere[40]. Diagnoses were based on ICD-10 [39] and DSM-IV [41] diagnostic nomenclatures. Families with known associated medical conditions or chromosomal abnormalities were excluded from the study. A total of 106 families included 400 individuals for whom genotype data was available. Of these, 111 had a diagnosis of infantile autism and 13 a diagnosis of Asperger syndrome. All families were Finnish, except for one family where the father was Turkish.

Iranian Trio samples—Eligible participants in this study were Iranian families with at least one child affected with ASD, including cases of autistic disorder, Asperger syndrome and pervasive developmental disorder-not otherwise specified (PDD-NOS). Eighty families (282 individuals) from Iran were ascertained and assessed. This sample was ascertained by screening and diagnostic testing of over 90,000 preschool children from Tehran in 2004. Diagnoses of

children were made according to DSM-IV criteria via the ADI-R and the ADOS. Patients with abnormal karyotypes and dysmorphic features were excluded. Most of the families were father-mother-child trios but some had more than one affected child. All affected biological siblings were assessed with the same diagnostic tools. We have ascertained and assessed 80 families (282 individuals) from Iran.

AFFYMETRIX GENOTYPING

The AGRE samples were genotyped on Affymetrix 5.0 chips at the Genetic Analysis Platform of the Broad Institute, using standard protocols. The 5.0 chip was designed to genotype nearly 500,000 SNPs across the genome in order to enable genome-wide association studies [26,27]. The NIMH controls were genotyped at the Broad Institute using the Affymetrix 500K Sty and Nsp chips, using a similar protocol [6]. The Autism Consortium and Montreal replication samples were also genotyped at the Broad Institute under the same conditions. The NIMH autism samples were genotyped at the Johns Hopkins Center for Complex Disease on the Affymetrix 500K (Nsp and Sty) and 5.0 platforms using similar standard protocols.

Genotype calling for the 5.0 arrays was performed by Birdseed [26,27] and for the 500K arrays was performed by BRLMM. As basic QC filters for the data generated at the Broad Institute, we required that genotyping was >95% complete for each individual, and that each family had fewer than 10,000 Mendelian inheritance errors across the genome. We also required that each SNP had >95% genotyping, fewer than 15 Mendelian errors, Hardy-Weinberg Equilibrium $P > 10^{-10}$, and minor allele frequency greater than 1%. For the AGRE sample, this left 2,883 high quality individuals genotyped for 399,147 SNPs with 99.6% average call rate. The basic filters for the data generated at Johns Hopkins were individual call rates > 95% for 5.0 arrays and > 90% for 500K arrays data, fewer than 5,000 Mendelian errors per family. Only monomorphic SNPs and those with greater than 50% missing data were dropped, for 498,216 SNPs. Our combined dataset had nearly 365,000 SNPs passing QC.

SEQUENOM GENOTYPING

SNPs were assayed using Sequenom technology for the AGP samples at three centers, namely Gulbenkian, Mt. Sinai, and Oxford: DNA from 1,629 families representing numerous recruiting sites was genotyped for 54 SNPs. SNPs with >3% missing data, namely rs4690464, rs10513025, and rs17088296, were excluded from analysis. The next step in our quality control process was to remove families with ≥4 Mendelian errors, out of 51 remaining loci, under the assumption that this indicated pedigree errors. Data from 110 families were removed due to Mendelian errors. Thereafter, SNPs were removed if they showed excessive Mendelian errors (>16) in the remaining families. Using this criterion, two more SNPs, rs155437 and rs1925058, were removed from analysis. It was apparent that DNA quality varied by study site and could be responsible for concomitant genotype quality differences. Therefore, we also evaluated rate of missing genotypes per locus and study site. Our analyses showed that DNA from a few population samples showed excess missingness for two SNPs, rs4742408 and rs7869239, relative to the remaining population samples. Specifically three population samples showed more than 7% missing genotypes for rs4742408 and rs7869239 whereas the remaining population samples had about 1% or less missing genotypes. Therefore, for these loci we deleted genotypes only from the samples showing excess missingness. As a final quality control step, we then evaluated missing genotypes for the remaining loci. If more than five loci were missing genotypes, the individual's data was removed from analysis. By this criterion 76 additional families became uninformative for family-based association analysis, leaving 1,443 families for association analysis. The Finnish autism samples were genotyped in the Peltonen lab, and the Iranian trios were genotyped at the Broad Institute using very similar protocols. All samples were genotyped using aliquots from the same pooled primers and probes.

COPY NUMBER ANALYSIS

Because of previous reports of two large (>1 Mb), independent *de novo* deletions spanning this locus [42], we assessed the region surrounding rs10513025 and the entire *SEMA5A* locus for copy number variation that could either explain or provide independent evidence of the importance of this region to autism using Birdsuite [26] to analyze all Affymetrix 5.0 samples. Birdsuite genotypes previously annotated common copy number polymorphisms [27] and in parallel searches for novel copy number variants using an HMM. Probe coverage in the region was good, with no 50kb window having fewer than 10 probes and an average spacing between probes of 2.5 kb, allowing very good sensitivity for CNVs greater than 25kb. We found no deletions or duplications near this SNP, nor any overlapping the gene *SEMA5A*. The closest copy number variants upstream and downstream of this SNP appeared to be a rare (~2–3% frequency, previously annotated CNP) 40kb deletion from 288 kb from the 3′ end of *SEMA5A*, and a rare (~1% frequency, novel) 20kb deletion 356 kb upstream of the 5′ end of *SEMA5A*. Each of these appeared to be segregating polymorphisms, but fall far outside of the boundaries of *SEMA5A* and *TAS2R1* and far beyond the linkage disequilibrium block containing rs10513025.

EXPRESSION ANALYSIS

Fresh-frozen brain tissue samples dissected from the cortex (Brodmann area 19) were obtained through the Autism Tissue Program (www.atpportal.org) from the Harvard Brain Bank and the NICHD Brain and Tissue Bank at the University of Maryland from 20 samples with a primary diagnosis of autism, and 10 controls. Total RNA was extracted using TRIzol reagent (Invitrogen, Carlsbad, CA) according to the manufacturer's protocol. Complementary DNA (cDNA) was generated from 8μg of total RNA using the Superscript III First-Strand Synthesis kit (Invitrogen). cDNA was diluted 1:5 in 10mM Tris and 1μL of diluted cDNA was used per 10μL PCR reaction. Quantitative real-time PCR was performed on a Lightcycler 480 (Roche Applied Science, Indianapolis, IN) using 2X Taqman Gene Expression Master Mix and probes obtained from Applied Biosystems (ABI, Foster City, CA): *SEMA5A* (Hs01549381_m1), *MAP2* (Hs01103234_g1), *TBP* (Hs00920497_m1), *GAPDH* (4333764F). For multiplex reactions, 0.5μL FAM-labeled *SEMA5A* probe and 0.5μL VIC-labeled *MAP2* probe were used per 10μL reaction. The amount of *SEMA5A* relative to *MAP2* was determined for each case using the $\Delta\Delta C_t$ method [43]. Comparison of *SEMA5A* to *TBP* and *GAPDH* yielded similar results. Logistic regression was performed on autism status, adjusting for age at death, post-mortem interval, sex, and *SEMA5A* expression, with a 1-sided *P*-value reported for the association of lower *SEMA5A* expression with autism status.

DETERMINATION OF SIGNIFICANCE

To determine an appropriate experimental threshold for genome-wide significance, permutation was performed on this dataset by gene-dropping, and genome-wide significance was estimated by taking the lowest *P*-value from each of 1000 permuted datasets and using the 50th as a threshold for *P* < 0.05 experiment-wide significance ($P < 2.5 \times 10^{-7}$). To calculate an estimate of the effective number of tests (T_{eff}), we used the following algorithm:

1. Start with the most 5′ SNP on a chromosome ($SNP_{i,j}$), where i=chromosome, and j=SNP position, and calculate pairwise LD with all downstream SNPs within 1 Mb ($r^2[SNP_{1,1} \times SNP_{1,n}]$)

2. For $SNP_{1,1}$, $T_{eff(1,1)} = 1 - \max(r^2[SNP_{1,1} \times SNP_{1,n}])$

3.

For chromosome i, $\quad T_{eff(i)} = \sum_{j=1}^{m} T_{eff(i,j)}$, where m=the total number of SNPs on a chromosome

4.

$$T_{eff} = \sum_{i=1}^{23} T_{eff(i)}$$

Since this algorithm only accounts for pair-wise LD, it provides a conservative estimate of the number of effective tests.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Abrahams BS, Geschwind DH. Advances in autism genetics: on the threshold of a new neurobiology. Nat Rev Genet 2008;9:341–355. [PubMed: 18414403]

2. Weiss LA, et al. Association between microdeletion and microduplication at 16p11.2 and autism. N Engl J Med 2008;358:667–675. [PubMed: 18184952]

3. Geschwind DH, et al. The autism genetic resource exchange: a resource for the study of autism and related neuropsychiatric conditions. Am J Hum Genet 2001;69:463–466. [PubMed: 11452364]

4. Abecasis GR, Wigginton JE. Handling marker-marker linkage disequilibrium: pedigree analysis with clustered markers. Am J Hum Genet 2005;77:754–767. [PubMed: 16252236]

5. Lander E, Kruglyak L. Genetic dissection of complex traits: Guidelines for interpreting and reporting linkage results. Nature Genetics 1995;11:241–247. [PubMed: 7581446]

6. Sklar P, et al. Whole-genome association study of bipolar disorder. Mol Psychiatry 2008;13:558–569. [PubMed: 18317468]

7. Melin M, et al. Constitutional downregulation of SEMA5A expression in autism. Neuropsychobiology 2006;54:64–69. [PubMed: 17028446]

8. Gilbert SJ, Bird G, Brindley R, Frith CD, Burgess PW. Atypical recruitment of medial prefrontal cortex in autism spectrum disorders: an fMRI study of two executive function tasks. Neuropsychologia 2008;46:2281–2291. [PubMed: 18485420]

9. Hirota H, et al. Williams syndrome deficits in visual spatial processing linked to GTF2IRD1 and GTF2I on chromosome 7q11.23. Genet Med 2003;5:311–321. [PubMed: 12865760]

10. Edelmann L, et al. An atypical deletion of the Williams-Beuren syndrome interval implicates genes associated with defective visuospatial processing and autism. J Med Genet 2007;44:136–143. [PubMed: 16971481]

11. van Hagen JM, et al. Contribution of CYLN2 and GTF2IRD1 to neurological and cognitive symptoms in Williams Syndrome. Neurobiol Dis 2007;26:112–124. [PubMed: 17270452]

12. Wang K, et al. Common genetic variants on 5p14.1 associate with autism spectrum disorders. Nature 2009;459:528–533. [PubMed: 19404256]

13. Zafeiriou DI, Ververi A, Vargiami E. Childhood autism and associated comorbidities. Brain Dev 2007;29:257–272. [PubMed: 17084999]

14. Szatmari P, et al. Mapping autism risk loci using genetic linkage and chromosomal rearrangements. Nat Genet 2007;39:319–328. [PubMed: 17322880]

15. Bhangale TR, Rieder MJ, Nickerson DA. Estimating coverage and power for genetic association studies using near-complete variation data. Nat Genet 2008;40:841–843. [PubMed: 18568023]

16. Risch NJ. Searching for genetic determinants in the new millennium. Nature 2000;405:847–856. [PubMed: 10866211]

17. Arking DE, et al. A common genetic variant in the neurexin superfamily member CNTNAP2 increases familial risk of autism. Am J Hum Genet 2008;82:160–164. [PubMed: 18179894]

18. Alarcon M, et al. Linkage, association, and gene-expression analyses identify CNTNAP2 as an autism-susceptibility gene. Am J Hum Genet 2008;82:150–159. [PubMed: 18179893]

19. Bakkaloglu B, et al. Molecular cytogenetic analysis and resequencing of contactin associated protein-like 2 in autism spectrum disorders. Am J Hum Genet 2008;82:165–173. [PubMed: 18179895]

20. Campbell DB, et al. Disruption of cerebral cortex MET signaling in autism spectrum disorder. Ann Neurol 2007;62:243–250. [PubMed: 17696172]

21. Campbell DB, et al. A genetic variant that disrupts MET transcription is associated with autism. Proc Natl Acad Sci U S A 2006;103:16834–16839. [PubMed: 17053076]

22. Di Rienzo A. Population genetics models of common diseases. Curr Opin Genet Dev 2006;16:630–636. [PubMed: 17055247]

23. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet 2002;30:97–101. [PubMed: 11731797]

24. Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559–575. [PubMed: 17701901]

25. Lord C, Rutter M, Le Couteur A. Autism Diagnostic Interview - Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. Journal of Autism and Developmental Disorders 1994;24:659–685. [PubMed: 7814313]

26. Korn JM, et al. Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. Nat Genet 2008;40:1253–1260. [PubMed: 18776909]

27. McCarroll SA, et al. Integrated detection and population-genetic analysis of SNPs and copy number variation. Nat Genet 2008;40:1166–1174. [PubMed: 18776908]

28. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 2005;6:95–108. [PubMed: 15716906]

29. Mitchell AA, Cutler DJ, Chakravarti A. Undetected genotyping errors cause apparent overtransmission of common alleles in the transmission/disequilibrium test. Am J Hum Genet 2003;72:598–610. [PubMed: 12587097]

30. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447:661–678. [PubMed: 17554300]

31. Gauthier J, et al. Autism spectrum disorders associated with X chromosome markers in French-Canadian males. Mol Psychiatry 2006;11:206–213. [PubMed: 16261168]

32. Lord C, et al. The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. Journal of Autism and Developmental Disorders 2000;30:205–223. [PubMed: 11055457]

33. Berument SK, Rutter M, Lord C, Pickles A, Bailey A. Autism screening questionnaire: diagnostic validity. Br J Psychiatry 1999;175:444–451. [PubMed: 10789276]

34. Le Couteur A, et al. Autism Diagnostic Interview: A standardized investigator-based instrument. J Autism Develop Disord 1989;19:363–388.

35. Tyrer, PJ. Personality disorders: diagnosis, management, and course. Wright; 1988.

36. Landa R, et al. Social language use in parents of autistic individuals. Psychol Med 1992;22:245–254. [PubMed: 1574562]

37. Mattila ML, et al. An epidemiological and diagnostic study of Asperger syndrome according to four sets of diagnostic criteria. J Am Acad Child Adolesc Psychiatry 2007;46:636–646. [PubMed: 17450055]

38. Wechsler, D. Wechsler Intelligence Scale for Children. Vol. 3. The Psychological Corporation; 1991.

39. World Health Organization. The ICD-10 Classification of Mental and Behavioural Disorders. Diagnostic Criteria for Research. WHO; 1993.

40. Auranen M, et al. A genomewide screen for autism-spectrum disorders: evidence for a major susceptibility locus on chromosome 3q25–27. Am J Hum Genet 2002;71:777–790. [PubMed: 12192642]

41. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (4th edn) (DSM-IV). Vol. 4. APA; 1994.

42. Marshall CR, et al. Structural variation of chromosomes in autism spectrum disorder. Am J Hum Genet 2008;82:477–488. [PubMed: 18252227]

43. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(−Delta Delta C(T)) Method. Methods 2001;25:402–408. [PubMed: 11846609]

Author List

Writing Group: Lauren A. Weiss, Dan E. Arking, Mark J. Daly, Aravinda Chakravarti

JHU – NIMH Genome Scan Team: Dan E. Arking[1], Camille W. Brune[2], Kristen West[1], Ashley O'Connor[1], Gina Hilton[1], Rebecca L. Tomlinson[76], Andrew B. West[76], Edwin H. Cook Jr.[2], Aravinda Chakravarti[1]

Autism Consortium - AGRE Genome Scan Team: Lauren A. Weiss[3,4,a], Todd Green[3,4], Shun-Chiao Chang[3], Stacey Gabriel[4], Casey Gates[4], Ellen Hanson[5], Andrew Kirby[3,4], Joshua Korn[3,4], Finny Kuruvilla[3,4], Steven McCarroll[3,4], Eric Morrow[3,4,6,b], Benjamin Neale[3,4], Shaun Purcell[3,4], Roksana Sasanfar[3,7], Carrie Sougnez[4], Christine Stevens[4], David Altshuler[3,4], James Gusella[3,4], Susan L. Santangelo[3], Pamela Sklar[3,4], Rudolph Tanzi[3], Mark J. Daly[3,4]

Replication Teams

Autism Genome Project Consortium (listed alphabetically): Richard Anney[28], Anthony J. Bailey[8], Gillian Baird[9], Agatino Battaglia[56], Tom Berney[11], Catalina Betancur[12], Sven Bölte[13], Patrick F. Bolton[14], Jessica Brian[15], Susan E. Bryson[16], Joseph D. Buxbaum[17], Ines Cabrito[51], Guiqing Cai[17], Rita M. Cantor[18], Edwin H. Cook Jr.[2], Hilary Coon[19], Judith Conroy[25], Catarina Correia[51], Christina Corsello[20], Emily L. Crawford[46], Michael L. Cuccaro[21], Geraldine Dawson[60], Maretha de Jonge[23], Bernie Devlin[24], Eftichia Duketis[13], Sean Ennis[25], Annette Estes[22], Penny Farrar[38], Eric Fombonne[26], Christine M. Freitag[13], Louise Gallagher[28], Daniel H. Geschwind[29], John Gilbert[21], Michael Gill[28], Christopher Gillberg[53], Jeremy Goldberg[30], Andrew Green[25], Jonathan Green[31], Stephen J. Guter[2], Jonathan L. Haines[32], Joachim Hallmayer[33], Vanessa Hus[20], Sabine M. Klauck[34], Olena Korvatska[55], Janine A. Lamb[35], Magdalena Laskawiec[8], Marion Leboyer[54], Ann Le Couteur[11], Bennett L. Leventhal[2], Xiao-Qing Liu[15, 44], Catherine Lord[20], Linda Lotspeich[33], Elena Maestrini[58], Tiago Magalhaes[51], William Mahoney[36], Carine Mantoulan[37], Helen McConachie[11], Christopher J McDougle[57], William M. McMahon[19], Christian R. Marshall[44], Judith Miller[19], Nancy J. Minshew[2], Anthony P. Monaco[38], Jeff Munson[22], John I Nurnberger Jr[57], Guiomar Oliveira[52], Alistair Pagnamenta[38], Katerina Papanikolaou[39], Jeremy R. Parr[8], Andrew D. Paterson[15, 44], Margaret A Pericak-Vance[21], Andrew Pickles[40], Dalila Pinto[44], Joseph Piven[41], David J Posey[57], Annemarie Poustka[34,x], Fritz Poustka[13], Regina Regan[25], Jennifer Reichert[17], Katy Renshaw[8], Wendy Roberts[15], Bernadette Roge[37], Michael L. Rutter[42], Jeff Salt[2], Gerard D. Schellenberg[59], Stephen W. Scherer[44], James S. Sutcliffe[46], Peter Szatmari[30], Katherine Tansey[28], Ann P. Thompson[30], John Tsiantis[39], Herman Van Engeland[23], Astrid M. Vicente[51], Veronica J. Vieland[10], Fred Volkmar[47], Simon Wallace[8], Thomas H. Wassink[48], Ellen M. Wijsman[49], Kirsty Wing[38], Kerstin Wittemeyer[37], Brian L. Yaspan[46], Lonnie Zwaigenbaum[50]

The Homozygosity Mapping Collaborative for Autism (HMCA): Eric M. Morrow[3,4,6,61,b], Seung-Yun Yoo[4,6,61], Robert Sean Hill[4,6,61], Nahit M. Mukaddes[62], Soher Balkhy[63], Generoso Gascon[63,64], Samira Al-Saad[66], Asif Hashmi[65], Janice Ware[5], Robert M. Joseph[67], Elaine LeClair[5], Jennifer N. Partlow[6,61], Brenda Barry[6,61], and Christopher A. Walsh[4,6,61]

MGH-Oulu Study: David Pauls[3], Irma Moilanen[68], Hanna Ebeling[68], Marja-Leena Mattila[68], Sanna Kuusikko[68], Katja Jussila[68], Jaakko Ignatius[68]

MGH-Iran Study: Roksana Sasanfar[3,7], Ala Tolouei[7], Majid Ghadami[7], Maryam Rostami[69], Azam Hosseinipour[7], and Maryam Valujerdi[7]

MGH-EDSP Study: Susan L. Santangelo[3], Kara Andresen[3,70], Brian Winkloski[3], Stephen Haddad[3]

Children's Hospital Boston: Lou Kunkel[6], Zak Kohane[6], Tram Tran[6], Sek Won Kong[6], Stephanie Brewster O'Neil[6], Ellen M. Hanson[5], Rachel Hundley[5], Ingrid Holm[6], Heather Peters[6], Elizabeth Baroni[6], Aislyn Cangialose[6], Lindsay Jackson[6], Lisa Albers[6], Ronald Becker[6], Carolyn Bridgemohan[6], Sandra Friedman[6], Kerim Munir[6], Ramzi Nazir[6], Judith Palfrey[6], Alison Schonwald[6], Esau Simmons[6], Leonard A. Rappaport[5]

Montreal: Julie Gauthier[71], Laurent Mottron[72], Ridha Joober[26], Eric Fombonne[26], Guy Rouleau[72]

Finland: Karola Rehnstrom[73,74], Lennart von Wendt[73,74], Leena Peltonen[73,74,75]

AFFILIATION LIST

[1]Center for Complex Disease Genomics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD 21205; [2]Institute for Juvenile Research, Department of Psychiatry, University of Illinois at Chicago, Chicago, IL; [3]Center for Human Genetic Research, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114; [4]Broad Institute of MIT and Harvard, Cambridge, MA 02142; [5]Developmental Medicine Center, Children's Hospital Boston, Boston, MA, 02115; [6]Division of Genetics, Children's Hospital Boston and Harvard Medical School, Boston, MA, 02115; [7]Special Education Organization, Tehran, IRAN; [8]University Department of Psychiatry, Warneford Hospital, Headington, Oxford, UK. [9]Newcomen Centre, Guy's Hospital, London, UK. [10]Battelle Center for Mathematical Medicine, The Research Institute at Nationwide Children's Hospital and The Ohio State University, Columbus, Ohio, USA. [11]Child and Adolescent Mental Health, University of Newcastle, Sir James Spence Institute, Newcastle upon Tyne, UK. [12]INSERM U952, Université Pierre et Marie Curie, Paris, France. [13]Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, J.W Goethe University Frankfurt, Frankfurt, Germany. [14]Department of Child and Adolescent Psychiatry, Institute of Psychiatry, London, UK. [15]Autism Research Unit, The Hospital for Sick Children and Bloorview Kids Rehabilitation, University of Toronto, Toronto, Ontario, Canada. [16]Department of Pediatrics and Psychology, Dalhousie University, Halifax, Nova Scotia, Canada. [17]Laboratory of Molecular Neuropsychiatry, Seaver Autism Center for Research and Treatment, Departments of Psychiatry, Genetics and Genomic Sciences, and Neuroscience, Mount Sinai School of Medicine, New York, NY 10025, [18]Department of Human Genetics, University of California - Los Angeles School of Medicine, Los Angeles, California, USA. [19]Psychiatry Department, University of Utah Medical School, Salt Lake City, Utah, USA. [20]Autism and Communicative Disorders Centre, University of Michigan, Ann Arbor, Michigan, USA. [21]Miami Institute for Human Genomics, University of Miami, Miami, Florida, USA. [22]Depts. of Psychology and Psychiatry, University of Washington, Seattle, Washington, USA. [23]Department of Child Psychiatry, University Medical Center, Utrecht, The Netherlands. [24]University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA. [25]School of Medicine and Medical Science University College, Dublin, Ireland. [26]Division of Psychiatry, McGill University, Montreal, Quebec, Canada. [27]Department of Child and Adolescent Psychiatry, Saarland University Hospital, Homburg, Germany. [28]Autism Genetics Group, Department of Psychiatry, School of Medicine, Trinity College Dublin, Ireland. [29]Department of Neurology, University of California - Los Angeles School of Medicine, Los Angeles, California, USA. [30]Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton, Ontario, Canada. [31]Academic Department of Child Psychiatry, Booth Hall of Children's Hospital, Blackley, Manchester, UK. [32]Centre for Human Genetics Research, Vanderbilt University Medical Centre, Nashville, Tennessee, USA. [33]Child and Adolescent Psychiatry and Child Development, Stanford University School of Medicine, Stanford, California, USA. [34]Deutsches Krebsforschungszentrum, Molekulare Genomanalyse, Heidelberg,

Germany. [35]Centre for Integrated Genomic Medical Research, University of Manchester, Manchester, UK. [36]Department of Pediatrics, McMaster University, Hamilton, Ontario, Canada. [37]Centre d'Eudes et de Recherches en Psychopathologie, University de Toulouse Le Miral, Toulouse, France. [38]Wellcome Trust Centre for Human Genetics, University of Oxford, UK. [39]University Department of Child Psychiatry, Athens University, Medical School, Agia Sophia Children's Hospital, Athens, Greece. [40]Department of Medicine, School of Epidemiology and Health Science, University of Manchester, Manchester, UK. [41]Carolina Institute for Developmental Disabilities, University of North Carolina, Chapel Hill, North Carolina, USA. [42]Social, Genetic and Developmental Psychiatry Centre, Institute Of Psychiatry, London, UK. [43]Veterans Affairs, Department of Medicine, University of Washington, Seattle, Washington, USA. [44]The Centre for Applied Genomics and Program in Genetics and Genome Biology, The Hospital for Sick Children and Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada. [45]Department of Pediatrics and Howard Hughes Medical Institute Carver College of Medicine, University of Iowa, Iowa City, Iowa, USA. [46] Vanderbilt Kennedy Center and Center for Molecular Neuroscience, Vanderbilt University, Nashville, Tennessee, USA. [47]Child Study Centre, Yale University, New Haven, Connecticut, USA. [48]Department of Psychiatry, Carver College of Medicine, Iowa City, Iowa, USA. [49]Department of Biostatistics and Medicine, University of Washington, Seattle, Washington, USA. [50]Department of Pediatrics, University of Alberta, Edmonton, Alberta, Canada. [51]Instituto Nacional de Saude Dr Ricardo Jorge Instituto Gulbenkian de Cîencia Lisbon, Portugal. [52]Hospital Pediatrico de Coimbra Coimbra, Portugal. [53]Department of Child and Adolescent Psychiatry, Goteborg University, Goteborg, Sweden. [54]INSERM U995, Department of Psychiatry, Groupe hospitalier Henri Mondor-Albert Chenevier, AP-HP, Créteil, France. [55]Department of Medicine, University of Washington, Seattle, Washington, USA. [56]Stella Maris Institute, Department of Child and Adolescent Neurosciences, Italy. [57]Department of Psychiatry, Indiana University School of Medicine, Indianapolis, USA [58]Department of Biology, University of Bologna, Italy, [59]Pathology and Laboratory Medicine, University of Pennsylvania, Pennsylvania, USA. [60]Autism Speaks, USA. [61]Department of Neurology and Howard Hughes Medical Institute, Beth Israel Deaconess Medical Center, Boston, MA. [62]Department of Child Psychiatry, Istanbul Faculty of Medicine, Istanbul University, Istanbul, Turkey. [63]Department of Neurosciences and Pediatrics, King Faisal Specialist Hospital and Research Centre, Jeddah, Kingdom of Saudi Arabia. [64]Clinical Neurosciences & Pediatrics, Brown University School of Medicine, Providence, Rhode Island. [65]Department of Neurology, Combined Military Hospital, Lahore, Pakistan. [66]Kuwait Center for Autism, Kuwait City, Kuwait. [67]Department of Anatomy and Neurobiology, Boston University School of Medicine, Boston, MA. [68]Department of Child Psychiatry and Department of Clinical Genetics, Oulu University Hospital and Oulu University, Oulu, FINLAND. [69]Medical Genetic Diagnosis Department, National Institute for Genetic Engineering and Biotechnology, Tehran, Iran. [70]Casa de Corazon, Taos, New Mexico. [71]Centre de recherche du CHUM, Hôpital Notre-Dame, Montreal, Quebec, Canada. [72]Sainte-Justine Hospital Research Center, Universite de Montreal, Montréal, Quebec, Canada. [73]Department of Medical Genetics, University of Helsinki, Helsinki, Finland. [74]Department of Molecular Medicine, National Public Health Institute, Helsinki, Finland. [75]Wellcome Trust Sanger Institute, Hinxton, United Kingdom, [76]Center for Neurodegeneration and Experimental Therapeutics, University of Alabama School of Medicine, Birmingham, AL.

[a]Current affiliation: Department of Psychiatry, Institute for Human Genetics, Center for Neurobiology and Psychiatry, UCSF, San Francisco, CA.

[b]Current affiliation: Deptartment of Molecular Biology, Cell Biology and Biochemistry, and Institute for Brain Science, Brown University, Providence, RI

[x]deceased
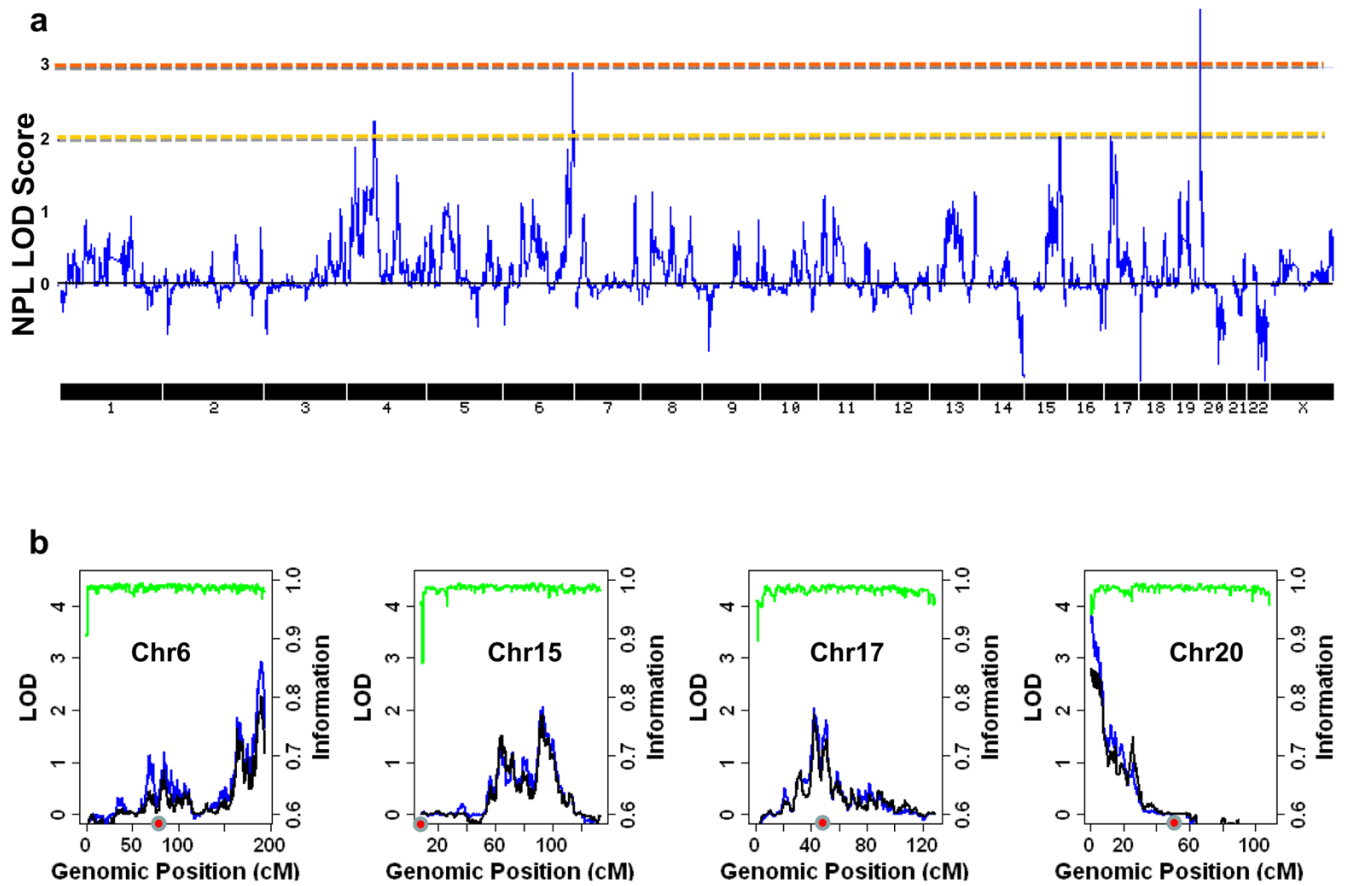
Figure 1.
Genome-wide Linkage Results.

The genome-wide linkage results are shown in 1a, with the orange line indicating LOD 3. The four chromosomes with LOD > 2 are shown in 1b. The black and blue lines indicate results from families with both parents genotyped and all families, respectively. The green line indicates information content. The red circle indicates the position of the centromere.

Figure 2.
SEMA5A Expression in Autism Brains
SEMA5A gene expression is shown relative to *MAP2*. Yellow diamonds indicate individual expression levels for each sample; error bars indicate standard error (SE).

Table 1

Top TDT results and replication data.

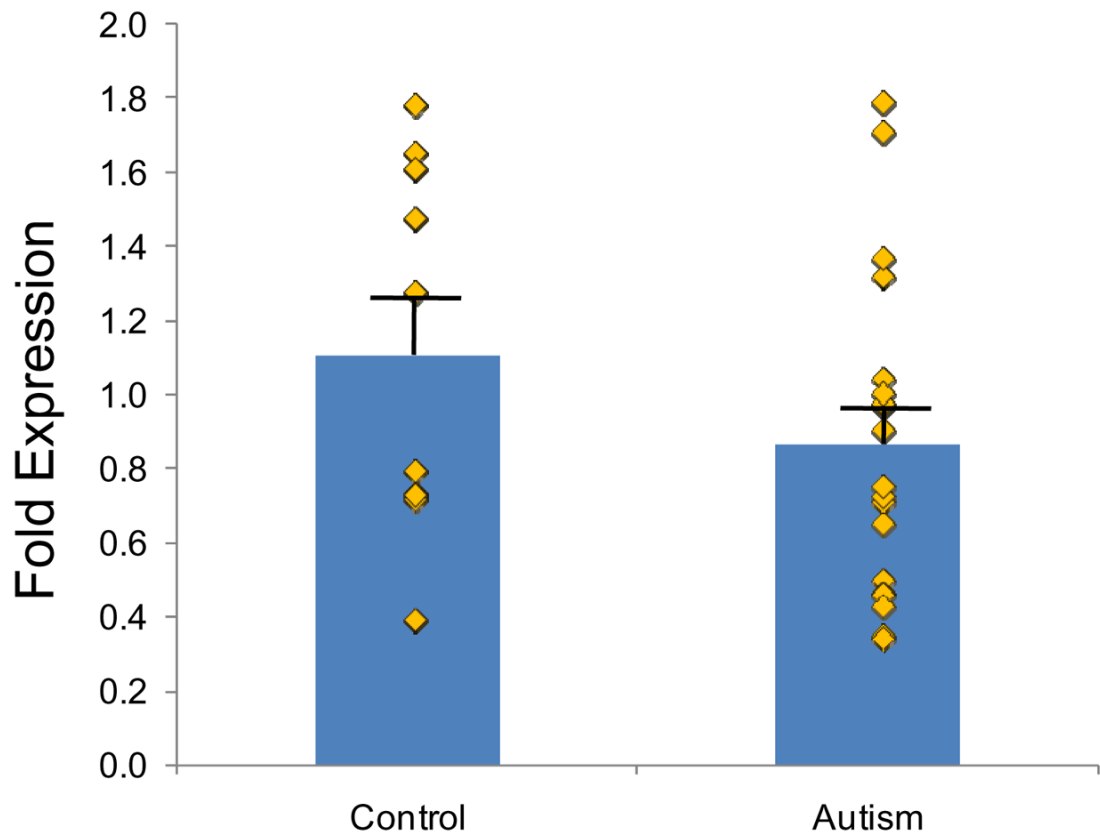| Locus Chr	Position	SNP	LD, proxy	Scan				Replication					Meta-analysis	
				T	U	OR	P	P [w case-control]	T	U	OR	P [1-sided]	P meta	P proxy
4	68019960	rs17088254		137	219	0.63	1.4×10^{-5}	$\mathbf{8.5 \times 10^{-6}}$	48	38	1.3	NA	0.011	4.8×10^{-3}
4	68189460	rs2632453	$r^2 = 0.67$, rs17088254	171	245	0.70	2.9×10^{-4}	2.4×10^{-4}	248	234	1.06	NA	0.022	
5	9676622	rs10513025		84	152	0.55	9.6×10^{-6}	$\mathbf{1.7 \times 10^{-6}}$	152	199	0.76	$\mathbf{6.1 \times 10^{-3}}$	$\mathbf{2.1 \times 10^{-7}}$	
6	15365718	rs13208655	$r^2 = 0.74$, rs7766973	NA	NA	NA	NA	NA	831	829	1.00	0.48	0.48	
6	15376030	rs7766973		631	811	0.78	2.1×10^{-6}	$\mathbf{6.8 \times 10^{-7}}$	139	142	0.98	0.43	2.0×10^{-4}	2.8×10^{-4}
9	7763723	rs4742408		591	739	0.80	4.9×10^{-5}	2.7×10^{-4}	241	224	1.08	NA	0.030	
9	7764180	rs4742409		499	645	0.77	1.6×10^{-5}	$\mathbf{7.9 \times 10^{-6}}$	77	87	0.89	0.22	1.6×10^{-4}	3.6×10^{-4}
9	7764774	rs6477233	$r^2=0.6$, rs4742409	NA	NA	NA	NA	NA	734	752	0.98	0.32	0.32	
9	86471331	rs952834		656	825	0.80	1.1×10^{-5}	$\mathbf{7.8 \times 10^{-6}}$	173	160	1.08	NA	5.4×10^{-3}	
10	68842909	rs7923367		89	160	0.56	6.8×10^{-6}	$\mathbf{3.4 \times 10^{-6}}$	18	25	0.72	0.14	4.1×10^{-5}	
11	22775950	rs12293188		449	327	1.37	1.2×10^{-5}	$\mathbf{1.1 \times 10^{-6}}$	486	513	0.95	NA	3.0×10^{-3}	
11	22785182	rs16910190		421	308	1.37	2.8×10^{-5}	1.4×10^{-5}	55	67	0.82	NA	0.014	
11	22785488	rs16910194		444	330	1.35	4.2×10^{-5}	$\mathbf{3.7 \times 10^{-6}}$	80	75	1.07	0.34	2.8×10^{-4}	
11	22791645	rs3763947		429	320	1.34	6.8×10^{-5}	3.4×10^{-5}	57	57	1.00	NA	2.4×10^{-3}	

Top results from the combined TDT and case-control analysis are shown ($P < 10^{-5}$), with replication data, where it exists. For Sequenom genotyping that used a proxy SNP, that SNP and its LD (r^2) with the SNP of interest is shown. Transmitted (T) and untransmitted (U) counts and odds ratios (OR) for the minor allele are shown for each SNP. Replication results are shown for additional autism family data using Affymetrix and Sequenom genotyping technology. The meta-analysis P-value is shown, and the P-value for meta-analysis where proxy SNP data was included. Bold font, $P < 10^{-5}$ TDT/case-control analysis, $P < 0.05$ replication, $P < 2.5 \times 10^{-7}$ meta-analysis.

Table 2

Chromosome 5p15 SNPs.

SNP	Chr	Position	MAF	OR	P	Replication P
rs10513025	5	9676622	0.041	0.5526	0.000009581	0.006059
rs10513026	5	9677106	0.040	0.53	4.50E-06	na
rs16883317	5	9701592	0.038	0.53	7.20E-05	na

Three SNPs in the chromosome 5p15 association locus genotyped by Sequenom iPLEX are shown, with minor allele frequency (MAF), odds ratio (OR), and P-value in the AGRE and NIMH sample, as well as replication data from all available samples for rs10513025 (see Methods).