# A Genome Scan for Loci Shared by Autism Spectrum Disorder and Language Impairment

Christopher W. Bartlett, Ph.D.

Liping Hou, Ph.D.

Judy F. Flax, Ph.D.

Abby Hare, B.S.

Soo Yeon Cheong, Ph.D.

Zena Fermano, M.A.

Barbie Zimmerman-Bier, M.D.

Charles Cartwright, M.D.

Marco A. Azaro, Ph.D.

Steven Buyske, Ph.D.

Linda M. Brzustowicz, M.D.

**Objective:** The authors conducted a genetic linkage study of families that have both autism spectrum disorder (ASD) and language-impaired probands to find common communication impairment loci. The hypothesis was that these families have a high genetic loading for impairments in language ability, thus influencing the language and communication deficits of the family members with ASD. Comprehensive behavioral phenotyping of the families also enabled linkage analysis of quantitative measures, including normal, subclinical, and disordered variation in all family members for the three general autism symptom domains: social, communication, and compulsive behaviors.

**Method:** The primary linkage analysis coded persons with either ASD or specific language impairment as "affected." The secondary linkage analysis consisted of quantitative metrics of autism-associated behaviors capturing normal to clinically severe variation, measured in all family members.

**Results:** Linkage to language phenotypes was established at two novel chromosomal loci, 15q23–26 and 16p12. The secondary analysis of normal and disordered quantitative variation in social and compulsive behaviors established linkage to two loci for social behaviors (at 14q and 15q) and one locus for repetitive behaviors (at 13q).

**Conclusion:** These data indicate shared etiology of ASD and specific language impairment at two novel loci. Additionally, non-language phenotypes based on social aloofness and rigid personality traits showed compelling evidence for linkage in this study group. Further genetic mapping is warranted at these loci.

Autism is a severe neurodevelopmental disorder characterized by altered functioning in three domains: 1) social interaction, 2) communication, and 3) stereotyped behavior and/or restricted interests and activities. Separately, there is a well-known classification in language disorders for children who have difficulties in acquiring language but are otherwise neurologically and psychologically normal, known as specific language impairment. Given that both disorders have a language component to their diagnosis, previous work has suggested that specific language impairment and autism could have shared genetic contributors. This hypothesis is supported by a series of genetic mapping studies examining the relationship between autism and language impairment in complementary ways (1–11).

Autism linkage mapping studies have examined the relationship of language and autism with two paradigms. The first paradigm used the language delay status of probands with autism spectrum disorder (ASD) to stratify families into two groups, most often based on presence or absence of phrase speech by 36 months of age. Stratification on phrase speech delay is based on the premise that there are subgroups of autism that can be genetically differentiated by language status. The language stratification paradigm has yielded autism findings on chromosomes 2q24–32 (5, 6), 7q22–32 (4), and 13q21–22 (4); this 13q region is also linked to specific language impairment in non-ASD groups (2, 3, 7). At all three locations, stratifying ASD families on phrase speech delay in an a priori design nearly perfectly separated families that were linked to the given locus from families that were not. Some loci were linked only in families including a person with ASD and phrase speech delay, and other loci were linked only in families without phrase speech delay. In contrast, a replication study used both stratification of families and determination of affection status in non-ASD parents based on self-report of language impairment history in childhood (8). This study found that the coding of parents as affected on the basis of a history of language intervention or language delay minimally influenced the results, and while the group with phrase speech delay did have more evidence of linkage, the chromosome 2, 7, and 13 findings were not replicated.

The second paradigm used language phenotypes to directly map loci for either a quantitative language trait or language impairment as a dichotomous trait within families with ASD. Genome-wide locus mapping for the quantitative language trait in ASD was performed by using two items from the Autism Diagnostic Interview–Revised

(12), "age in months of onset of first word" and "age in months of onset of phrase speech"; both showed linkage to 7q36 (9, 10). *CNTNAP2*, a gene also associated with both specific language impairment and normal language development, was later shown to be the most likely candidate responsible for this linkage (11), thus providing evidence that genetic variation relevant to both ASD and language impairment can occur within the same gene.

We present here what we believe to be the first linkage/association study to use a novel paradigm that complements both the stratification and quantitative trait locus approaches, in order to better understand the relevance of language variation in ASD. We selected families for the presence of both autism and specific language impairment under the hypothesis that if the two conditions are etiologically related, then this selection scheme will enrich our study group for loci related to language difficulties in autism and will also reduce the genetic heterogeneity of ASD. Within these families, we collected state-of-the-art language phenotypes in relatives and, whenever possible, individuals with ASD. This study thus likely represents the most comprehensive family-based language phenotyping in a molecular genetic study of ASD.

We tested the hypothesis of genetic overlap by performing our analysis assuming that ASD and language impairments have the same underlying genetic etiology (i.e., we considered phenotypes from both disorders as equally affected) and then performed genome-wide linkage scans. Our analyses coded individuals with oral or written language impairments without ASD, as well as individuals with similar types of oral and written language impairments and ASD, as "affected." Additionally, persons with ASD who could not be evaluated by means of quantitative language measures were incorporated into the analysis by using a method for censored data (i.e., systematically missing data), which in this case was based on the assumption that the censoring was due to low language ability (13) (see also the supplementary methodological information in the data supplement accompanying the online version of this article). Our study design is most effective for mapping loci that are etiologically relevant to both ASD and language impairment. If ASD and language impairment are genetically unrelated in these families, then coding both disorders as "affected" in the same analysis will reduce evidence for linkage or association. Positive findings were formally tested to determine whether autism or specific language impairment or both contributed to detected linkage signals.

The phenotypic battery also included 17 quantitative population-normed language assessments and quantitative measures of social and compulsive behaviors. As a secondary objective, we performed linkage analysis using these quantitative data to capture both normal variation and clinically severe variation on the same scale, an analysis generally considered to be more powerful than analyses that classify subjects only as affected or unaffected. To our knowledge, these analyses included the first use of the Yale-Brown Obsessive Compulsive Scale (YBOCS) (14–17) in a molecular genetic study of autism.

## Method

### Overview of Design

Our primary goal was to find genetic variation relevant to both language impairments and ASD, using a set of previously described families recruited for the presence of persons with autism and of separate individuals with specific language impairment (18). To accomplish our goal, we created subgroups from 79 families (see Table S1 in the online data supplement) according to phenotypic characteristics. We established three groups of families that we denoted tiers 1–3. Tier 1 (N=46) consisted of families with both an autism proband and a different proband with specific language impairment, as defined in this study group previously (18), or, in a few cases, one autism proband and one ASD proband with low language ability (19, 20). Individuals with ASD plus language impairment were contrasted with ASD probands who had normal language ability and ASD probands who were nonverbal; this last category is often ignored in the literature since language cannot be quantitatively assessed.

Tier 1 had an internal contrast since any observed linkage could be further examined by excluding either the language-impaired proband or the autism proband to understand the connection of each disorder to the linkage signal on a locus-by-locus basis. The other two tiers included ASD families that did not have a language-impaired proband or a proband with ASD plus language impairment according to direct testing (in the literature this is often called ALI for "autism-language impaired"). Tier 2 (N=15) consisted of multiplex ASD families without a language-impaired proband in which neither ASD proband was operationally defined as language impaired (neither having structural language deficits nor nonverbal). Tier 3 contained families (N=9) with an ASD proband and at least one relative who scored in the impaired range on either the Social Responsiveness Scale, a well-studied inventory of social functioning (21, 22), or the YBOCS, used to evaluate obsessive-compulsive disorder (OCD) in psychiatric evaluations (14–17). For association analysis, an additional nine ASD trios (N=9) were added to tier 3.

Prior to behavioral testing all subjects gave informed consent conforming to the guidelines for treatment of human subjects at Rutgers University. All family members, including higher-functioning family members with ASD, received age-appropriate measures of language and reading (see Table S2 in the online data supplement). Descriptive statistics by diagnostic group (specific language impairment, autism, and other) have been previously published (18). Observed correlations between measures are in Table S3 of the online data supplement.

For the purpose of categorical phenotype linkage or association analysis, we defined oral language impairment, called "LI" in our previous articles (2, 3, 7), as either an age-appropriate Clinical Evaluation of Language Fundamentals–Fourth Edition (CELF−4) core standard score of 85 or less, or at least 1 SD below peers on at least 60% of all oral language subtest scores, and a significant history of language or reading difficulties, defined as more than 2 years of intervention and/or a childhood diagnosis of language and/or reading impairment. For the purpose of finding loci that jointly influence oral language impairment and ASD, we defined "LI*" as a phenotype that includes persons affected with our definition of language impairment as well as persons affected with ASD (i.e., etiological equivalence).

In our previous studies of multiplex language impairment families, we observed many instances of semicompensated adults with a childhood diagnosis of language problems and

currently demonstrating weak language skills who did not cross the threshold for language impairment but did meet the criteria for reading impairment (2, 3, 7). On the basis of our prior successful mapping of a locus for specific language impairment with that reading impairment phenotype (2, 3, 7), we defined written language (reading) impairment, called "RI" in our previous publications (2, 3, 7, 18), as at least 1 SD below the population mean on 60% of all reading tests and subtests. For the purpose of finding shared loci that jointly influence written language impairment and ASD, we defined "RI*" as a phenotype that includes persons affected with our definition of reading impairment as well as persons affected with ASD (i.e., etiological equivalence). Throughout this article, LI* and RI* refer to our specific diagnostic definitions of language impairment and/or ASD and reading impairment and/or ASD, respectively, while the term "language impairment" is meant in a more general sense to apply to oral and/or written language impairments, in context.

### Genotyping

Affymetrix Axiom 1.0 arrays (Affymetrix, Santa Clara, Calif.) were used to generate 567,893 single-nucleotide polymorphism (SNP) genotypes on 440 individuals from the 79 families. Quality control on the SNP genotypes was conducted as described previously (7), with additional details included in the supplementary methodological information in the online data supplement, according to individual/SNP genotype completion, relationship checking, Mendelian errors, and ancestry. A subset of 8,086 SNPs was chosen for linkage analysis to minimize marker-to-marker linkage disequilibrium and retain a high frequency of minor alleles to provide suitable genomic coverage of recombination events in the pedigrees. The association analysis used all SNPs that met quality control standards and had a frequency of minor alleles greater than 0.05, yielding 529,874 SNPs. Validation genotyping was conducted on a Luminex 200 machine (Luminex, Austin, Tex.) by means of a custom oligonucleotide ligation assay (23), with allele calling and quality control as described elsewhere (7, 24).

### Statistical Analysis

**Overall data analysis plan.** We first conducted genome-wide linkage scans with follow-up association analysis in the linkage regions. We also conducted a genome-wide association analysis over the remainder of the genome. Given the depth of phenotyping, it was not considered reasonable to perform univariate analyses of all 21 cognitive measures on a genome-wide basis because of the difficulty of interpreting results from analyses of many correlated traits. Instead, we opted for a mix of empirically and theoretically driven phenotypes. We used two categorical phenotypes, LI* for oral language impairment and RI* for written language (reading) impairment, the latter being a strong indicator of an unresolved oral language impairment in multiplex pedigrees with specific language impairment (reading deficits caused by an underlying language deficit) according to our previous studies (2, 3, 7). We also derived three quantitative traits using a factor analysis (details appear in the online data supplement) to reduce the phenotypic data, which we called factors 1 through 3 (factor loadings are shown in Table S4). To elucidate possible shared etiology in language-phenotype-linked regions, follow-up analyses were conducted to assess whether ASD or language impairment or both were required to detect the linkage peak. The nonlanguage traits were analyzed as quantitative traits, which included scores on the YBOCS and the Social Responsiveness Scale. Scores on the Social Responsiveness Scale were also analyzed as a dichotomous trait based on a threshold indicating mild impairment (see online data supplement).

**Linkage/association analysis methods.** Linkage and association analyses were conducted with the KELVIN 2.3.3 package (25;

http://kelvin.mathmed.org/). KELVIN implements the posterior probability of linkage (PPL) metric to measure the probability that a genetic location is linked with the trait of interest and the combined posterior probability of linkage disequilibrium (cPPLD) metric to measure the probability that a SNP is in linkage disequilibrium with the trait of interest conditional on the evidence for linkage at a given locus. It is important to note that the cPPLD uses a pedigree likelihood that explicitly accounts for family structure while assessing the evidence that linkage disequilibrium between the SNP and the disease is present (26, 27). In quantitative trait analysis for which some ASD subjects lacked data because of an inability to participate in some cognitive tests, the measures for those individuals were treated as censored data, meaning the true scores were unknown but were known to be below a threshold. These analyses were handled by using the PPL model for censored data (13, 28), as described in the online data supplement. The sex-averaged marker map for linkage was obtained from the Rutgers Combined Linkage-Physical Map of the Human Genome (29).

Primary linkage analysis of language phenotypes was conducted on each of the three tiers separately, and the linkage evidence was sequentially updated across the three tiers to provide a single metric for linkage evidence. Follow-up family-based association analysis was conducted similarly; however, families that contained persons that were not of European ancestry (N=1 in tier 3) were dropped from the association analysis since combining samples with different genetic ancestries can generate false positive results (24). For nonlanguage phenotypes, all families were run as a single data set.

**Statistical correction for multiple phenotypes.** In order to assess the effect of performing multiple genome scans using correlated phenotypic traits, we simulated 3,000 genomes without regard to phenotypes to create an empirical null distribution for estimating p values. We simulated chromosomes using the same SNP allele frequency and genetic distances as in our linkage data set. For each simulated genome, we conducted an analysis with the five phenotypes, saving the overall maximum PPL per replicated genome. This list of maximum PPLs is the null distribution accounting for analysis with our five phenotypes, using our specific pedigree configuration and patterns of missing data, etc. After correction for multiple phenotypes, a PPL of 0.347 or greater retained a genome-wide error rate consistent with p<0.001, a PPL of 0.269 corresponded to p<0.01, and a PPL of 0.104 corresponded to p<0.05. These threshold values are slightly higher than those in two previous studies of the false positive rate of the PPL (3, 30) because of correction for multiple phenotypes in the present study.

## Results

### Initial Linkage Results for Language

We performed a genome-wide linkage analysis using two categorical definitions of language impairment (affected and unaffected status) and three quantitative language scores from a factor analysis of all 21 language measures, which we denoted factor 1, factor 2, and factor 3 (see Table S4 in the online data supplement). The first trait, abbreviated LI*, defined individuals as affected if they possessed either an oral language impairment or ASD. The second trait, abbreviated RI*, defined individuals as affected if they possessed either a written language impairment or ASD. LI* produced clear evidence of linkage to chromosome 15, and RI* showed linkage to chromosome 16, as shown in Figure 1

**FIGURE 1. Genome-Wide Linkage Analysis of Five Language-Related Traits in 70 Families With Both Autism Spectrum Disorder (ASD) and Specific Language Impairment[a]**
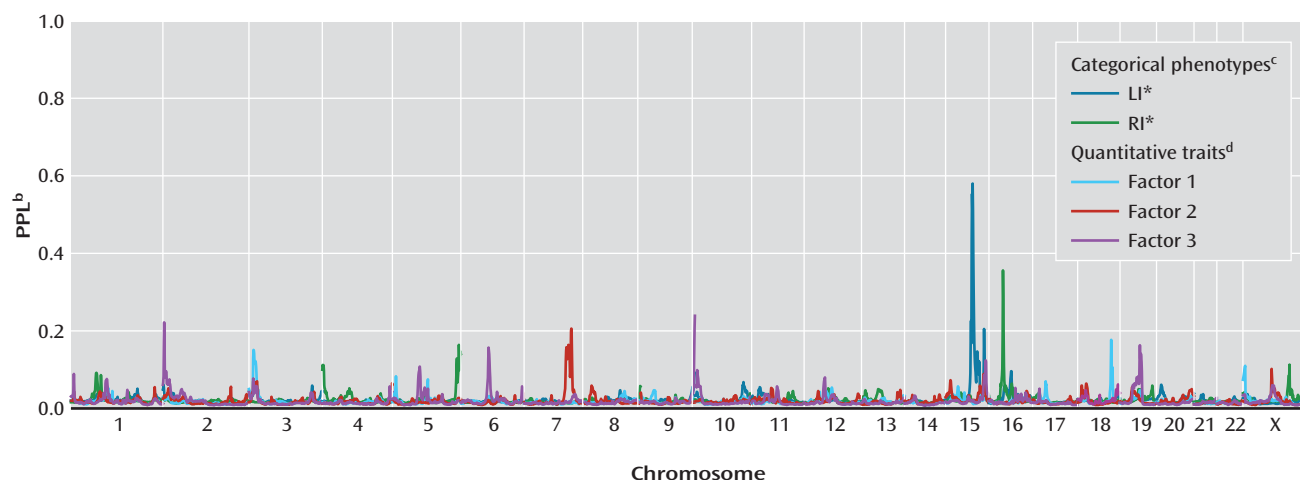


[a] The PPL is scaled such that values below 0.02 represent evidence against linkage to that location while values higher than 0.02 represent evidence for linkage to that location. A PPL of exactly 0.02 indicates that the data are not informative for linkage. The peaks on chromosomes 15 and 16, which represent the LI* and RI* phenotypes, respectively, clearly stand out from the rest of the genome, and overall the PPL displays a high signal-to-noise ratio for linkage mapping. While the three factor scores lack strong peaks, several regions of potential interest are identified.
[b] PPL, posterior probability of linkage.
[c] LI* represents oral language impairment and/or ASD. RI* represents reading impairment and/or ASD.
[d] The factors were derived from 21 standardized measures of language, as described in the data supplement accompanying the online version of this article. Factor loadings are presented in Table S4 in the data supplement.

(with a summary of all large linkage peaks given in Table 1). The PPL is scaled such that the genome-wide plots show very clear signal-to-noise ratios, as seen in the figure. The magnitudes of both signals (>0.35) have type I error rates appropriate for establishing linkage in a genome-wide scan, even after accounting for testing of multiple phenotypes (see Method section). Table 1 also includes the mod score, the fully maximized lod score (logarithm of the odds ratio for linkage), to allow comparison of the PPL to a statistic with a more commonly used scale; the estimated disease gene frequency and risk probabilities by genotype (penetrances) for the linkage peak are also included in the table, including the estimate of the proportion of families linked to a given locus ($\alpha$).

Chromosome 15q23–26.2 was linked to LI* with a maximum PPL of 0.57 and implicating a region of 24.2 Mb from 73 cM to 106 cM. As can be seen on the left side of Figure 2, the linked region has a 15-cM high-confidence linkage region with a much larger low-confidence region (to the right) that accounts for about half of the total implicated region. Nonverbal IQ was not linked at this locus, indicating a dissociation of language and intelligence at this locus, as expected, since specific language impairment does not include deficits in intelligence. We next wanted to assess how the level of language impairment in the ASD subjects (language impairment, normal language, or nonverbal) in each family modulates the linkage signal. We therefore defined a metric to quantify the relative contribution of the three language levels, which indicated homogeneous contributions from all three groups to the linkage at this peak

(in the online data supplement, see "Assessing the relative contribution of the three proband types to the final PPL").

Family-based association analysis of all available SNPs with frequencies of minor alleles above 0.05 in the linked region yielded only weak evidence of association, with a maximum cPPLD of 0.06. These results are not consistent with strong evidence for association that would account for the observed linkage. However, the Axiom array SNPs in the region successfully haplotype tagged only 48% of the common variation (as described in the online supplemental methods, with results in Table S5); thus, follow-up of the region based on this SNP genotyping platform should be considered incomplete.

Chromosome 16p12.1–12.3 was linked to RI* over 8.9 Mb from 35 cM to 49 cM, with a maximum PPL of 0.36 (Figure 2, right). All three ASD language levels contributed equally to the PPL (see "Assessing the relative contribution of the three proband types to the final PPL" in the online data supplement). Similar to the chromosome 15 results, analysis of nonverbal IQ as a quantitative trait yielded evidence against linkage at this locus (PPL=0.017, or below the prior probability of linkage, which is 0.02), indicating a dissociation of language and intelligence at this locus as well. Follow-up cPPLD analysis to find SNPs that accounted for the linkage signal yielded a maximum cPPLD of 0.08 at a single SNP (next highest cPPLD=0.02). Genotyped SNPs in this region successfully tagged only 55% of the common variation (Table S5 in the online data supplement).

Additional language-related peaks of interest (PPL>0.20) were observed on chromosome 7 with factor 2 (the second

TABLE 1. Large Peaks[a] in Linkage Analysis of Language Impairment in 70 Families With Autism Spectrum Disorder (ASD) and Specific Language Impairment

| Phenotype | Chromo-some | cM | PPL | Band Range | Width (Mb) | Fully Maximized Lod Score[b] | Estimated Genotypic Effect for Locus[c] | | | Disease Gene Frequency | Heterogeneity Parameter in Admixture Likelihood (α)[d] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | −/− | +/− | +/+ | | |
| Yale-Brown Obsessive Compulsive Scale | 13 | 58 | 0.36 | q14.3–21.3 | 17.3 | 4.2 | −3.0 | 0.00 | 3.00 | 0.300 | 0.7 |
| Social Responsiveness Scale dichotomous trait | 14 | 117 | 0.37 | q32.2–32.23 | 7.6 | 3.5 | 0.0 | 0.10 | 0.80 | 0.100 | 1.0 |
| LI* trait[e] | 15 | 83 | 0.57 | q23–26.2 | 24.2 | 4.1 | 0.0 | 0.70 | 0.99 | 0.001 | 1.0 |
| Social Responsiveness Scale quantitative trait | 15 | 120 | 0.52 | q26.2–26.3 | 6.2 | 4.5 | −2.0 | 1.00 | 2.00 | 0.100 | 0.9 |
| RI* trait[e] | 16 | 43 | 0.36 | p12.1–12.3 | 8.9 | 4.6 | 0.0 | 0.00 | 0.90 | 0.200 | 1.0 |

[a] A large peak was defined as a posterior probability of linkage (PPL) above 0.35.
[b] Sometimes referred to as a mod score.
[c] For categorical analysis these quantities are penetrances, and for quantitative traits they are genotypic means on a z-score scale.
[d] Estimate of the proportion of families linked to a given locus.
[e] LI* represents oral language impairment and/or ASD. RI* represents reading impairment and/or ASD.

FIGURE 2. Follow-Up Association Analysis of SNPs Under Genetic Linkage Peaks for Oral Language and Reading Impairment Phenotypes in 79 Families With Both Autism Spectrum Disorder (ASD) and Specific Language Impairment
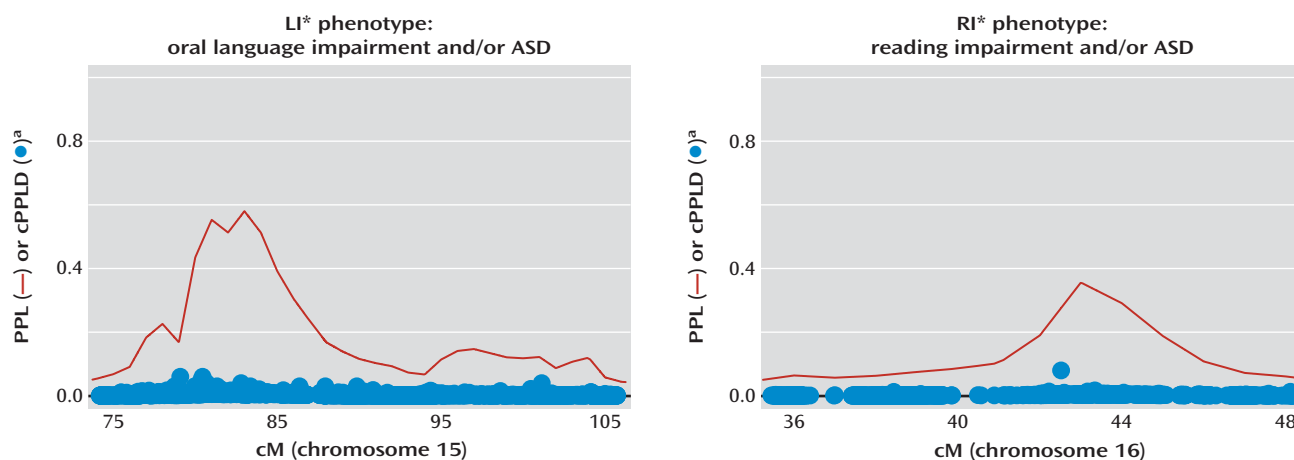


[a] PPL, posterior probability of linkage. cPPLD, combined posterior probability of linkage disequilibrium.

trait defined from the factor analysis of all 21 language tests) and chromosomes 2 and 9 with factor 3 (the third trait defined from our factor analysis). The linkage to chromosome 7 (PPL=0.21), located over the region containing *CNTNAP2*, has been replicated in both ASD and specific language impairment. This PPL may be considered appropriately large to replicate the *CNTNAP2* locus.

### Further Characterization of the Roles of Language and Autism at Linked Loci

We sought to define the relative contributions of autism and language impairment to the linkage results on 15q and

16p, i.e., to assess whether the two disorders contribute equally to those linkage findings as an indication of the equivalence of the disorders. We restricted these analyses to tier 1 since only this subset of families contains both autism and language-impaired probands in each pedigree. Each locus was assessed separately. We assessed the specificity of chromosome 15 for language impairment versus autism by removing the autism proband from each pedigree and repeating the linkage analysis, then doing a separate and equivalent analysis after removing the language-impaired proband and retaining the autism proband. In both cases, the linkage signal was greatly reduced (PPLs dropped to 0.02 and 0.04). Similarly, on

FIGURE 3. Genome-Wide Linkage Analysis of Three Non-Language-Related Traits in 70 Families With Both Autism Spectrum Disorder (ASD) and Specific Language Impairment[a]
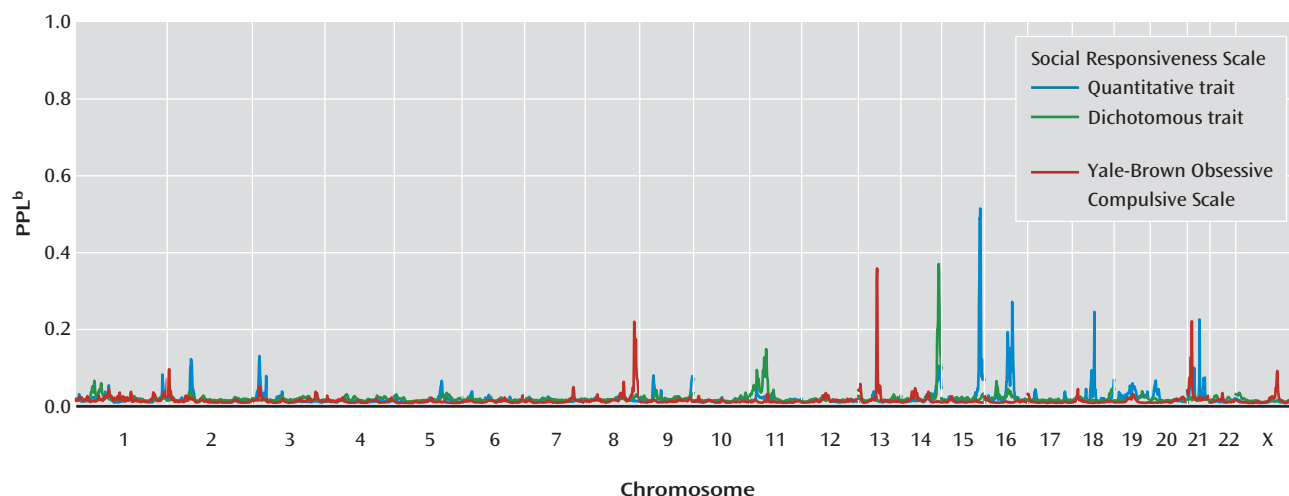


[a] The largest signals for the three traits were on chromosomes 13, 14, and 15. No overlap was observed between the analyses of the Social Responsiveness Scale quantitative and dichotomous traits.
[b] PPL, posterior probability of linkage.

chromosome 16 we removed the autism proband from each pedigree, yielding a PPL of 0.03, and removed one non-ASD reading-impaired subject from the linkage analysis, giving a PPL of 0.06. Therefore, both the 15q and 16p loci are sensitive to the presence of both autism and language impairment. However, the reduction in PPL signals could have been the result of lower power due to inclusion of fewer affected individuals in the analysis or loss of specific and relevant disease information. To rule out the former, we then used a permutation study to assess whether removing language-impaired and/or ASD subjects from the analysis induced a greater average drop in the PPL than removing subjects randomly (see also "Assessing the within-family relative contributions of language impairment and ASD to the final PPL" in the online data supplement). For chromosome 15, the permutation test indicated significantly greater effects of specifically removing subjects with language impairment and ASD on the PPL than other combinations (p<0.01). This was not the case for chromosome 16, where the test was not significant, indicating that low power cannot be ruled out as a confound when interpreting the contributions of reading impairment and autism to this linkage peak.

To test whether our phenotypic definitions of LI* and RI* were too restricted, we also repeated the linkage analysis on chromosomes 15 and 16 using a combined phenotypic definition of both oral and written language impairment, where persons were defined as affected if they were either LI* or RI* (or both). For both chromosomes, the PPL was attenuated to less than half the original linkage signal (dropping to 0.19 and 0.12) and the linkage region was greatly broadened (data not shown).

### Linkage Analysis of Nonlanguage Phenotypes

All families were ascertained only for language phenotypes in the individuals without ASD, not for any non-language characteristics of the broad autism phenotype. However, we quantitatively assessed social responsiveness and aspects of obsessive-compulsiveness in all pedigree members able to participate in an assessment, and we used the phenotypes for linkage analysis. Results are summarized in Figure 3. We analyzed the YBOCS score as a quantitative trait yielding a PPL of 0.36 on 13q14.3–21.33 (55 to 63 cM), with the peak directly over *PCDH20*, close to loci in previous linkage studies of ASD and language impairment (2–4). The linked region is 17.3 Mb in size. Analysis of all available SNPs with minor allele frequencies above 0.05 did not yield any cPPLD results over 0.02 (Figure 4, left). Two smaller peaks of interest occurred on chromosomes 8 (PPL=0.22) and 21 (PPL=0.22) with no cPPLDs under those peaks greater than 0.03.

Our analysis of the Social Responsiveness Scale was conducted two ways. The first was as a quantitative trait, using the numerical scores generated by the test. This yielded a linkage peak PPL of 0.52 on chromosome 15q26.2–26.3 (113 to 133 cM) that does not overlap with the LI* peak described above (Figure 4, center). We observed three SNPs spanning 3 Mb (rs12440787, rs7170868, rs9672677) with cPPLD values of 0.07 under this peak, but none was in linkage disequilibrium with another and so they do not represent a single coherent signal. In addition, we observed two smaller peaks on chromosomes 16 (PPL=0.27) and 18 (PPL=0.25). The second analysis used the "mild" impairment threshold (see online data supplement) to dichotomize scores on the Social Responsiveness Scale for analysis. We chose to create this simple distinction between mild impairment and no impairment since this

FIGURE 4. Follow-Up Association Analysis of SNPs Under Genetic Linkage Peaks for Three Non-Language-Related Traits in 79 Families With Both Autism Spectrum Disorder (ASD) and Specific Language Impairment[a]
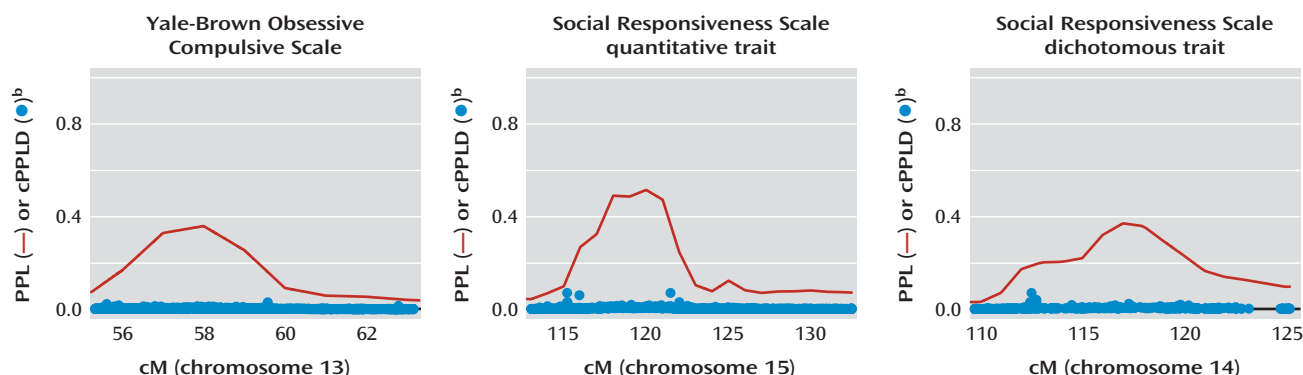


[a] No cPPLD values were observed that could account for the linkage signals with the Yale-Brown Obsessive Compulsive Scale or the Social Responsiveness Scale as a quantitative or dichotomous trait.
[b] PPL, posterior probability of linkage. cPPLD, combined posterior probability of linkage disequilibrium.

classification is commonly used in the broader autism phenotype literature. We saw the peak PPL of 0.37 on chromosome 14q32.2–32.33 (110–126 cM) encompassing 7.7 Mb (Figure 4, right). No cPPLD from available SNPs in these regions was greater than 0.02.

### Genome-Wide Association Analysis

Linkage analysis requires only a limited number of SNPs to attain essentially full information, and thus it used only 8,086 of the 529,874 SNPs that passed quality control filtering. We conducted association analyses on the 529,874 genome-wide SNPs, yielding 19 SNPs with cPPLD values above 0.10 (SNP names are given in Table S6 in the online data supplement). We then performed follow-up genotyping, using a different platform, of these SNPs and SNPs in strong linkage disequilibrium ($r^2>0.95$) for validation, and we conducted cPPLD analysis using both linkage information from the families and linkage disequilibrium with the trait. The highest cPPLD value, 0.20, was located at rs3792495 for the factor 2 trait.

## Discussion

We identified two linkage peaks for language impairment in families with both language impairment and ASD (15q25.1 and 16p12.3) that do not overlap with previously discovered ASD or language impairment loci. The two linkage signals showed specificity for oral language impairments for 15q and for written language impairment for 16p. This specificity was evidenced by attenuation of the linkage signals when a combined oral/written language impairment phenotype was applied, suggesting there is a subset of individuals with reading problems who do not have comorbid oral language deficits. Additionally, there was no evidence that either locus is primarily related to language impairment alone or ASD alone; rather, it appears that each locus is jointly related to both language impairment

and ASD. These findings are in keeping with the goal of the study, to find genetic variation that is relevant to both disorders. To our knowledge, this is the first molecular genetic study of families that have both a language-impaired proband and an autism proband, and we hypothesize that these families have a high genetic loading for impairments in language ability, further influencing the language and communication deficits of the ASD probands. In our previous study (18), we also observed that the average scores on a standardized test of the social use of language (pragmatics) were similar in the language-impaired and ASD subjects in these families. Pragmatic impairment is not part of the defined deficits in specific language impairment but is commonly seen in ASD, suggesting that the two disorders, as seen in the families selected by our ascertainment criteria and recruitment methods, may be on an etiological continuum.

There were also several compelling linkage peaks for nonlanguage traits even though the ascertainment scheme did not include any requirements regarding nonlanguage traits beyond the ASD proband. The Social Responsiveness Scale peak on 15q26.3 was the second largest in the study (PPL=0.52) and had the narrowest linkage region (6.2 Mb). This region was implicated in a meta-analysis of ASD and schizophrenia (31). An additional Social Responsiveness Scale peak was noted when we used a cut-off for mild impairment to create a categorical affection status (14q32; PPL=0.37). The 14q32 region has been associated with autism through cytogenetic abnormalities and copy number variation (32, 33). While social skills and communication are fundamentally related, it is unclear from our data whether ascertainment for performance language assessments increased the power to detect social behavior (34–37) because of that relationship. A conservative interpretation is that the deep phenotyping performed in this study was simply more likely to find multiple strong effects across phenotypic domains than were studies with fewer phenotypic data. The

Social Responsiveness Scale is a good quantitative metric for mapping autism loci, as it has yielded strong findings in other studies (38, 39). However, this appears to be the first gene mapping study to also use a cut-off for mild impairment on the Social Responsiveness Scale, which is more analogous to procedures in the broader literature on the autism phenotype, where the affected/unaffected distinction is commonly applied. We also believe this is the first use of the YBOCS as a quantitative trait in genome-wide analysis of ASD. The empirical performance appears quite good according to our data, suggesting that wider use of this measure may be warranted in ASD research, especially in family genetic studies, where specific behaviors may be apparent in some family members but not severe enough to impair activities of daily living. Our specific finding on 13q does not coincide with results in OCD studies but does align with those in previous autism genetic studies (4, 40, 41).

While the linkage analysis showed several strong peaks with different language- and non-language-related traits, there were no strong association signals either under the peaks or across the remaining genome. As mentioned in the Results section, the regions under the linkage peaks are not adequately tagged for comprehensive association analysis by the Axiom 1.0 array, as is the case for much of the genome, and this greatly decreases the chance of observing associations. The modest number of subjects also limits the power to detect common variants of small effect. Additionally, recent evidence suggests that autism has significant allelic heterogeneity, and association analyses have generally not been replicated except for variants with very small effects (42). Studies of rare variants indicate widespread heterogeneity (43–46), though it remains unclear what proportions of genetic mechanisms for ASD involve rare, infrequent, or common variants.

The issue of allelic heterogeneity across disorders is still uncharacterized from both theoretical and empirical points of view, but it is quite relevant to the debate on genetic overlap between disorders. Until reasonably inferred functional alleles are associated in each disorder, it is not possible to directly address the issue of whether the same genetic variants within the same gene are relevant to both specific language impairment and ASD (11, 47). Hence, it is possible that the origins of language impairment and ASD have some of the same key genes but not the same underlying variants or molecular mechanisms. However, if functional variants are the same for both disorders, then it remains to be explained why some members of the family develop specific language impairment and not ASD. It could be that individuals manifesting language impairment have a smaller genetic load for such variants or that language impairment is a truly dissociable subcomponent in at least some forms of ASD. We will continue susceptibility allele mapping in our pedigrees with autism and language impairment to disentangle these complicated mechanisms.

## References

1. Warburton P, Baird G, Chen W, Morris K, Jacobs BW, Hodgson S, Docherty Z: Support for linkage of autism and specific language impairment to 7q3 from two chromosome rearrangements involving band 7q31. Am J Med Genet 2000; 96:228–234

2. Bartlett CW, Flax JF, Logue MW, Smith BJ, Vieland VJ, Tallal P, Brzustowicz LM: Examination of potential overlap in autism and language loci on chromosomes 2, 7, and 13 in two independent samples ascertained for specific language impairment. Hum Hered 2004; 57:10–20

3. Bartlett CW, Flax JF, Logue MW, Vieland VJ, Bassett AS, Tallal P, Brzustowicz LM: A major susceptibility locus for specific language impairment is located on 13q21. Am J Hum Genet 2002; 71:45–55

4. Bradford Y, Haines J, Hutcheson H, Gardiner M, Braun T, Sheffield V, Cassavant T, Huang W, Wang K, Vieland V, Folstein S, Santangelo S, Piven J: Incorporating language phenotypes strengthens evidence of linkage to autism. Am J Med Genet 2001; 105:539–547

5. Buxbaum JD, Silverman JM, Smith CJ, Kilifarski M, Reichert J, Hollander E, Lawlor BA, Fitzgerald M, Greenberg DA, Davis KL: Evidence for a susceptibility gene for autism on chromosome 2 and for genetic heterogeneity. Am J Hum Genet 2001; 68: 1514–1520

6. Shao Y, Raiford KL, Wolpert CM, Cope HA, Ravan SA, Ashley-Koch AA, Abramson RK, Wright HH, DeLong RG, Gilbert JR, Cuccaro ML, Pericak-Vance MA: Phenotypic homogeneity provides increased support for linkage on chromosome 2 in autistic disorder. Am J Hum Genet 2002; 70:1058–1061

7. Simmons TR, Flax JF, Azaro MA, Hayter JE, Justice LM, Petrill SA, Bassett AS, Tallal P, Brzustowicz LM, Bartlett CW: Increasing genotype-phenotype model determinism: application to bivariate reading/language traits and epistatic interactions in language-impaired families. Hum Hered 2010; 70:232–244

8. Spence SJ, Cantor RM, Chung L, Kim S, Geschwind DH, Alarcon M: Stratification based on language-related endophenotypes in autism: attempt to replicate reported linkage. Am J Med Genet B Neuropsychiatr Genet 2006; 141B:591–598

9. Alarcón M, Yonan AL, Gilliam TC, Cantor RM, Geschwind DH: Quantitative genome scan and ordered-subsets analysis of autism endophenotypes support language QTLs. Mol Psychiatry 2005; 10:747–757

10. Alarcón M, Cantor RM, Liu J, Gilliam TC, Geschwind DH; Autism Genetic Research Exchange Consortium: Evidence for a language quantitative trait locus on chromosome 7q in multiplex autism families. Am J Hum Genet 2002; 70:60–71

11. Alarcón M, Abrahams BS, Stone JL, Duvall JA, Perederiy JV, Bomar JM, Sebat J, Wigler M, Martin CL, Ledbetter DH, Nelson SF, Cantor RM, Geschwind DH: Linkage, association, and gene-expression analyses identify CNTNAP2 as an autism-susceptibility gene. Am J Hum Genet 2008; 82:150–159

12. Lord C, Rutter M, Le Couteur A: Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. J Autism Dev Disord 1994; 24:659–685

13. Hou L, Wang K, Bartlett CW: Evaluation of a Bayesian model integration-based method for censored data. Hum Hered 2012; 74:1–11

14. Goodman WK, Price LH, Rasmussen SA, Mazure C, Fleischmann RL, Hill CL, Heninger GR, Charney DS: The Yale-Brown Obsessive Compulsive Scale, I: development, use, and reliability. Arch Gen Psychiatry 1989; 46:1006–1011

15. Goodman WK, Price LH, Rasmussen SA, Mazure C, Delgado P, Heninger GR, Charney DS: The Yale-Brown Obsessive Compulsive Scale, II: validity. Arch Gen Psychiatry 1989; 46:1012–1016

16. Scahill L, Riddle MA, McSwiggin-Hardin M, Ort SI, King RA, Goodman WK, Cicchetti D, Leckman JF: Children's Yale-Brown Obsessive Compulsive Scale: reliability and validity. J Am Acad Child Adolesc Psychiatry 1997; 36:844–852

17. Scahill L, McDougle CJ, Williams SK, Dimitropoulos A, Aman MG, McCracken JT, Tierney E, Arnold LE, Cronin P, Grados M, Ghuman J, Koenig K, Lam KS, McGough J, Posey DJ, Ritz L, Swiezy NB, Vitiello B; Research Units on Pediatric Psychopharmacology Autism Network: Children's Yale-Brown Obsessive Compulsive Scale modified for pervasive developmental disorders. J Am Acad Child Adolesc Psychiatry 2006; 45:1114–1123

18. Bartlett CW, Flax JF, Fermano Z, Hare A, Hou L, Petrill SA, Buyske S, Brzustowicz LM: Gene × gene interaction in shared etiology of autism and specific language impairment. Biol Psychiatry 2012; 72:692–699

19. De Fossé L, Hodge SM, Makris N, Kennedy DN, Caviness VS Jr, McGrath L, Steele S, Ziegler DA, Herbert MR, Frazier JA, Tager-Flusberg H, Harris GJ: Language-association cortex asymmetry in autism and specific language impairment. Ann Neurol 2004; 56:757–766

20. Lindgren KA, Folstein SE, Tomblin JB, Tager-Flusberg H: Language and reading abilities of children with autism spectrum disorders and specific language impairment and their first-degree relatives. Autism Res 2009; 2:22–38

21. Constantino JN, Todd RD: Intergenerational transmission of subthreshold autistic traits in the general population. Biol Psychiatry 2005; 57:655–660

22. Constantino JN, Davis SA, Todd RD, Schindler MK, Gross MM, Brophy SL, Metzger LM, Shoushtari CS, Splinter R, Reich W: Validation of a brief quantitative measure of autistic traits: comparison of the Social Responsiveness Scale with the Autism Diagnostic Interview-Revised. J Autism Dev Disord 2003; 33:427–433

23. Bruse S, Moreau M, Azaro M, Zimmerman R, Brzustowicz L: Improvements to bead-based oligonucleotide ligation SNP genotyping assays. Biotechniques 2008; 45:559–571

24. Hou L, Phillips C, Azaro M, Brzustowicz LM, Bartlett CW: Validation of a cost-efficient multi-purpose SNP panel for disease based research. PLoS ONE 2011; 6:e19699

25. Vieland VJ, Huang Y, Seok SC, Burian J, Catalyurek U, O'Connell J, Segre A, Valentine-Cooper W: KELVIN: a software package for rigorous measurement of statistical evidence in human genetics. Hum Hered 2011; 72:276–288

26. Huang Y, Vieland VJ: Association statistics under the PPL framework. Genet Epidemiol 2010; 34:835–845

27. Yang X, Huang J, Logue MW, Vieland VJ: The posterior probability of linkage allowing for linkage disequilibrium and a new estimate of disequilibrium between a trait and a marker. Hum Hered 2005; 59:210–219

28. Vieland VJ, Huang Y, Bartlett C, Davies TF, Tomer Y: A multilocus model of the genetic architecture of autoimmune thyroid disorder, with clinical implications. Am J Hum Genet 2008; 82:1349–1356

29. Matise TC, Chen F, Chen W, De La Vega FM, Hansen M, He C, Hyland FC, Kennedy GC, Kong X, Murray SS, Ziegle JS, Stewart WC, Buyske S: A second-generation combined linkage physical map of the human genome. Genome Res 2007; 17:1783–1786

30. Logue MW, Vieland VJ, Goedken RJ, Crowe RR: Bayesian analysis of a previously published genome screen for panic disorder reveals new and compelling evidence for linkage to chromosome 7. Am J Med Genet B Neuropsychiatr Genet 2003; 121B:95–99

31. Chagnon YC: Shared susceptibility region on chromosome 15 between autism and catatonia. Int Rev Neurobiol 2006; 72:165–178

32. Merritt JL 2nd, Jalal SM, Barbaresi WJ, Babovic-Vuksanovic D: 14q32.3 deletion syndrome with autism. Am J Med Genet A 2005; 133A:99–100

33. Qiao Y, Tyson C, Hrynchak M, Lopez-Rangel E, Hildebrand J, Martell S, Fawcett C, Kasmara L, Calli K, Harvard C, Liu X, Holden JJ, Lewis SM, Rajcan-Separovic E: Clinical application of 2.7M Cytogenetics array for CNV detection in subjects with idiopathic autism and/or intellectual disability. Clin Genet 2013; 83:145–154

34. Bolte S, Poustka F, Constantino JN: Assessing autistic traits: cross-cultural validation of the Social Responsiveness Scale (SRS). Autism Res 2008; 1:354–363

35. Constantino JN, Gruber CP, Davis S, Hayes S, Passanante N, Przybeck T: The factor structure of autistic traits. J Child Psychol Psychiatry 2004; 45:719–726

36. Kamio Y, Inada N, Moriwaki A, Kuroda M, Koyama T, Tsujii H, Kawakubo Y, Kuwabara H, Tsuchiya KJ, Uno Y, Constantino JN: Quantitative autistic traits ascertained in a national survey of 22 529 Japanese schoolchildren. Acta Psychiatr Scand 2013; 128: 45–53

37. Wigham S, McConachie H, Tandos J, Le Couteur AS; Gateshead Millennium Study core team: The reliability and validity of the Social Responsiveness Scale in a UK general child population. Res Dev Disabil 2012; 33:944–950

38. Coon H, Villalobos ME, Robison RJ, Camp NJ, Cannon DS, Allen-Brady K, Miller JS, McMahon WM: Genome-wide linkage using the Social Responsiveness Scale in Utah autism pedigrees. Mol Autism 2010; 1:8

39. Duvall JA, Lu A, Cantor RM, Todd RD, Constantino JN, Geschwind DH: A quantitative trait locus analysis of social responsiveness in multiplex autism families. Am J Psychiatry 2007; 164:656–662

40. Barrett S, Beck JC, Bernier R, Bisson E, Braun TA, Casavant TL, Childress D, Folstein SE, Garcia M, Gardiner MB, Gilman S, Haines JL, Hopkins K, Landa R, Meyer NH, Mullane JA, Nishimura DY, Palmer P, Piven J, Purdy J, Santangelo SL, Searby C, Sheffield V, Singleton J, Slager S, Struchen T, Svenson S, Vieland V, Wang K, Winklosky B: An autosomal genomic screen for autism: collaborative linkage study of autism. Am J Med Genet 1999; 88: 609–615; correction, 2001; 105:805

41. Steele MM, Al-Adeimi M, Siu VM, Fan YS: Brief report: a case of autism with interstitial deletion of chromosome 13. J Autism Dev Disord 2001; 31:231–234

42. Anney R, Klei L, Pinto D, Almeida J, Bacchelli E, Baird G, Bolshakova N, Bölte S, Bolton PF, Bourgeron T, Brennan S, Brian J, Casey J, Conroy J, Correia C, Corsello C, Crawford EL, de Jonge M, Delorme R, Duketis E, Duque F, Estes A, Farrar P, Fernandez BA, Folstein SE, Fombonne E, Gilbert J, Gillberg C, Glessner JT, Green A,

Green J, Guter SJ, Heron EA, Holt R, Howe JL, Hughes G, Hus V, Igliozzi R, Jacob S, Kenny GP, Kim C, Kolevzon A, Kustanovich V, Lajonchere CM, Lamb JA, Law-Smith M, Leboyer M, Le Couteur A, Leventhal BL, Liu XQ, Lombard F, Lord C, Lotspeich L, Lund SC, Magalhaes TR, Mantoulan C, McDougle CJ, Melhem NM, Merikangas A, Minshew NJ, Mirza GK, Munson J, Noakes C, Nygren G, Papanikolaou K, Pagnamenta AT, Parrini B, Paton T, Pickles A, Posey DJ, Poustka F, Ragoussis J, Regan R, Roberts W, Roeder K, Roge B, Rutter ML, Schlitt S, Shah N, Sheffield VC, Soorya L, Sousa I, Stoppioni V, Sykes N, Tancredi R, Thompson AP, Thomson S, Tryfon A, Tsiantis J, Van Engeland H, Vincent JB, Volkmar F, Vorstman JA, Wallace S, Wing K, Wittemeyer K, Wood S, Zurawiecki D, Zwaigenbaum L, Bailey AJ, Battaglia A, Cantor RM, Coon H, Cuccaro ML, Dawson G, Ennis S, Freitag CM, Geschwind DH, Haines JL, Klauck SM, McMahon WM, Maestrini E, Miller J, Monaco AP, Nelson SF, Nurnberger JI Jr, Oliveira G, Parr JR, Pericak-Vance MA, Piven J, Schellenberg GD, Scherer SW, Vicente AM, Wassink TH, Wijsman EM, Betancur C, Buxbaum JD, Cook EH, Gallagher L, Gill M, Hallmayer J, Paterson AD, Sutcliffe JS, Szatmari P, Vieland VJ, Hakonarson H, Devlin B: Individual common variants exert weak effects on the risk for autism spectrum disorders. Hum Mol Genet 2012; 21:4781–4792

43. Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J, Yamrom B, Lee YH, Narzisi G, Leotta A, Kendall J, Grabowska E, Ma B, Marks S, Rodgers L, Stepansky A, Troge J, Andrews P, Bekritsky M, Pradhan K, Ghiban E, Kramer M, Parla J, Demeter R, Fulton LL, Fulton RS, Magrini VJ, Ye K, Darnell JC, Darnell RB, Mardis ER, Wilson RK, Schatz MC, McCombie WR, Wigler M: De novo gene disruptions in children on the autistic spectrum. Neuron 2012; 74:285–299

44. Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A, Lin CF, Stevens C, Wang LS, Makarov V, Polak P, Yoon S, Maguire J, Crawford EL, Campbell NG, Geller ET, Valladares O, Schafer C, Liu H, Zhao T, Cai G, Lihm J, Dannenfelser R, Jabado O, Peralta Z, Nagaswamy U, Muzny D, Reid JG, Newsham I, Wu Y, Lewis L, Han Y, Voight BF, Lim E, Rossin E, Kirby A, Flannick J, Fromer M, Shakir K, Fennell T, Garimella K, Banks E, Poplin R, Gabriel S, DePristo M, Wimbish JR, Boone BE, Levy SE, Betancur C, Sunyaev S, Boerwinkle E, Buxbaum JD, Cook EH Jr, Devlin B, Gibbs RA, Roeder K, Schellenberg GD, Sutcliffe JS, Daly MJ: Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature 2012; 485:242–245

45. O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, Levy R, Ko A, Lee C, Smith JD, Turner EH, Stanaway IB, Vernot B, Malig M, Baker C, Reilly B, Akey JM, Borenstein E, Rieder MJ, Nickerson DA, Bernier R, Shendure J, Eichler EE: Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature 2012; 485:246–250

46. Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, Ercan-Sencicek AG, DiLullo NM, Parikshak NN, Stein JL, Walker MF, Ober GT, Teran NA, Song Y, El-Fishawy P, Murtha RC, Choi M, Overton JD, Bjornson RD, Carriero NJ, Meyer KA, Bilguvar K, Mane SM, Sestan N, Lifton RP, Günel M, Roeder K, Geschwind DH, Devlin B, State MW: De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature 2012; 485:237–241

47. Chapman NH, Estes A, Munson J, Bernier R, Webb SJ, Rothstein JH, Minshew NJ, Dawson G, Schellenberg GD, Wijsman EM: Genome-scan for IQ discrepancy in autism: evidence for loci on chromosomes 10 and 16. Hum Genet 2011; 129:59–70