Molecular Psychiatry (2005) 10, 747–757

www.nature.com/mp

ORIGINAL RESEARCH ARTICLE

Quantitative genome scan and Ordered-Subsets Analysis of autism endophenotypes support language QTLs

M Alarcón[1], AL Yonan[2], TC Gilliam[2,3], RM Cantor[4] and DH Geschwind[1]

[1]*Department of Neurology, UCLA School of Medicine, Center for Neurobehavioral Genetics and Neuropsychiatric Research Institute, Los Angeles, CA, USA;* [2]*Columbia Genome Center, Department of Genetics and Development, New York, NY, USA;* [3]*Department of Psychiatry, Columbia University, New York, NY, USA;* [4]*Departments of Human Genetics and Pediatrics, UCLA School of Medicine, Los Angeles, CA, USA*

Autism is a neurodevelopmental syndrome with early childhood onset and deficits in three behavioral and cognitive dimensions: language, social skills and repetitive or restrictive behaviors. We hypothesized that using these endophenotypes would provide more power to detect linkage than the diagnosis of autism. Previously, we reported results for a nonparametric quantitative trait locus (QTL) genome scan in 152 families with autism, which revealed a linkage peak related to spoken language on 7q35. Here, we present the results of a nonparametric QTL scan of autism endophenotypes in 291 multiplex families, including the original 152. The strongest evidence for an 'age at first word' QTL was on chromosomes 3q at 147 cM ($Z = 3.10$, $P < 0.001$), and 17q at 93 cM ($Z = 2.84$, $P = 0.002$), both represent novel susceptibility loci for autism endophenotypes. There was also support for a previously identified autism peak on chromosome 17 at 43 cM ($Z = 2.22$, $P = 0.013$) with 'age at first phrase'. The 7q35 language peak was attenuated ($Z = 2.05$, $P = 0.02$) compared with the original finding. To explore the possibility of increased heterogeneity resulting from the addition of 135 families to the sample, we conducted an Ordered-Subsets Analysis on chromosome 7; these results suggest that the 132 autism families with the earliest average age at first word are responsible for the QTL on 7q35. This locus on 7q35 may harbor a gene contributing variability in spoken language that is not uniquely related to language delay in autism.
Molecular Psychiatry (2005) **10,** 747–757. doi:10.1038/sj.mp.4001666; published online 12 April 2005

Keywords: autism spectrum disorder; language; linkage; nonparametric; sibpair

Autism (MIM 209850) is one of the most common neurodevelopmental disorders and is characterized by language impairments, social and communicative deficits, and repetitive or stereotyped behaviors (RSB) or activities. It has an onset prior to 3 years of age and, by most recent estimates, affects more than three in 1000 individuals.[1] Evidence for the role of genes in the development of autism includes high heritability estimates,[2] a high sibling risk compared to the general population,[3] and chromosomal anomalies reported in individuals with the disorder.[4] Overlapping regions of suggestive linkage have been reported in more than two studies on chromosomes 1, 2q, 3q, 7q, 16p, 17q and Xq,[5–11] but no genes have been conclusively identified yet.[12]

Several factors, such as phenotypic and genetic heterogeneity, and multiple interacting genes, have been postulated to complicate the identification of autism loci leading to inconsistent results. The phenotypic expression of autism varies greatly among the affected.[13] Mild variations of the main characteristics or endophenotypes[14] of autism, such as social or communicative impairments, RSB and, especially, language deficits, are also often observed in the first- and second-degree relatives of autistic probands.[15] The presence of these autism-related traits in the probands' relatives, irrespective of their own affection status, suggests that the traits are also familial and may be genetically transmitted.

Endophenotypes represent simple, biologically based aspects of a disease that may be governed by fewer susceptibility genes. To optimize the search for these genes, endophenotypes must be clearly associated with the disease, must be heritable, and are often observed in unaffected relatives at a higher rate than in the general population.[16] Although the concept of psychiatric endophenotypes was described over 30 years ago,[17] their use in genetic studies of autism has gained popularity only recently. Investigators have begun to use these autism endophenotypes as covariates in linkage analyses in an attempt to increase both phenotypic and genetic

Correspondence: DH Geschwind, MD, PhD and M Alarcón, PhD, Department of Neurology, UCLA School of Medicine, Center for Neurobehavioral Genetics and Neuropsychiatric Research Institute, 710 Westwood Plaza, RNRC 1-145, Los Angeles, CA 90095, USA. E-mail: dhg@ucla.edu, malarcon@ucla.edu
Received 22 September 2004; revised 10 January 2005; accepted 14 February 2005

homogeneity of the affected sample.[18–21] For example, Buxbaum *et al*[19] reported that most of the linkage evidence for a putative autism susceptibility locus on chromosome 2q was from the families with a language delay. These results were supported in an independent sample of 82 families with autism.[21] Bradford *et al*[18] investigated two chromosomes, 7 and 13, and suggested that language-delayed families were responsible for most of the evidence for linkage to autism to these chromosomes. Shao and colleagues[20] applied the Ordered-Subsets Analysis method to a covariate representing 'insistence on sameness' derived from a principal components analysis, and identified a subset of homogeneous families with autism that were responsible for linkage to a previously reported region on 15q11–13.

As an alternative approach in the search for autism susceptibility genes, we defined several familial endophenotypes from the Autism Diagnostic Interview-Revised (ADI-R)[22] and used them to perform a genomewide quantitative trait locus (QTL) analysis.[23] Several suggestive peaks were identified including the most significant language-related peak on 7q35. Here we continue this approach, performing a nonparametric QTL analysis of familial autism endophenotypes in this recently expanded sibpair sample. Results of the present quantitative analysis, which included linkage peaks on chromosomes 3, 17 and a more modest peak on 7q35, led us to attempt to investigate the heterogeneity in our sample through the use of the Ordered-Subsets Analysis approach.[24,25]

Materials and methods

Participants

The AGRE sample was ascertained for nuclear families with at least two children having a possible diagnosis of autism spectrum disorder (autism, pervasive developmental disorder or Asperger's syndrome). Families were recruited through physician referrals and the Cure Autism Now foundation, a nonprofit organization established by parents, clinicians and researchers to fund biomedical research of autism. AGRE has human subjects' approval and our use of the AGRE repository was approved by the UCLA and Columbia University IRBs.

For the present report, phenotypic information was available from 436 AGRE families; genotypic and phenotypic data were publicly available from the AGRE website[26] for a total of 345 families. The total number of individuals in the sample was 1794, and included 1034 offspring. Among the 701 individuals affected with autism spectrum, the ratio of males to females was 3.25 (536 boys and 165 girls), consistent with other reports of an increased prevalence of the disorder in boys.[5,7,10] The mean age of the children at the time of testing was 7.3 years. Cases with confirmed Fragile X syndrome or other chromosomal abnormalities were excluded from the analysis. In addition, the specific age at first word was not available for 115 children due to various reasons: 18 had a few words and then lost them, 78 had not reached the milestone of having said their first word and for the rest the value was unknown. However, we did not expect that these missing data would bias the results. The families from the present report were included in a recent genome scan of the diagnosis of autism.[27] However, Yonan and colleagues' qualitative approach only required the individuals' diagnosis to be included in the analysis; thus, affected individuals with missing quantitative variables were included in their report and can not be included here.

Measures

The ADI-R[22] is a semistructured interview for the caregivers of individuals that may be affected with autism spectrum disorders. A trained tester administered the ADI-R, which is based on the ICD-10 and DSM-IV criteria for autism diagnosis, in the family home. We analyzed three items from the ADI-R that quantified the age at which the subjects spoke their first word (WORD), their age at speaking their first phrase (PHRASE) and a composite measure of RSB: ADI variables A12, A13 and DD total, respectively.[23] Children from 291 families had no missing data for at least one of these three items from the ADI-R. The majority of the families ($N=258$) had sibships of size two, 30 families had sibships of size three and three families had sibships of size four. Data from a total of 618 offspring were included in the quantitative linkage analysis. From this sample, WORD, PHRASE and RSB items were available for 515, 430 and 618 individuals from multiplex families, respectively. The families used for this analysis are available as supplemental data (geschwindlab.medsch.ucla.edu).

For the OSA, individuals with a broad diagnosis of autism were considered affected. Individuals were considered affected with narrow autism if they had an onset prior to 3 years of age and met the criteria in quality of social interaction, communication and language, and repetitive, restricted and stereotyped interests and behavior.[22] In the genotyped sample, the broad autism group included individuals diagnosed with formal autism based on the ADI-R ($N=505$), individuals categorized as Not Quite Autism (NQA, one point short of meeting the autism criteria; $N=40$), and those who are in the Spectrum (individuals show either (a) severe deficit on at least one domain or (b) moderate deficits in at least two domains; or (c) minimal deficits in all three domains; $N=96$). Thus, the sample available for the OSA includes genotyped individuals with a variable expression of the disorder: from mild to severe impairment.

Families with at least two genotyped offspring with broad autism with at least one of these offspring measured for the quantitative traits were included in the OSA. The OSA covariates were the means, minimum values and maximum values for the language-related traits—WORD and PHRASE. Since the covariates represent an entire family and not an individual, one estimate was used per family using the available data from all siblings within that family.

The OSA WORD and PHRASE analyses included 289 and 273 families, respectively.

Genotyping

The laboratory and genotyping protocols for the AGRE study have been described previously.[9] DNA samples from the parents and offspring were genotyped at Columbia University using 335 microsatellite markers comprising a modified version of the Weber 8.0 marker set. An additional 73 microsatellite markers were genotyped to follow-up regions of interest identified through a qualitative analysis of 156 families (results reported in Alarcón *et al*[23] and Yonan *et al*[27]). Of these 73 follow-up markers, 30 span a region of about 60 cM on chromosome 7q. PCR amplification of microsatellite markers has been described previously.[9,28] The average heterozygosity of markers used in this study was 0.77 and the average density was 10 cM.

Statistical and genetic analysis

The SAS software (version 8; SAS Institute, Cary, NC, USA) was used to merge data files, calculate descriptive statistics and to prepare the input files for the genetic analyses. PedCheck[29] was used to find Mendelian genotype errors. When a genotype error was identified, the genotype was recoded as missing. Marker allele frequencies were obtained by counting parental genotypes; however, 90% of the parents were genotyped and therefore frequencies would not have a major impact on linkage results. Map distances were obtained from the Center for Medical Genetics, Marshfield Medical Research Foundation (http://research.marshfieldclinic.org/genetics/).

We applied a nonparametric linkage approach to identify QTL that may contribute to the expression of the autism endophenotypes of WORD, PHRASE and RSB. Linkage analysis was performed using the program Mapmaker/Sibs[30] within the Genehunter 2.1 software package[31] that conducts multipoint sibling-pair linkage analysis of quantitative traits. All pairs of affected sibs were used in the analysis, that is, option 3 includes all possible pairs within a sibship and weighs them by $2/N$ (where N is the number of affecteds) to account for the correlation among pairs from the same family.

The nonparametric QTL statistic is based on the Wilcoxon rank-sum test. Sib pair trait differences are ranked and multiplied by a function of the number of alleles shared IBD; the ratio of the statistic to its standard deviation provides a Z-score that follows a standard normal distribution.[30,32] This method is robust to violations of the normality assumption.[30]

Since Genehunter does not have an analysis program to test X-linkage on quantitative traits, we used Mapmaker/Sibs on the discrete trait of 'affection status'. We selected those individuals whose trait value exceeded the mean for their sex and classified them as 'affected'. LOD scores were reported for brother–brother, brother–sister and sister–sister pairs on a single multipoint plot for each trait. In this report, we highlighted linkage regions with nominal P-values of <0.01, and if additional traits show evidence for linkage with a P-value of <0.05 in the same region, we also presented those results.

Follow-up analyses explored the possibility that increased heterogeneity in our large sample reduced the original QTL signal on chromosome 7q35. Although the possibility exists that the original signal was simply a false positive, the results from independent studies supporting the presence of susceptibility genes for autism[33–35] and for language disorders[36] within 20 cM of our QTL signal suggested that an alternative explanation existed for the attenuation of the linkage peak. To evaluate this, we applied an OSA approach[24,25] to the data from chromosome 7. The OSA method consists of performing a qualitative linkage analysis to obtain a linkage statistic for each family. Then, the families are ranked according to the covariate of interest and the proportion of families responsible for the linkage signal is estimated by adding the families one to several at a time and running the qualitative linkage analysis. This procedure is repeated until maximal evidence for linkage is obtained; then, empirical significance values are estimated. Since our original QTL result for chromosome 7q was linked to a language trait, WORD and PHRASE were used in these analyses as covariates for ranking the families. The sib averages of these traits, the minimum trait values within a sibship and the maximum trait values within a sibship were used as the covariates for the OSA. Since we did not make *a priori* assumptions whether the families with the earliest or the most delayed language development were responsible for the linkage result, we ranked the families in both ascending and descending order by each of their covariate values, resulting in ordering the sibships in 12 ways.

Results

Sib correlations of language and repetitive behavior measures

Sibling correlations for the three ADI-R items that assessed language and repetitive behavior were significant and are consistent with those reported previously in a subset of the sample.[23] As shown in Table 1, the modest but significant WORD ($r=0.27$), PHRASE ($r=0.30$) and RSB ($r=0.31$) sibling correlations suggested that the variability in these traits was familial and may have been due to genetic influences; however, the effects of a common sibling environment may have also contributed to these relationships.[37]

Trait distributions

Children with normal language development typically say their first word prior to 12 months of age and their first phrase by 24 months. In this sample, the age at first word ranged from 5 to 96 months and the first phrase occurred between the ages of 9 and 114 months. While the majority of the offspring in the AGRE sample shows either a word or phrase

Table 1 Sibling correlations of autism endophenotypes

Endophenotype	*ADI-R*[a] *Item*	*N Sibpairs*	*Correlation*[b]	*P-value*
Age at first word	A12	202	0.27	<0.0001
Age at first phrase	A13	146	0.30	0.003
Repetitive or stereotyped behavior	DD total	286	0.31	<0.0001

[a]ADI-R.[22]
[b]Spearman correlation coefficients.

Table 2 Descriptives for age at first word (WORD), age at first phrase (PHRASE) and repetitive or stereotyped behavior (RSB) stratified by sex

Measure	*N*	*Mean*[a]	*SD*	*Median*[a]	*Range*[a]	*Kurtosis*	*Skewness*
WORD[b]							
Males	418	30.75	18.77	27	5–96	0.18	0.86[c]
Females	130	26.96	15.66	24	8–72	−0.53	0.73[c]
PHRASE							
Males	348	43.81	19.10	42	9–114	0.45	0.61[c]
Females	110	41.35	18.96	39	12–96	−0.21	0.50[c]
RSB[b]							
Males	505	5.60	2.63	6	0–12	−0.18	0.17[c]
Females	153	4.79	2.50	5	0–12	0.07	0.41

[a]In months for WORD and PHRASE.
[b]Males and females are significantly different ($P<0.05$).
[c]Significantly non-normal ($P<0.05$).

speech delay, many also met normal language milestones; this illustrates the wide variability in expression of language-related endophenotypes present in Autism.

As shown in Table 2, males had a significantly greater delay in WORD and more RSB symptoms. WORD and PHRASE exhibited significant positive skewing irrespective of sex, although they were not kurtotic (ie the distributions did not have abnormally large or small tails). These results support the use of nonparametric genetic analyses that do not assume distributional normality. Thus, all genetic analyses made no distributional assumptions and all analyses were run using the untransformed scores.

Genome scan of language and repetitive behavior-related QTL

A complete quantitative multipoint genome scan analysis for autism endophenotypes in nuclear families having at least two children affected with autism spectrum disorder was performed. Although only phenotypic data from affected siblings were included in these analyses, the parental genotypes permitted more precise estimation of allele sharing (ie IBD). The nonparametric genome-wide QTL results for the three quantitative traits are shown in Figures 1 and 2, and are summarized in Table 3. As shown (Table 3), 7 chromosomes had evidence for nominal linkage at the 1% level (ie, $Z>2.18$, nominal $P\leq0.01$). The largest Z-score was obtained for WORD at 147 cM on chromosome 3 between markers D3S3045 and D3S1763: $Z=3.10$, one-sided $P<0.001$ (Figure 1). A smaller peak for PHRASE ($Z=1.76$, nominal $P=0.04$) at 170 cM slightly overlaps the linked region for WORD.

Another suggestive linkage peak for WORD was observed on chromosome 17q at 93 cM ($Z=2.84$, nominal $P=0.002$) between markers D17S1290 and D17S1301. RSB showed a similar result ($Z=2.31$, nominal $P=0.01$) at the same position. Results from Auranen *et al*[35] also support a QTL on 17q, although their peak location was 23 cM telomeric from the current peak. However, these results are not directly comparable because Auranen *et al*[35] analyzed markers that were more telomeric to the currently available scan markers. Interestingly, one of the most significant linkage peaks for PHRASE was on this chromosome, at 43 cM, between markers D17S1298 and D17S1299. This result overlaps with several reports of autism linkage between 35 and 49 cM.[8,10,11,27] Adding four additional markers in this 17q region (between 48 and 53 cM), increased the peak Z-score for PHRASE to almost 3.0.

The largest Z-score for RSB was on the short arm of chromosome 16 at 8 cM ($Z=2.5$, $P=0.006$), and there may be a QTL for language (WORD $Z=2.38$, $P=0.009$,

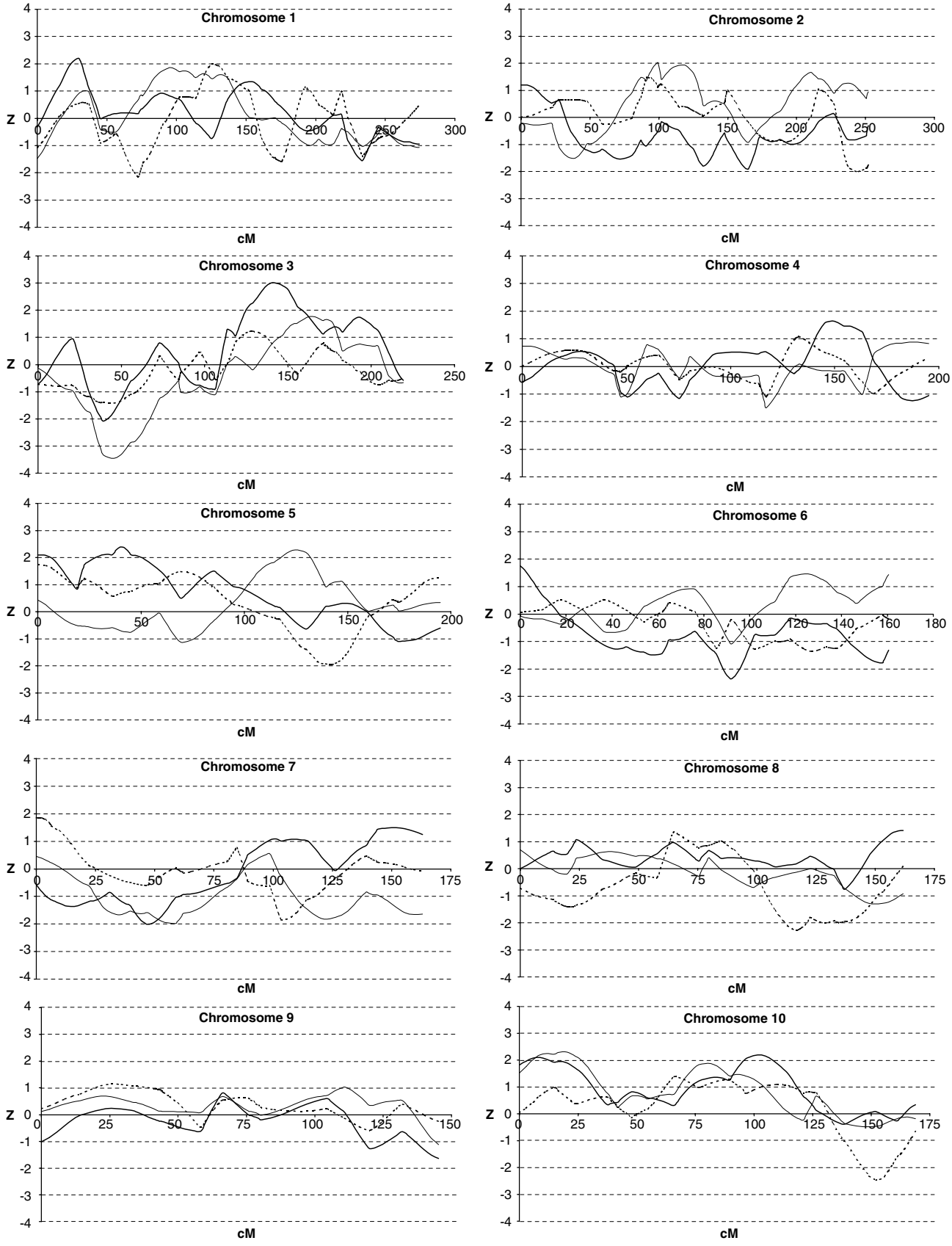

Figure 1 Nonparametric QTL results of an autosomal genome scan for age at first word (bold line), age at first phrase (solid line) and the composite measure of repetitive behaviors (dotted line).

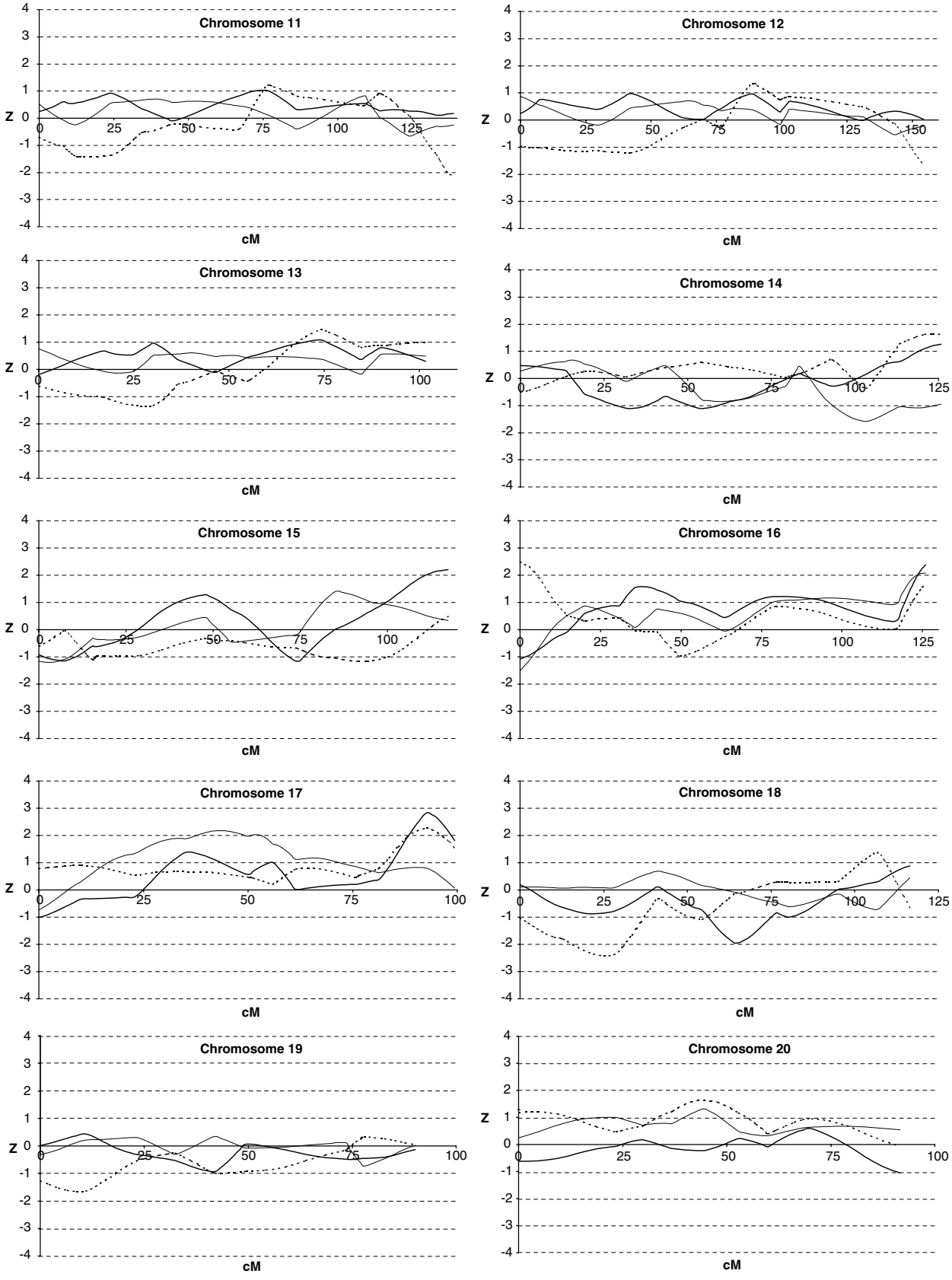

Figure 1 Continued.

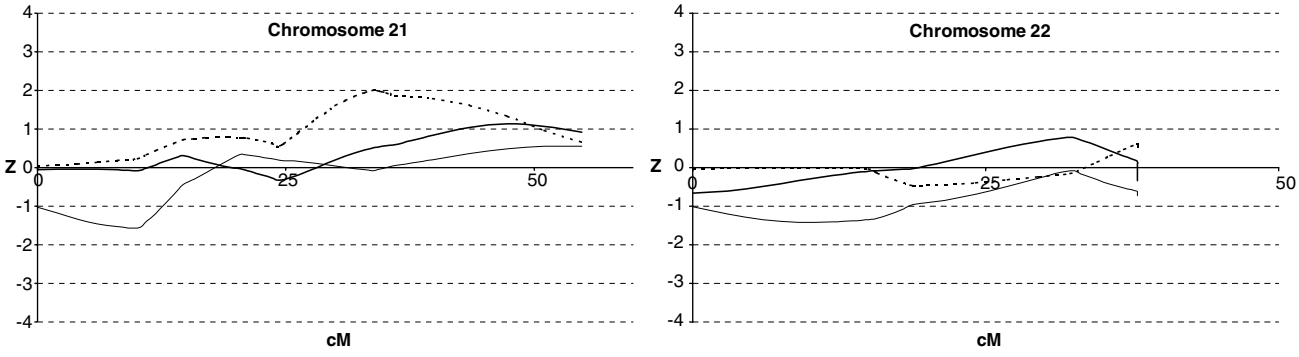

Figure 1 Continued.

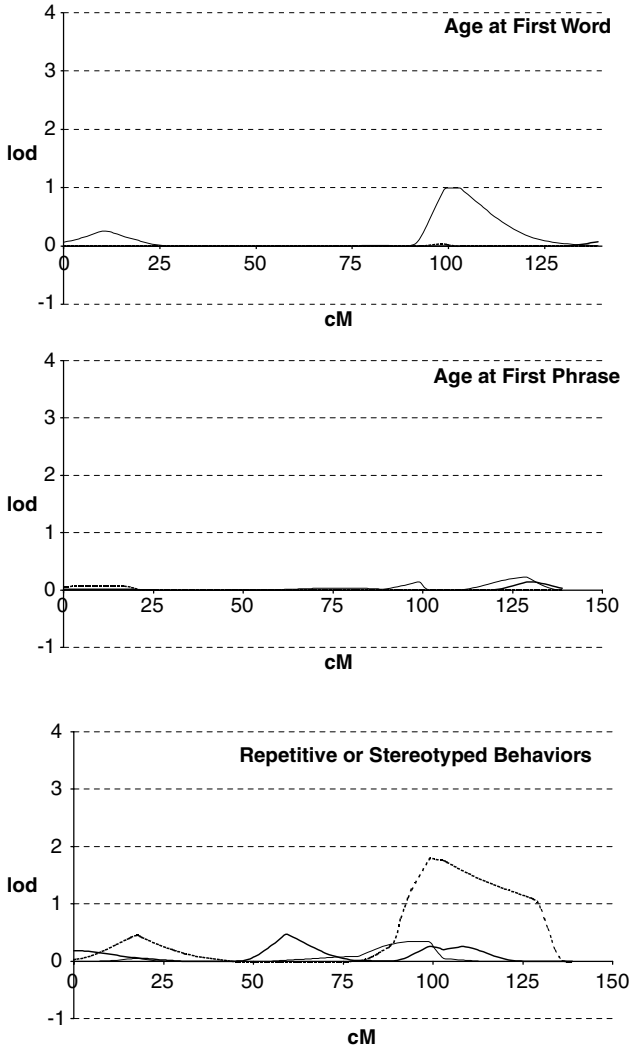

Figure 2 Nonparametric X-linkage results for age at first word, age at first phrase and the composite measure of repetitive behaviors for brother–brother (bold line), brother–sister (solid line) and sister–sister (dotted line) pairs.

134 cM; PHRASE $Z = 2.08$, $P = 0.019$, 134 cM) on the long arm. Existence of a language QTL on 16q is supported by a report of linkage of specific language impairment to a region (between 92 and 126 cM;

Table 3 Summary of multipoint quantitative genome scan in 287 families in chromosome order of significance

Chr	*Peak (cM)*	*Region*[a] *(cM)*	*Z*	*P-value*[b]	*Trait*[c]
3	147	126–170	3.10	<0.01	WORD
17	45	13–96	2.22	0.01	PHRASE
	93	86	2.84	<0.01	WORD
	93	78	2.31	0.01	RSB
16	8	0–20	2.5	0.01	RSB
	134	127	2.38	0.01	WORD
	134	78	2.08	0.02	PHRASE
5	40	0–67	2.39	0.01	WORD
	125	98–152	2.28	0.01	PHRASE
10	23	0–43	2.31	0.01	PHRASE
	107	72–126	2.19	0.01	WORD
1	30	9–41	2.20	0.01	WORD
15	122	98	2.20	0.01	WORD

[a]Based on the 90% confidence interval.
[b]Chromosomal regions included if nominal *P*-value was ≤ 0.01; additional peaks included if *P*-value was < 0.05.
[c]WORD = age at first word; PHRASE = age at first phrase; RSB = restrictive or repetitive behaviors.

Consortium[38]) proximal to the current linkage interval (78 cM to the end of the chromosome).

Language-related QTL on 7q and heterogeneity analysis

The peak *Z*-score for WORD on chromosome 7 in this updated sample was smaller than the one in the original scan result ($Z = 2.98$, $P = 0.001$, 150–174 cM; Alarcón *et al*[23]), however, the location of the QTL for WORD on 7q35 remained the same. One possibility was that the original peak was a false positive. However, linkage studies of autism[33–35] and language disorders[36] provide evidence for susceptibility genes in this region. Moreover, results of a regional meta-analysis of four published autism genome scans show that linkage to the chromosome 7q region is significant at the genomewide level.[39] An alternative explanation for the reduced evidence for the 7q

language-related QTL is an increase in genetic heterogeneity in the additional 135 families. To explore this possibility, we applied an OSA approach to the chromosome 7 scan data for the language-related traits of WORD and PHRASE. The OSA allows the identification of a subset of families that is responsible for the maximal evidence for linkage to a region in the presence of heterogeneity; without the requirement of the prior specification of the subset.[24] In this way, OSA selects a subset of families with linkage homogeneity that can later be characterized. In the present report, the autism families were ranked according to scores for language endophenotypes and the subset of families that provided the linkage evidence to chromosome 7 was identified; a summary of the significant OSA results for chromosome 7 is shown in Table 4. The peak linkage results for WORD and PHRASE were at 163 cM, and in both cases, the subsets of families that contributed to the linkage evidence had offspring with the earliest language development compared to the overall autism sample. These results were not surprising given that WORD and PHRASE are correlated traits ($r = 0.66$, $P < 0.001$) and that there was a 78% overlap in the subsets of families contributing to the OSA linkage peaks for both traits. That is, when PHRASE was covaried, 52 of the 67 families that were responsible for the chromosome 7 linkage were also in the subset of families responsible for the linkage when WORD was covaried. The substantial increase in the LOD score for WORD was obtained using the ascending sib average to rank the families and 132 of them were responsible for the change. Interestingly, only 70 of these families were included in the original QTL report. Thus, the newest families in the updated sample also provide evidence for a language QTL on 7q.

To assess whether the offspring in the subset of families responsible for the chromosome 7q linkage differed in language, cognitive or behavioral measures as opposed to those in the unlinked subset of families, we compared their scores on WORD, PHRASE and RSB from the ADI-R, subtests from the Vineland Adaptive Behavior Scales, and the summary scores for the Peabody Picture Vocabulary Test (PPVT) and the Ravens Colored Progressive Matrices. The Vineland evaluates the socialization, daily living and communication skills of the affected individuals. The PPVT measures receptive vocabulary and verbal ability; and the Ravens is a test of nonverbal ability and measures the capacity to recognize geometric patterns and designs. To avoid assumptions of distributional normality, we used the Wilcoxon rank-sum test for the comparisons. As expected, the sibships responsible for the chromosome 7q35 linkage had significantly lower averages ($P < 0.0001$) of WORD (16.5 months) and PHRASE (35.2 months) than the unlinked sibships (40.3 and 49.6 months, respectively). With the exception of the PPVT, all other comparisons of language, cognitive or behavioral measures were not significantly different. As shown in Table 5, the two groups only differed marginally in the PPVT ($P < 0.04$) with the linked group having a slightly higher mean vocabulary score (96.46) than the unlinked group (81.80). Interestingly, the PPVT, a test of receptive language, was negatively correlated with WORD and PHRASE in the complete sample

Table 4 OSA summary—increases in LOD score due to selection of family subsets based on language traits and linkage to chromosome 7

Trait[a]	*Peak (cM)*	*Max LOD*	*Δ*	*P-value*[b]	*No. of families*	*Order*
WORD						
Sib average	163	2.57	2.13	0.035	132	Ascending
PHRASE						
Minimum	163	2.76	2.11	0.022	67	Ascending

[a]WORD = age at first word; PHRASE = age at first phrase.
[b]Strength of the relationship between the covariate and the linkage information for a subset of families.[24]

Table 5 Comparison of families responsible for chromosome 7 linkage with unlinked families

Measure	*Families linked to Chr 7*		*Families unlinked to Chr 7*		
	Mean[a]	*N*	*Mean*[a]	*N*	*P-value*[c]
Age at first word	16.5	132	40.3	173	<0.0001
Age at first phrase	35.2	117	49.6	165	<0.0001
Peabody Picture Vocabulary[b]	96.5	67	81.8	80	0.04

[a]Mean calculated within each sibship, regardless of number of siblings.
[b]Raw scores.
[c]Wilcoxon two-sample test, normal approximation, two-sided probability.

(Spearman's $r=-0.25$, $P<0.005$ and $r=-0.31$, $P<0.005$, respectively). These correlations between a receptive language assessment, PPVT and the ADI-R questionnaire items WORD and PHRASE were reassuring and suggested that the retrospective interview data are valid assessments of items on psychometric language tests.

Lastly, to assess whether our most significant linkage result on chromosome 3 was independent of linkage to chromosome 7, we tested linkage of the quantitative trait WORD to chromosome 3 in the following two subsets identified through the OSA: the 132 families linked to chromosome 7 and the 155 unlinked families. Irrespective of linkage to chromosome 7, there was modest evidence of linkage to chromosome 3: for the families linked to chromosome 7, $Z=1.64$, $P=0.101$, 112–192 cM (peak = 145 cM), and for the unlinked families, $Z=2.50$, $P=0.012$, 124–169 cM (peak = 153 cM). These results provide additional evidence for heterogeneity in autism and support the existence of at least two QTL for autism-related traits on chromosomes 3 and 7 that may contribute to language development independently.

Discussion

In the first report of a quantitative genome scan of autism, we described a potential language-related QTL on chromosome 7q.[23] Since that time, the AGRE sample has expanded and phenotypic information is now available for 436 families; 291 of these have quantitative traits and have also been genotyped and were included in our updated genome scan. By 2005, AGRE is expected to have genotypes from an independent sample of at least 100 additional autism families that will serve to test putative QTLs identified in this scan and the quantitative scan reported by Yonan and colleagues.[27]

As in the original report, nonparametric linkage analysis of three autism endophenotypes (WORD, PHRASE and RSB) revealed several regions that may harbor QTL for these traits. The most significant result was obtained for WORD on chromosome 3. As shown in Figure 1 and Table 3, the peak nonparametric *Z*-score for WORD was 3.10 ($P<0.001$) at 147 cM, suggesting that a language-related susceptibility locus or QTL may reside on the long arm of chromosome 3. Although Auranen *et al*[35] also obtained a peak on chromosome 3q, the distance between the peaks is quite large, around 43 cM, suggesting that the peaks are independent. Interestingly, Fisher and colleagues[40] reported a modest linkage peak of ADHD only 22 cM away from the current 3q WORD peak. Moreover, there is evidence for a QTL for phonological memory in families with speech-sound disorder[41] that is about 9 cM from the 3q region linked to WORD in the present autism sample. These results suggest that a general QTL for susceptibility to language-related developmental disorders may reside in this 3q region. Fine mapping in this region is currently underway for the AGRE sample.

The current linkage peak for WORD on chromosome 7 was attenuated compared to the results from the original report. The 7q linkage result, coupled with literature support for an autism susceptibility gene within 15 cM of the original QTL region,[5,7,10,11,34] prompted us to investigate the possibility of increased heterogeneity in the enlarged sample. Based on the novel work by Shao *et al*,[20] who described the use of phenotypic subtypes to identify homogeneous subsets of autism families, we applied the OSA approach to the language traits and obtained encouraging results. First, the linkage statistic for the qualitative analysis increased substantially by ranking the families according to the language covariates. Without ranking the families according to their language traits, the 7q region would not have been considered worthy of follow-up based on the qualitative results. Second, the subset of 132 families responsible for the linkage evidence had offspring with the *earliest* language development. Therefore, the language QTL does not appear to confer susceptibility for severe language *delay* in this sample, but rather may reflect the effects of a more general language-related locus. This is supported by linkage of other language-related traits and disorders to this region.[42,43] Third, of the 132 families responsible for the evidence in support of linkage of WORD to 7q, only 40% were from the original QTL analysis. Thus, similar percentages of families from each cohort contributed; the new families that joined the AGRE sample since the initial scan[23] contributed to the linkage evidence. These results suggest that there has not been systematic ascertainment for families that support a WORD QTL on 7q. The present results support the previous hypothesis[23] that one or more language loci can act alone or in combination with a more modest autism gene on chromosome 7 to provide risk for language disorders, or the broader autism phenotype.

The present linkage results were obtained from the largest collection of autism families available. Alternative approaches to extend the analysis of autism endophenotypes currently include incorporating language development information from unaffected siblings and parents of the probands into the linkage analysis, examining the prospective language and social development of young children that have not yet been diagnosed with autism, and analyzing additional psychometric measures of behavior and cognition in affected children. Owing to the complex etiology of the disorder and the mounting evidence for genetic heterogeneity, several strategies and large samples for replication will be necessary to identify the genes involved with autism.

Acknowledgements

We gratefully acknowledge the contributions of the AGRE families who participated in this study and have made the resource possible. In addition, we thank the AGRE Consortium (see Appendix A) for their oversight of the resource. This project was

supported by NIMH Grant R01 MH 64547 (to DHG), the UCLA Center for Autism Research and Treatment and the Cure Autism Now foundation (to AGRE). TCG also acknowledges support from the generous contribution of Dr Judith P Sulzberger.

Electronic-Database Information

Autism Genetic Research Exchange, http://agre.org/
Cure Autism Now, http://canfoundation.org/
Genome Database, The, http://gdbwww.gdb.org/
Human Biological Database Interchange, http://www.hbdi.org/
Marshfield Medical Research Foundation, The, http://research.marshfieldclinic.org/genetics/
Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/omim/ (for Autism Disorder (MIM 209850))

References

1 Yeargin-Allsopp M, Rice C, Karapurkar T, Doernberg N, Boyle C, Murphy C. Prevalence of autism in a US metropolitan area. *JAMA* 2003; **289**: 49–55.
2 Bailey A, Le Couteur A, Gottesman I, Bolton P, Simonoff E, Yuzda E *et al.* Autism as a strongly genetic disorder: evidence from a British twin study. *Psychol Med* 1995; **25**: 63–77.
3 Smalley SL, Asarnow RF, Spence A. Autism and genetics. *Arch Gen Psychiatry* 1988; **45**: 953–961.
4 Wassink TH, Piven J, Patil SR. Chromosomal abnormalities in a clinic sample of individuals with autistic disorder. *Psychiatr Genet* 2001; **11**: 57–63.
5 Barrett S, Beck JC, Bernier R, Bisson E, Braun TA, Casavant TL *et al.* An autosomal genomic screen for autism. Collaborative linkage study of autism. *Am J Med Genet* 1999; **88**: 609–615.
6 Lamb JA, Moore J, Bailey A, Monaco AP. Autism: recent molecular genetic advances. *Hum Mol Genet* 2000; **9**(6): 861–868.
7 IMGSAC. A full genome screen for autism with evidence for linkage to a region on chromosome 7q. International Molecular Genetic Study of Autism Consortium. *Hum Mol Genet* 1998; **7**: 571–578.
8 IMGSAC. A genomewide screen for autism: strong evidence for linkage to chromosomes 2q, 7q, and 16p. *Am J Hum Genet* 2001; **69**: 570–581.
9 Liu J, Nyholt DR, Magnussen P, Parano E, Pavone P, Geschwind DH *et al.* A genomewide screen for autism susceptibility loci. *Am J Hum Genet* 2001; **69**: 327–340.
10 Philippe A, Martinez M, Guilloud-Bataille M, Gillberg C, Råstam M, Sponheim E *et al.* Genome-wide scan for autism susceptibility genes. Paris Autism Research International Sibpair Study. *Hum Mol Genet* 1999; **8**: 805–812.
11 Risch N, Spiker D, Lotspeich L, Nouri N, Hinds D, Hallmayer J *et al.* A genomic screen of autism: evidence for a multilocus etiology. *Am J Hum Genet* 1999; **65**: 493–507.
12 Veenstra-VanderWeele J, Cook EH. Molecular genetics of autism spectrum disorder. *Mol Psychiatry* 2004; **9**: 819–832.
13 Bryson SE, Smith IM. Epidemiology of autism: prevalence, associated characteristics, and implications for research and service delivery. *Ment Retard Dev Disabilities Res Rev* 1998; **4**: 97–103.
14 Stoltenberg SF, Burmeister M. Recent progress in psychiatric genetics-some hope but no hype. *Hum Mol Genet* 2000; **9**: 927–935.
15 Piven J, Palmer P, Landa R, Santangelo S, Jacobi D, Childress D. Personality and language characteristics in parents from multiple-incidence autism families. *Am J Med Genet* 1997; **74**: 398–411.
16 Gottesman I, Gould TD. The endophenotype concept in psychiatry: etymology and strategic intentions. *Am J Psychiatry* 2003; **160**: 636–645.
17 Gottesman I, Shields J. *Schizophrenia and Genetics: A Twin Study Vantage Point*. New York: Academic Press, 1972.
18 Bradford Y, Haines J, Hutcheson H, Gardiner M, Braun T, Sheffield V *et al.* Incorporating language phenotypes strengthens evidence of linkage to autism. *Am J Med Genet* 2001; **105**: 539–547.
19 Buxbaum JD, Silverman JM, Smith CJ, Kilifarski M, Reichert J, Hollander E *et al.* Evidence for a susceptibility gene for autism on chromosome 2 and for genetic heterogeneity. *Am J Hum Genet* 2001; **68**: 1514–1520.
20 Shao Y, Cuccaro ML, Hauser ER, Raiford KL, Menold MM, Wolpert CM *et al.* Fine mapping of autistic disorder to chromosome 15q11–q13 by use of phenotypic subtypes. *Am J Hum Genet* 2003; **72**: 539–548.
21 Shao Y, Raiford KL, Wolpert CM, Cope HA, Ravan SA, Ashley-Koch AA *et al.* Phenotypic homogeneity provides increased support for linkage on chromosome 2 in autistic disorder. *Am J Hum Genet* 2002; **70**: 1058–1061.
22 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.
23 Alarcón M, Cantor RM, Liu J, Gilliam TC, Consortium A, Geschwind DH. Evidence for a language QTL on chromosome 7q in multiplex autism families. *Am J Hum Genet* 2002; **70**: 60–71.
24 Hauser ER, Watanabe RM, Duren WL, Bass MP, Langefeld CD, Boehnke M. Ordered subset analysis in genetic linkage mapping of complex traits. *Genet Epidemiol* 2004; **27**: 53–63.
25 Scott WK, Hauser ER, Schmechel DE, Welsh-Bohmer KA, Small GW, Roses AD *et al.* Ordered-subsets linkage analysis detects novel Alzheimer disease loci on chromosomes 2q34 and 15q22. *Am J Hum Genet* 2003; **73**: 1041–1051.
26 Geschwind DH, Sowinski J, Lord C, Iversen P, Shestack J, Jones P *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.
27 Yonan AL, Alarcon M, Cheng R, Magnusson PK, Spence SJ, Palmer AA *et al.* A genomewide screen of 345 families for autism-susceptibility loci. *Am J Hum Genet* 2003; **73**: 886–897.
28 Aita VM, Christiano AM, Gilliam TC. Mapping complex traits in diseases of the hair and skin. *Exp Dermatol* 1999; **8**: 439–452.
29 O'Connell JR, Weeks DE. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* 1998; **63**: 259–266.
30 Kruglyak L, Lander ES. Complete multipoint sib-pair analysis of qualitative and quantitative traits. *Am J Hum Genet* 1995; **57**: 439–454.
31 Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES. Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* 1996; **58**: 1347–1363.
32 Conover WJ. *Practical Nonparametric Statistics*, 2nd edn. John Wiley and Sons Inc: New York, 1980.
33 Shao Y, Wolpert CM, Raiford KL, Menold MM, Donnelly SL, Ravan SA *et al.* Genomic screen and follow-up analysis for autistic disorder. *Am J Med Genet* 2002; **114**: 99–105.
34 Ashley-Koch A, Wolpert CM, Menold MM, Zaeem L, Basu S, Donnelly SL *et al.* Genetic studies of autistic disorder and chromosome 7. *Genomics* 1999; **61**: 227–236.
35 Auranen M, Vanhala R, Varilo T, Ayers K, Kempas E, Ylisaukko-Oja T *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.
36 Kaminen N, Hannula-Jouppi K, Kestila M, Lahermo P, Muller K, Kaaranen M *et al.* A genome scan for developmental dyslexia confirms linkage to chromosome 2p11 and suggests a new locus on 7q32. *J Med Genet* 2003; **40**: 340–345.
37 Silverman JM, Smith CJ, Schmeidler J, Hollander E, Lawlor BA, Fitzgerald M *et al.* Symptom domains in autism and related conditions: evidence for familiality. *Am J Med Genet* 2002; **114**: 64–73.
38 Consortium S. A genomewide scan identifies two novel loci involved in specific language impairment. *Am J Hum Genet* 2002; **70**: 384–398.
39 Badner JA, Gershon ES. Regional meta-analysis of published data supports linkage of autism with markers on chromosome 7. *Mol Psychiatry* 2002; **7**: 56–66.
40 Fisher SE, Francks C, McCracken JT, McGough JJ, Marlow AJ, MacPhie IL *et al.* A genomewide scan for loci involved in

attention-deficit/hyperactivity disorder. *Am J Hum Genet* 2002; **70**: 1183–1196.
41 Stein CM, Schick JH, Gerry Taylor H, Shriberg LD, Millard C, Kundtz-Kluge A *et al*. Pleiotropic effects of a chromosome 3 locus on speech-sound disorder and reading. *Am J Hum Genet* 2004; **74**: 283–297.
42 Lai CS, Fisher SE, Hurst JA, Levy ER, Hodgson S, Fox M *et al*. The SPCH1 region on human 7q31: genomic characterization of the critical interval and localization of translocations associated with speech and language disorder [see comments]. *Am J Hum Genet* 2000; **67**: 357–368.
43 Fisher SE, Vargha-Khadem F, Watkins KE, Monaco AP, Pembrey ME. Localisation of a gene implicated in a severe speech and language disorder [published erratum appears in Nat Genet 1998 Mar;18(3):298]. *Nat Genet* 1998; **18**: 168–170.

Appendix A

AGRE Consortium

Daniel H Geschwind, University of California at Los Angeles, Los Angeles; Maja Bucan, University of Pennsylvania, Philadelphia; W Ted Brown, New York State Institute for Basic Research in Developmental Disabilities, Long Island; Joseph D Buxbaum, Mt Sinai School of Medicine, New York; Rita M Cantor, University of California, Los Angeles; John N Constantino, Washington University School of Medicine, St Louis; T Conrad Gilliam, Columbia Genome Center, New York; Clara Lajonchere, Cure Autism Now, Los Angeles; David H Ledbetter, Emory University, Atlanta; Christa Lese-Martin, Emory University, Atlanta; Janet Miller, Cure Austism Now, Los Angeles; Stanley F Nelson, University of California at Los Angeles School of Medicine, Los Angeles; Gerard D Schellenberg, University of Washington and Veterans Affairs Medical Center, Seattle; Carole Samango-Sprouse, Children's National Medical Center, Baltimore; Sarah J Spence, University of California, Los Angeles; Rudolph E Tanzi, Massachusetts General Hospital, Boston.