

## ORIGINAL ARTICLE

# Evidence for multiple loci from a genome scan of autism kindreds

GD Schellenberg<sup>1,2,3,4</sup>, G Dawson<sup>5,6</sup>, YJ Sung<sup>7</sup>, A Estes<sup>6</sup>, J Munson<sup>5</sup>, E Rosenthal<sup>8</sup>, J Rothstein<sup>8</sup>, P Flodman<sup>9</sup>, M Smith<sup>10</sup>, H Coon<sup>10</sup>, L Leong<sup>1,2</sup>, C-E Yu<sup>1,2</sup>, C Stodgell<sup>11</sup>, PM Rodier<sup>11</sup>, MA Spence<sup>9</sup>, N Minshew<sup>12</sup>, WM McMahon<sup>10</sup> and EM Wijsman<sup>7,8</sup>

<sup>1</sup>Geriatrics Research Education and Clinical Center, Puget Sound Veterans Affairs Medical Center, Seattle, WA, USA;

<sup>2</sup>Gerontology and Geriatric Medicine, Department of Medicine, University of Washington, Seattle, WA, USA; <sup>3</sup>Department of Neurology, University of Washington, Seattle, WA, USA; <sup>4</sup>Department of Pharmacology, University of Washington, Seattle, WA, USA; <sup>5</sup>Department of Psychology and the Center on Human Development and Disability, University of Washington, Seattle, WA, USA; <sup>6</sup>Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA, USA;

<sup>7</sup>Department of Medicine, Division of Medical Genetics, University of Washington, Seattle, WA, USA; <sup>8</sup>Department of Biostatistics, University of Washington, Seattle, WA, USA; <sup>9</sup>Department of Pediatrics, University of California, Irvine, CA, USA;

<sup>10</sup>Department of Psychiatry, Division of Child and Adolescent Psychiatry, University of Utah, Salt Lake City, UT, USA;

<sup>11</sup>Department of OB/GYN, University of Rochester Medical Center, Rochester, NY, USA and <sup>12</sup>Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA, USA

**We performed a genome-wide linkage scan using highly polymorphic microsatellite markers. To minimize genetic heterogeneity, we focused on sibpairs meeting the strict diagnosis of autism. In our primary analyses, we observed a strong linkage signal ( $P=0.0006$ , 133.16 cM) on chromosome 7q at a location coincident with other linkage studies. When a more relaxed diagnostic criteria was used, linkage evidence at this location was weaker ( $P=0.01$ ). The sample was stratified into families with only male affected subjects (MO) and families with at least one female affected subject (FC). The strongest signal unique to the MO group was on chromosome 11 ( $P=0.0009$ , 83.82 cM), and for the FC group on chromosome 4 ( $P=0.002$ , 111.41 cM). We also divided the sample into regression positive and regression negative families. The regression-positive group showed modest linkage signals on chromosomes 10 ( $P=0.003$ , 0 cM) and 14 ( $P=0.005$ , 104.2 cM). More significant peaks were seen in the regression negative group on chromosomes 3 ( $P=0.0002$ , 140.06 cM) and 4 ( $P=0.0005$ , 111.41 cM). Finally, we used language acquisition data as a quantitative trait in our linkage analysis and observed a chromosome 9 signal (149.01 cM) of  $P=0.00006$  and an empirical  $P$ -value of 0.0008 at the same location. Our work provides strong conformation for an autism locus on 7q and suggestive evidence for several other chromosomal locations. Diagnostic specificity and detailed analysis of the autism phenotype is critical for identifying autism loci.**

*Molecular Psychiatry* advance online publication, 1 August 2006; doi:10.1038/sj.mp.4001874

**Keywords:** autism; autism spectrum disorder; linkage; genome scan; chromosome 7

## Introduction

Autism (MIM209850) is a neurodevelopmental disorder that is manifested in early childhood and is characterized by impairments in reciprocal social interactions and language, and a restricted range of behaviors and interests. Whereas most children with autism develop symptoms in the first year of life, approximately a quarter appear to develop normally, then regress in the second or third year of life.<sup>1</sup> Autism is considered a spectrum disorder with heterogeneity in symptom presentation. Approxi-

mately 70% of individuals with autism have mental retardation. A strict diagnosis of autism requires symptoms in all three domains (social interaction, language, restricted range of activities) and onset early in life. The broader autism spectrum includes Asperger syndrome, where affected individuals have deficits in social interaction, and in repetitive and restricted behaviors, but have relatively intact language abilities.

Inheritance clearly plays a major role in susceptibility to autism. Monozygotic (MZ) twins are 69–95% concordant,<sup>2,3</sup> whereas dizygotic (DZ) twins are 0–24% concordant.<sup>2–6</sup> When a broader phenotype (BPH) is used, concordance increases for both MZ (88–91%) and DZ (9–30%) twins.<sup>2,5,6</sup> Reported sib risks are consistent with DZ twin concordance rates ranging from 2.8 to 7.0% whereas rates in control families are

Correspondence: Dr GD Schellenberg, Veterans Affairs Medical Center (182b), 1660 S. Columbian Road, Seattle, WA 98108, USA.  
E-mail: zachdad@u.washington.edu

Received 28 February 2006; revised 28 April 2006; accepted 1 May 2006

much lower.<sup>7–9</sup> In contrast, the rate of autism in relatives more distant than sibs is extremely low.<sup>10</sup> This risk pattern is consistent with autism susceptibility resulting from the presence of elevated-risk alleles at several genes.

We presently know little about the genes underlying autism susceptibility. In rare cases, autism is associated with an extra copy or copies of the Prader–Willi syndrome region of chromosome 15q11–q13. Autism is also associated with the relatively rare single gene disorders fragile X syndrome, tuberous sclerosis, neurofibromatosis and phenylketonuria. There are numerous reported cases of chromosomal abnormalities associated with autism, though the translocation sites are scattered throughout the genome.<sup>11</sup> For most autism cases, there is no identifiable gene or genetic change associated with the disorder. Efforts to identify autism genes have focused on linkage and candidate gene studies. To date, genome-wide linkage studies of autism in eight largely non-overlapping family collections have been published.<sup>12–25</sup> The number of families in these studies ranges from 39<sup>12</sup> to 345<sup>24</sup> families with most kindreds having two affected siblings. While no single region has been unambiguously associated with autism, coincident signals for specific chromosomal regions have been identified in multiple studies. These include 1p (~137–149 cM),<sup>12,19</sup> 2q (~177–198 cM),<sup>13,17,22,23</sup> 7q (104–145 cM)<sup>15–17,25</sup> 17q (61 cM)<sup>14,17,21,22</sup> and 19q (~60 cM),<sup>18,23</sup> with the 2q, 7q and 17q regions giving the strongest signals. Numerous candidate gene association studies have also been described but no association finding has been consistently and independently replicated across multiple studies.<sup>26</sup>

Several factors potentially contribute to inconsistent linkage and association findings. The extent of genetic heterogeneity in autism is unknown, and if extensive, could degrade power to detect susceptibility loci. Differences in results across linkage studies may reflect different study designs that affect the underlying genetic architecture of the sample. Diagnostic inclusion criteria vary between studies with some having broader inclusion criteria than others. Likewise, the diagnostic specificity can be substantially affected by the complement of methods used to evaluate affected individuals. The most rigorous approach, using the Autism Diagnostic Interview (ADI) and the Autism Diagnostic Observation Scale (ADOS), combined with a clinical diagnosis of individuals, is not always used. Also, because of pragmatic difficulties with accumulating large sample sizes of affected sib pairs (ASPs), no study to date has restricted enrollment to only sib pairs with the strictest diagnosis of autism in all affected individuals, although some studies have required that one sib meets the strict diagnosis. In addition, to attempt to reduce genetic heterogeneity, some groups have also defined phenotypic subsets of autism families using measures such as regression,<sup>27</sup> age at first word or phrase,<sup>28,29</sup> language delay,<sup>13,30</sup>

repetitive and stereotypic behaviors<sup>28–30</sup> and obsessive-compulsive behaviors.<sup>31</sup> These studies provide additional evidence for autism loci at 1q24.2, 7q, 13q, 17q and 21q.

To identify regions of the genome containing autism genes, we performed a genome-wide linkage scan using autism families not previously analyzed. To minimize the effects of heterogeneity, we analyzed 169 families where at least two sibs met the strict criteria for autism. We also analyzed a larger set of 222 families that, in addition to individuals with strictly defined autism, included affected participants who did not meet full criteria for autism but met criteria for broader autism spectrum disorder (ASD). For our primary analysis, the trait used is the affection status of family members. Based on recent findings by others, we also conducted secondary stratified and quantitative trait analyses of: (1) families with only male affected children (male-only, MO) versus families where there was at least one affected female child (female-containing, FC);<sup>17,20</sup> (2) families with or without developmental regression;<sup>27</sup> and (3) age at first word and phrase as quantitative traits.<sup>13,28–30</sup> Our results provide strong evidence for an autism locus on chromosome 7q with the evidence for linkage being strongest in the subset with the strictest diagnosis. This provides strong confirmation of previous results for chromosome 7q.<sup>16,17,25</sup> Likewise, we replicate a previously reported signal on chromosome 4 from FC sibships.<sup>20</sup> We obtained additional novel strong signals at the 9q terminal region using age at first word as a quantitative trait, chromosome 11 in MO sibships and chromosome 3 in regression-negative families. The fact that we were able to confirm signals from other studies strongly suggests that linkage analysis will lead to successful identification of at least some of the genes responsible for autism susceptibility.

## Materials and methods

### *Characterization of autism families, diagnostic criteria and phenotypes*

Families who identified themselves as having two or more children with autism (AUT), pervasive developmental disorder (PDD) and/or ASD were recruited as part of the National Institutes of Health Collaborative Programs of Excellence in Autism (CPEA). The 222 families used in the current genome scans were recruited through the University of Washington Autism Center (200 families), the University of California at Irvine (5 families), the University of Utah (10 families), the University of Rochester (2 families) and the University of Pittsburgh (5 families). The University of Washington recruitment network includes testing sites in Washington, Oregon, Florida, Tennessee, Arizona, Minnesota and Texas with approximately 50% of families living in the greater Seattle region. Fourteen families included multiple marriages, and two families were extended and included affected cousins. When possible, a blood

sample was obtained from parents, each affected sibling and an unaffected sibling. DNA from at least one parent was available from each family in all but one kindred, with 192 families providing DNA from both parents of affected children. Exclusionary criteria include age below 3 years, presence of a known genetic condition (e.g. neurofibromatosis (MIM + 162200), Tourette syndrome (MIM #137580), phenylketonuria (MIM + 261600), fragile X syndrome (MIM + 309550), Williams–Beuren syndrome (MIM #194050), velocardiofacial syndrome (MIM #192430)), history of serious head injury or neurological disease (e.g. encephalitis) or significant sensory or motor impairment.

For diagnostic evaluation, all affected individuals were administered the ADI – Revised (ADI-R)<sup>32</sup> and the ADOS – Generic (ADOS-G)<sup>33</sup> by clinicians trained to research reliability. The ADI-R and the ADOS-G are standardized measures used to diagnose ASDs across a range of ages and developmental levels. In addition, a clinical diagnosis, based on DSM-IV criteria,<sup>34</sup> of Autistic Disorder or Pervasive Developmental Disorder, Not otherwise specified (PDD-NOS) was made by the clinician evaluating the child. Children were administered a standardized intellectual evaluation appropriate for their age level (Mullen Scales of Early Learning,<sup>35</sup> Wechsler Preschool and Primary Scale of Intelligence – Revised, Wechsler Intelligence Scale for Children – Third Edition; Wechsler scales used a short form consisting of Vocabulary, Comprehension, Block Design and Object Assembly subtests). An estimated full scale IQ score for the Wechsler short forms was calculated as described by Sattler,<sup>36</sup> which takes into account subtest reliability and validity. Descriptive information regarding diagnosis and IQ is in Table 1. Full scale IQ scores for the affected children are distributed as follows: scores below 55, 26.7%; scores between 55 and 69, 14.9%; scores between 70 and 84, 16.8%; score 85 or higher, 40.7%. Ethnicities of the affected children are: White, Non-Hispanic, 85.7%; White, Hispanic, 3.1%; American Indian/Alaska Native, 2.2%; Asian, 1.0%; African-American, 1.0%; More than one race, 7.0%. Parents' report of their education level was as follows: Less than high school diploma, 1.6%; high school diploma, 7.8%; some college, 35.3%; college degree, 40.1%; and graduate degree, 15.2%.

The final diagnosis was based on the National Institute of Child Health and Development (NICHD) CPEA diagnostic criteria established by Catherine Lord and the CPEA diagnostic subcommittee in May 2003, which included operational definitions for research diagnoses of Autism, PDD and ASD. An individual was assigned a research diagnosis of autism if he or she met criteria for autism on the ADI-R and criteria for autism or ASD criteria on the ADOS. One exception, which applied to 21 children, was as follows: If an individual scored within 2 points of meeting criteria for autism on the ADI-R, met autism criteria on ADOS and also met DSM-IV criteria<sup>34</sup> for a clinical diagnosis of autism, the individual was assigned a research diagnosis of autism. For the research diagnosis of PDD, the individual met criteria for ASD on the ADOS-G and, on the ADI-R, met criteria for either the social or communication domain, and was at least within 2 points of meeting criteria on these domains. A research diagnosis of ASD was assigned if either the ADI-R or ADOS-G criteria were met. Children who displayed symptoms of autism but did not meet criteria for autism, PDD or ASD were referred to as BPH. All 222 families used had two or more children who each met criteria for autism, PDD or ASD. Some families included additional children with BPH. These children were not included in the affected sib-pair linkage analyses, but were included in quantitative trait analyses of language acquisition. For 169 of the 222 families, two or more children met criteria for autism, and for all but four of the 222 pedigrees, at least one child met criteria for autism (Table 2). These studies were approved by the University of Washington institutional review board and appropriate informed consent was obtained from all participants.

For some analyses, specific items from the ADI-R were used. Two quantitative language acquisition measures, age at first word and age at first phrase, were from ADI-R items 12 and 13, respectively. The presence of regression was determined by a score of 2 on item 38 (loss of spontaneous communicative speech), or item 39 (loss of words used spontaneously but without clear communicative intent) or item 95 (loss of other skills) of the ADI-R in any child with a diagnosis of AUT, PDD, ASD or BPH.

**Table 1** Characteristics for affected individuals

Diagnosis	Total N	Male:female	Mean age (years) at IQ (s.d.)	Mean full scale IQ (s.d.)
AUT	399	328:71	8.75 (4.23)	74.23 (25.95)
PDD	24	19:5	9.05 (6.24)	92.53 (27.29)
ASD	40	30:10	9.99 (4.38)	93.39 (21.73)
BPH	15	11:4	7.26 (2.03)	100.23 (14.71)
Total	478	388:90	8.82 (4.31)	77.53 (26.44)

Abbreviations: AUT, autism; ASD, autism spectrum disorder; BPH, broader autism phenotype; PDD, pervasive developmental disorder – not otherwise specified. Diagnoses were made as described under Materials and methods.

**Table 2** Families and individuals genotyped for linkage studies

Number of families	Type of family	Number of families with unaffected sibs					
		No unaffected sibs	One unaffected sib	Two unaffected sibs	Three unaffected sibs	One sib – missing diagnosis	One sib – Down syndrome
<i>Strict diagnosis families</i>							
147	AUT/AUT	93	48	3 <sup>a</sup>	2	0	1
3	AUT/AUT/ASD	1	2	0	0	0	0
7	AUT/AUT/AUT	4	1	1	0	1	0
1	AUT/AUT/AUT/AUT/AUT/plus <sup>b</sup>	1	0	0	0	0	0
1	AUT/AUT/AUT/BPH	1	0	0	0	0	0
1	AUT/AUT/AUT/plus <sup>b</sup>	1	0	0	0	0	0
6	AUT/AUT/BPH	6	0	0	0	0	0
1	AUT/AUT/BPH/BPH/BPH	1	0	0	0	0	0
1	AUT/AUT/PDD	1	0	0	0	0	0
1	AUT/AUT/PDD/PDD	1	0	0	0	0	0
Total: 169		110	51	4	2	1	1
<i>Additional families</i>							
16	AUT/PDD	12	1	1	1	1	0
1	AUT/PDD/ASD	1	0	0	0	0	0
1	AUT/PDD/BPH	1	0	0	0	0	0
29	AUT/ASD	20	7	1	0	1	0
2	AUT/ASD/BPH	1	1	0	0	0	0
1	PDD/ASD	1	0	0	0	0	0
2	PDD/ASD/BPH	2	0	0	0	0	0
1	ASD/ASD	1	0	0	0	0	0
Total: 53		39	9	2	1	2	0
Grand total: 222 broad diagnosis families							

<sup>a</sup>One AUT/AUT family also has a third sib with a missing diagnosis, but is not listed again in the 'One sib – missing diagnosis' column.

<sup>b</sup>'plus' indicates two sibships that are cousins. Abbreviations, see Table 1 footnote and diagnostic methods are those described in the Materials and methods section. Individuals with a diagnosis of BPH were not used in ASP linkage analysis.

### Genotyping

For 996 individuals (affected and non-affected siblings and parents), genotypes were determined by PCR amplification of polymorphic loci using primers labeled with fluorescent probes. The 410 markers used were primarily from the ABI Prism Linkage Mapping set, Version 2.5, HD5 (<http://home.applied-biosystems.com/>). The mean distance between markers was 9.07 cM, and the largest gap was 20.0 cM. DNA fragments were analyzed using an ABI377 or an ABI3100 DNA sequencing instrument and GeneScan and Genotyper software. Technicians performing the genotyping were blind to the diagnosis of the individuals. Genotyping was 97.1% complete overall, with 18.2% of individuals having 100% complete genotyping and 50% of individuals complete for >99.5% of markers.

### Data cleaning

We used RELPAIR, version 2.0<sup>37</sup> to check pedigree relationships and to identify sample swaps, using the

full genome scan data and an assumption of a 1% genotyping error rate. In the course of this analysis, eight MZ twin pairs were identified, and seven families with unresolvable sample swaps and/or relationship problems were identified. After resolving mis-specified relationships and removing MZ twin pairs and families with pedigree relationship problems, we used Loki<sup>38</sup> to identify all Mendelian-inconsistent genotypes in the remaining 222 families. For subsequent linkage analyses, 723 cases of Mendelian-inconsistent genotypes that could not be resolved by re-genotyping or through other data checks were coded as missing genotypes for all individuals in the pedigree(s) involved. These Mendelian-inconsistent errors represent an empirical incompatibility rate of 0.79% per pedigree.

### Linkage analysis

Multipoint analyses for genome scans were carried out for each sample stratum described below. We used the computer program Merlin<sup>39</sup> to carry out linkage

computations, with the Rutgers sex-averaged map<sup>40</sup> to provide genetic distances between markers. This map is expected to be the most accurate currently available because of the sample size used to construct the map. Analyses were carried out using a Haldane map and mapping function, but all results are presented here on the equivalent Kosambi map to facilitate comparison with other studies. For ASP analyses, the statistic  $S_{\text{all}}$ <sup>41</sup> was used to test for linkage, using the null distribution parameterized as suggested elsewhere<sup>42</sup> to prevent an overly-conservative test.  $S_{\text{all}}$  was used instead of  $S_{\text{pairs}}$  because  $S_{\text{all}}$  is statistically more efficient when there are multiple affected individuals in pedigrees, as we have in our sample. For these analyses, any children who were not explicitly coded as affected were not used. For quantitative language traits, we used variance component methods<sup>43</sup> by contrasting the likelihood under an additive variance major gene model with the likelihood under a polygenic model. For these quantitative trait analyses, we used language scores of all individuals for whom these were available; this included some individuals with BPH, but did not generally include unaffected individuals. All results are presented as the negative of  $\log_{10}$  of the  $P$ -value for a particular genomic location.

For a subset of the most positive linkage analysis results, empirical  $P$ -values were generated by simulation with Merlin.<sup>39</sup> Marker genotypes were simulated 20 000–40 000 times, depending on the asymptotic  $P$ -value, under the assumed marker map but under the null hypothesis of no linkage with the trait, and the resulting data were reanalyzed with the same multi-point methods used to analyze the original data. Simulation analysis was computationally intensive, requiring approximately 1–2 CPU week per chromosome, depending on the chromosome. We present normality-based  $P$ -values, except where explicitly noted because of the large computational burden needed to obtain empirical  $P$ -values combined with the similarity in results between empirical and normality-based asymptotic results for the ASP analyses.

#### Sample stratification

For the primary ASP scan, the families were divided into two groups based on the diagnosis of individuals in the families. The ‘strict’ category consisted of families with two or more sibs each with the diagnosis of autism (AUT, Table 2), the most stringent diagnostic category. For the ASP linkage analyses of the ‘strict’ families, only AUT–AUT pairs were considered as affected even if there were other individuals in the family with different diagnoses (PDD or ASD). The ‘broad’ diagnosis category included both the ‘strict’ families and 53 other families with other types of affected sibpairs also included in the analysis as affected individuals (e.g. AUT–PDD, etc). For analyses of the ‘broad’ category, all individuals in the complete set of 222 families with a diagnosis of AUT, ASD or PDD were considered affected. Individuals with a

diagnosis of BPH were not included as affected individuals in either ASP analysis.

Additional strata were also defined for some secondary analyses. One stratification criterion was based on gender of affected individuals, and we split the sample into MO and FC families. A family was included in the MO group only if all individuals with phenotypes AUT, ASD or PDD were male, and if female siblings, if present, were unaffected. All other families were treated as FC. A second stratification was based on a regression as defined above. Families with at least one individual showing regression were placed into the regression-positive group, with the remaining families placed into the regression-negative group. For the purpose of defining such strata, genotype data were not required to be available on the individuals relevant for stratification, since the purpose was to reduce genetic heterogeneity based on phenotype data. Nevertheless, all families in the regression-positive families had at least one genotyped affected individual who showed regression, and only four FC families (three in strict families) had no genotyped affected female subjects.

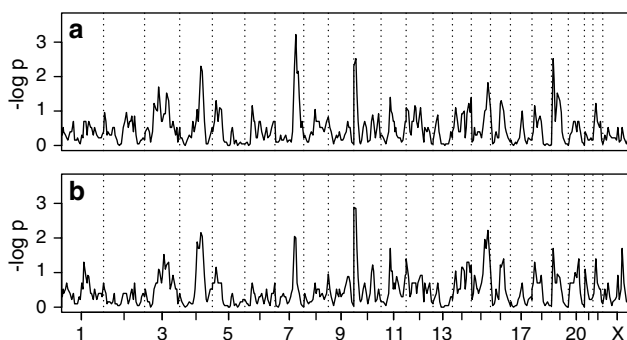
## Results

We performed a genome scan to identify autism susceptibility loci. Analysis methods that did not require a prespecified trait model were used because the exact mode of inheritance for autism susceptibility is unknown, but is consistent with multigenic inheritance. Because genetic heterogeneity can severely reduce power to detect linkage even with trait-model-free analyses, we sought to minimize phenocopy inclusion by focusing on families where two or more siblings have a strict diagnosis of autism (Table 2). In these ‘strict’ category families, only individuals with a diagnosis of autism were considered affected. Also, to insure diagnostic accuracy, the ADI-R was used, and all affected individuals were administered the ADOS-G and were clinically evaluated by a trained clinician. The final diagnosis was based on information from all three sources. This diagnostic protocol was used in our study because recent work shows that diagnostic specificity increases from 56.3–59.4% for the ADI alone to 75.0–85.8% for ADI plus ADOS, relative to the gold standard of a best estimate diagnosis based on ADI-R, ADOS-G and a clinical evaluation (Lord and Risi, personal communication).

Our analyses had three hierarchically related goals. The primary analysis was an affected sib-pair genome scan to detect autism loci using the diagnosis as the phenotype. Our secondary analyses attempted to replicate specific regional linkage signals reported by others who stratified their sample using gender,<sup>20</sup> developmental regression,<sup>27</sup> and language acquisition.<sup>29</sup> Our tertiary analyses explored the possibility of additional regions with relevant loci by completing the scan for the rest of the genome using strata defined by the same gender, developmental regression and language acquisition phenotypes.

### Primary genome scan

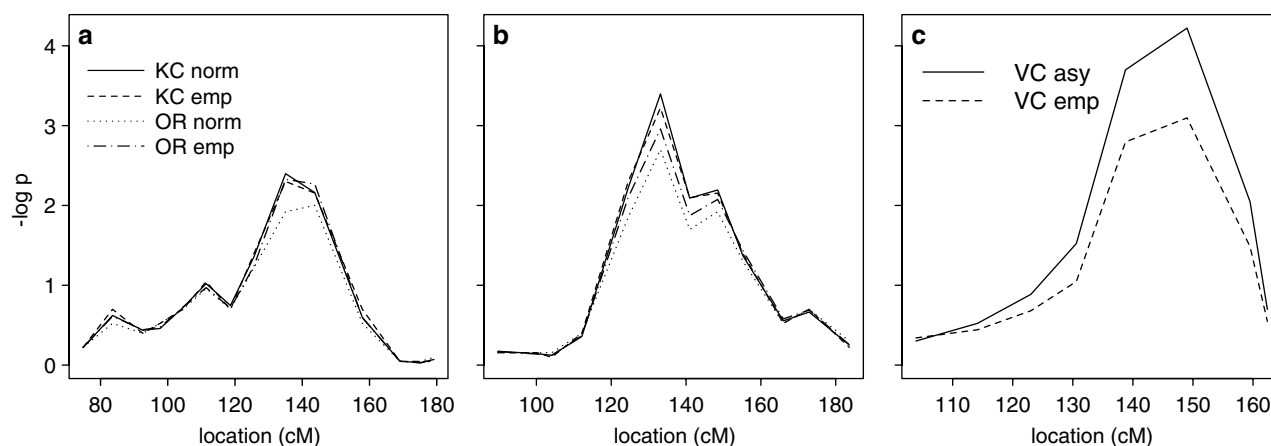
For the primary ASP scan, asymptotic  $P$ -values were computed with the adjustment suggested by Kong and Cox<sup>42</sup> for markers on all autosomes and for the X chromosome (Figure 1, see Supplementary Table S1 for  $P$ -values for the entire scan for all strata considered). We identified several regions with strong evidence for linkage, including both previously reported and novel regions. For the strict diagnostic category, the most significant multipoint  $P$ -values ( $P \leq 0.02$ , Table S2) are 0.0006 at D7S530 (133.16 cM), 0.003 at D19S209 (10.87 cM), 0.005 at D4S1575 (135.03 cM) and 0.02 at D3S1285 (89.56 cM). For the broad diagnostic category, the most significant multipoint  $P$ -values are 0.0014 at D10S591 (13.3 cM), 0.006 at D15S130 (106 cM), 0.02 at D11S935 (58.94 cM) and 0.02 at DXS1001 (123.29 cM). Regions with strong



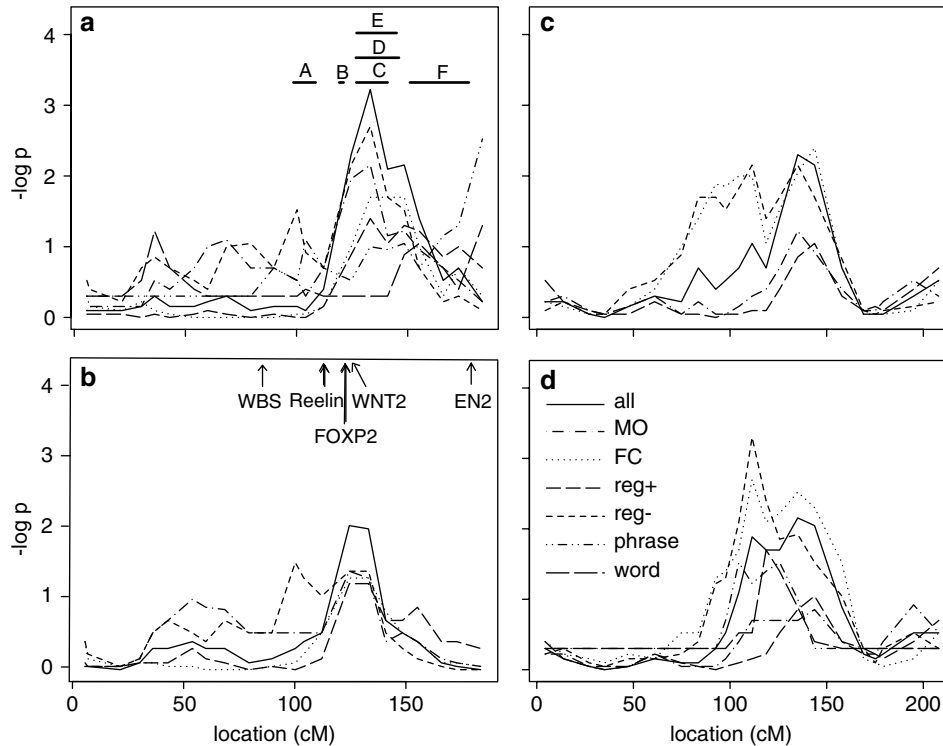
**Figure 1** Primary analysis genome-wide  $P$ -values for strict (a) and broad (b) categories of families. The  $-\log_{10}$  of the  $P$ -values are plotted on the Y-axis versus the genetic location along each chromosome in cM using the Rutgers Kosombi map. Chromosomes are noted by numbers along the X-axis, with tic-marks denoting additional chromosomes.  $P$ -values are normality-based asymptotic  $P$ -values computed as suggested by Kong and Cox.<sup>42</sup>

positive evidence of linkage were also checked with single-marker analyses, which are not sensitive to map assumptions, to verify that positive multipoint signals are not the result of differences between sex-averaged and sex-specific map distances.<sup>44</sup> Results from multipoint and single point analysis were comparable. Single point minimum  $P$ -values ( $P \leq 0.02$ ; Table S2) for the strict diagnostic category were 0.002 at D7S530 (133.16 cM), 0.002 at D4S1575 (135.03 cM), 0.003 at D19S209 (10.87 cM) and 0.011 at D15S130 (106.111 cM). For the broad diagnostic category, single point  $P$ -values were 0.0005 at D10S591 (13.3 cM), 0.0012 at D22S539 (18.78 cM), 0.01 at D11S935 (58.94 cM), 0.011 at D15S130 (106.111 cM) and 0.02 at D3S1278 (123.21 cM). To determine if the normality-based  $P$ -values computed with the Kong and Cox adjustment were accurate, we also computed empirical  $P$ -values for chromosomes 4 and 7. The results are very similar (Figure 2a and b; Table S3), suggesting that for these data, the Kong and Cox adjustment<sup>42</sup> method is accurate. In contrast, unadjusted  $S_{\text{all}}$  gave conservative  $P$ -values as noted by Kong and Cox.<sup>42</sup> Because simulation methods are computationally intensive, for the remaining analysis, we used  $P$ -values obtained directly from  $S_{\text{all}}$  with the Kong and Cox adjustment,<sup>42</sup> unless otherwise noted.

Our strongest signal was on chromosome 7 at 133.16 cM (Figure 3a and b). Evidence for linkage was considerably stronger for the strict diagnosis families ( $P = 0.0006$ ) when compared to the broad diagnosis group ( $P = 0.01$ ), even though the latter had more families (169 versus 222). The location of this signal overlaps with positive results from analysis from the International Molecular Genetic Study of Autism Consortium (IMGSAC)<sup>16,17,25</sup> and from others.<sup>15,19,23</sup> Thus, our data provides strong support for an autism locus on chromosome 7, both as an individual result, and in conjunction with previous results. Notably missing from our genome scan were



**Figure 2** Comparison of empirical  $P$ -values to asymptotic-based  $P$ -values. For chromosomes 4 (a) and 7 (b) empirical  $P$ -values were compared to normality-based (norm) asymptotic  $P$ -values. The empirical  $P$ -values matched the normality-based  $P$ -values better for Kong and Cox (KC)  $S_{\text{all}}$  statistics (dashed line for empirical, solid line for normality-based) than for ordinary (OR)  $S_{\text{all}}$  statistics (dot-dashed line for empirical, dotted line for normality-based). For chromosome 9 (c), empirical  $P$ -values from variable component (VC) analysis (dashed line) were less extreme than asymptotic (asy)  $P$ -values (solid line).



**Figure 3** *P*-values for chromosomes 7 and 4 for all analyses. The top row (**a**, **c**) is the strict and the bottom row (**b**, **d**) is the broad diagnostic group. Chromosome 7: left panels (**a**, **b**). Chromosome 4: right panels (**c**, **d**). Solid lines, all families (strict or broad); dot-dash lines, male only (MO) families; dots, FC families; short dashed, regression negative (reg<sup>-</sup>) families; medium dashed, regression positive (reg<sup>+</sup>) families; long-dashed, age at first word (word); dot-dot-dash, age at first phrase (phrase). Locations of chromosome 7 genes are indicated by arrows. WBS, Williams–Beuren syndrome region (85.34–87.4 cM). Bars at the top of panel a are locations of autism linkage signals (A, Barrett *et al.*<sup>15</sup>; C, Lamb *et al.*<sup>17</sup>; D, Shao *et al.*<sup>23</sup>; F, Alarcon *et al.*<sup>28</sup>), dyslexia (E, Kaminen *et al.*<sup>55</sup>), and specific language impairment (B, O’Brien *et al.*<sup>54</sup>). A color version of this figure is included in the Supplementary material.

signals corresponding to those reported by others on chromosome 2<sup>13,16,17,23,45</sup> and chromosome 17.<sup>14,20,24</sup>

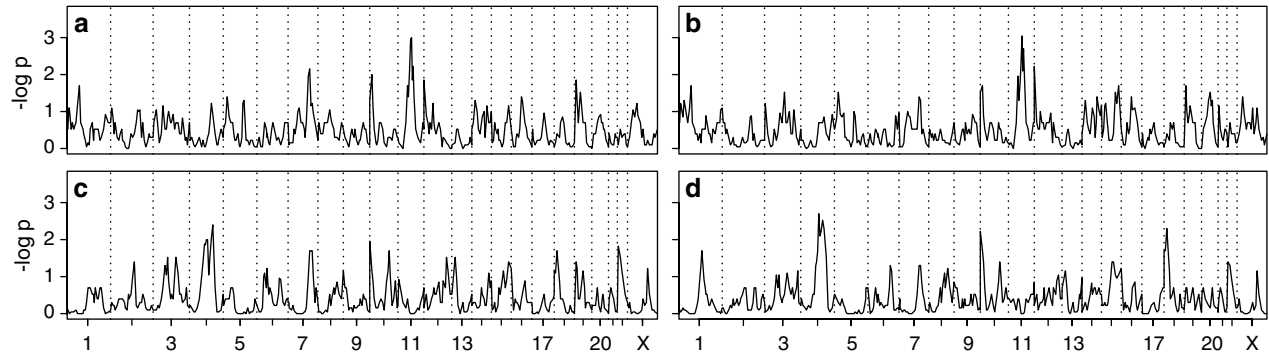
#### Gender and linkage analyses

Previous work by others suggested that linkage signals on chromosomes 4,<sup>14,20</sup> 7,<sup>17</sup> and 17<sup>14,20</sup> were dependent on the gender of the affected individuals in the families. To attempt to replicate these findings, we used the same stratification design, dividing our sample into families with only male affected individuals (MO families; 113 strict, 148 broad), and families with at least one affected female sib (FC families; 56 strict, 74 broad; Table S4). The strongest signal from the primary genome scan on chromosome 7 at 133.16 cM appears to originate from both types of families, though the MO families give a stronger result compared to the FC families (D7S530, strict: MO, 0.007; FC, 0.02; Figures 3a, b and 4; Table S5), as is consistent with a larger sample size in the MO group. This result is not entirely consistent with those from the IMGSAC who observed a MO specific peak at 133.2–141.2 cM.<sup>17</sup> Signals coming predominantly from MO families are at D11S1314 ( $P=0.0009$ , broad, 83.82 cM) and at D12S352 ( $P=0.006$ , broad, 0 cM) (Figures 4 and 5a). Other signals came from either the

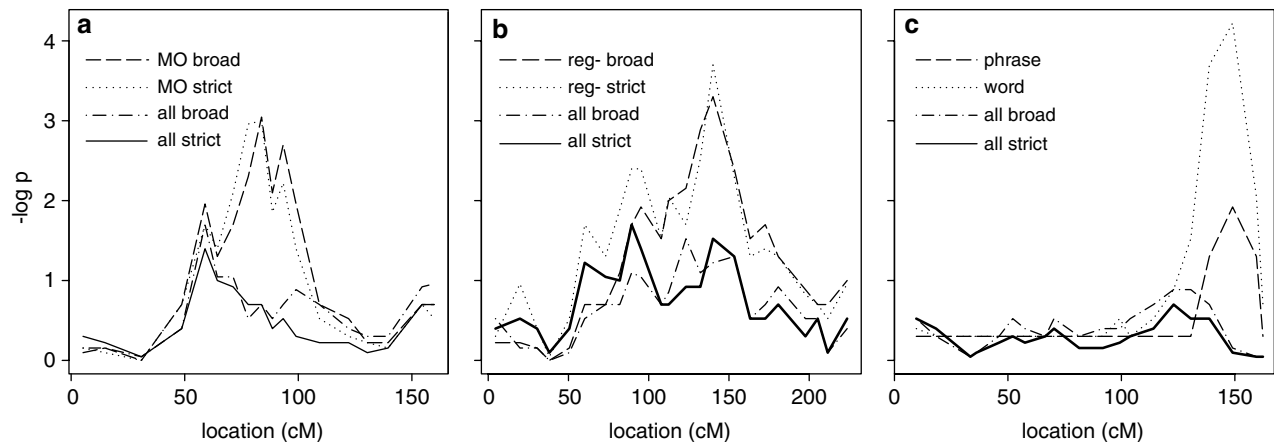
FC group or from both groups. Signals generated almost entirely by the FC group are at D4S1572 ( $P=0.002$ , broad, 111.41 cM), at D10S249 ( $P=0.006$ , broad, 0.0 cM) and at D18S452 ( $P=0.005$ , broad, 16.90 cM) (Figures 3c, d and 4; Table S5). Our evidence for linkage to chromosome 4 is comparable to findings from the AGRE consortium who reported a gender-dependent signal at approximately 160 cM.<sup>20</sup> In our study, other signals on chromosomes 10 (0–13.3 cM), 15 (86–106 cM) and 19 (10.9 cM) appear to derive from both types of families. We did not observe a signal in either gender-defined stratification on chromosome 17.<sup>20</sup>

#### Regression and linkage analysis

In previous work by others, analysis restricted to families containing individuals with developmental regression identified signals on chromosomes 7 and 21.<sup>27</sup> In an attempt to replicate these findings, we stratified our families into those with at least one affected individual with regression versus families with no cases of regression. There were 76 families meeting this criterion in the strict category and 93 families in the broad category (Table S4). The largest signals were at D10D249 ( $P=0.003$ , broad, 0.0 cM)



**Figure 4** Genome-wide  $P$ -values for gender-specific stratification of families. Top panels (**a** and **b**) are male-only families and the bottom panels (**c** and **d**) are FC families. The left column (**a** and **c**) is the 'strict' diagnosis and the right column (**b** and **d**) is the 'broad' diagnostic category.



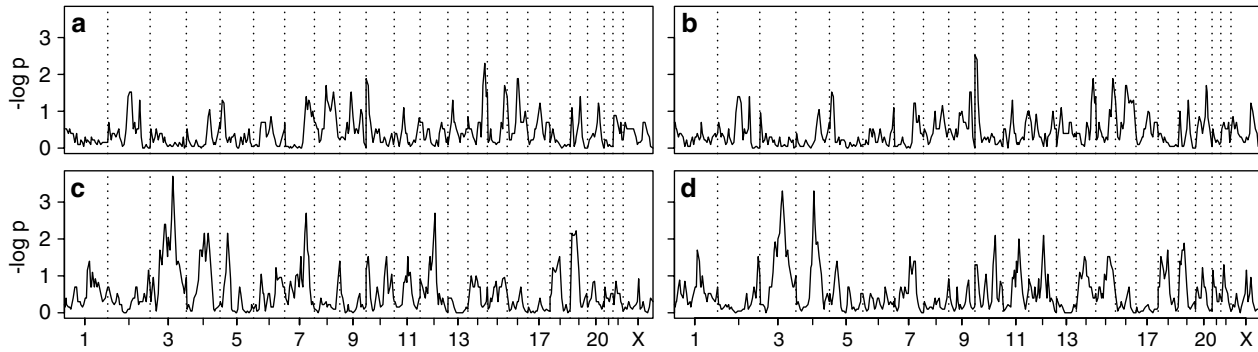
**Figure 5**  $P$ -values for chromosomes 3, 9 and 11. For all panels: solid lines, all families with strict diagnosis; dot-dash all families with broad diagnoses. For chromosome 11 (**a**), dotted line, male-only (MO) with strict diagnosis; dashed line, male-only with broad diagnosis. For chromosome 3 (**b**), dotted line, regression negative (reg-) with strict diagnosis; dashed line, regression negative with broad diagnosis. For chromosome 9 (**c**), dotted line, age at first word (word); dashed line, age at first phrase (phrase).

and D14S65 ( $P=0.005$ , strict, 104.2 cM; Figure 6; Table S6), but were of modest strength, possibly due to small sample size. We did not replicate signals reported by Molloy *et al.*<sup>27</sup> on chromosome 7 (165 cM) or chromosome 21 (0–30 cM). We also analyzed the families without regression (93 strict, 129 broad; Figure 6; Table S6). The most striking results were  $P$ -values of 0.0002 at D3S1292 (strict, 140.06 cM) (Figure 5b), and 0.0005 at D4S1572 (broad, 111.41 cM) (Figure 5c and d) with both being more significant than those in the primary genome scan. The chromosome 4 signal is coincident with the result of the FC family analysis (Figure 3c and d) and the chromosome 3 signal is coincident with results from both gender-strata (Figure 5b).

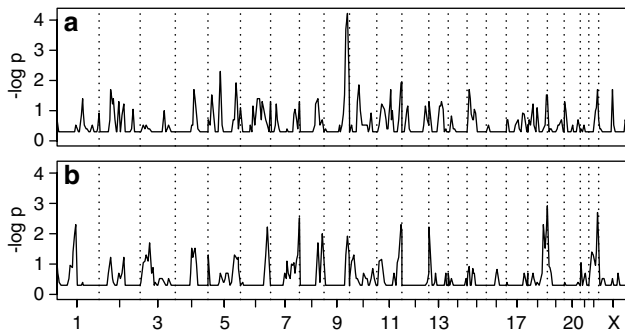
#### Language acquisition as a quantitative trait in linkage analysis

Language delay was used by others in autism linkage studies, either as a dichotomous trait<sup>13,30,45</sup> or as a

quantitative trait.<sup>28</sup> In these reports, language acquisition contributed to signals on chromosomes 2, 7 and 13. To attempt to replicate these studies, we performed variance-component quantitative trait linkage analyses for age at first word and for age at first phrase. For these analyses, we used all families for which we had language phenotypic data on at least two children. The most remarkable result was an asymptotic  $P$ -value of 0.00006 at D9S164 (149.01 cM) for age-at-first word analysis (Figure 7a; Table S7). However, re-simulation of chromosome 9 age-at-first word results gave an empirical  $P$ -value of 0.0008 at this same position, indicating that the significance of the asymptotic result in this case is likely to be inflated (Figure 5c; Table S7). The next most positive asymptotic  $P$ -value for age at first word was  $P=0.005$  at D5S647 (78.64 cM). For age at first phrase, the strongest asymptotic signals were  $P=0.0012$  at D18S70 (124.18 cM),  $P=0.002$  at D22S423 (55.5 cM),  $P=0.003$  at D7S2465 (183.3 cM),  $P=0.005$  at D1S206



**Figure 6** Genome-wide  $P$ -values for regression stratified groups. The top panels (a, b) are for regression-positive families, and the bottom panels (c, d) are for regression-negative families. The left panels (a, c) are for the strict diagnosis, and the right panels (b, d) are for the broad diagnosis. Chromosomes are noted by numbers along the X-axis, with tic-marks denoting additional chromosomes.



**Figure 7** Genome-wide  $P$ -values for variance-components analysis of age at first word (a) and age at first phrase (b). Chromosomes are noted by numbers along the X-axis, with tic-marks denoting additional chromosomes.

(134.31 cM),  $P=0.005$  at D11S1320 (154.48 cM) and  $P=0.006$  at D6S1581 (171.19 cM) (Figure 7b, Table S7). Note that the chromosome 7 signal at 183.63 cM for age at first phrase is 50 cM distal to the chromosome 7 signal from the primary genome scan at 133.16 cM (Figure 3a and b). This language signal overlaps with positive results obtained by Alarcon *et al.*<sup>28</sup> for the same trait.

There was only modest support for overlap between evidence for linkage for the language phenotypes and the primary scan. There is only a weak signal for age at first phrase ( $P=0.1$ ) at approximately the same chromosome 7 location as was obtained for the most significant result from the primary scan (Figure 3a and b). However, there is a modest signal from both language measures ( $P=0.03$  for word,  $P=0.04$  for phrase) that overlaps part of the chromosome 4 signal (Figure 3c and d).

## Discussion

The results presented here provide replication of a number of linkage findings for autism reported by

others. The most striking replication is for chromosome 7q, where the IMGSAC finds a strong signal at  $\sim 133\text{--}141$  cM<sup>17</sup> on the Rutgers genetic map. Modest evidence for a chromosome 7 locus at this approximate location was also observed in four other linkage analyses,<sup>15,18,19,23</sup> but not in three other studies.<sup>13,12,20</sup> Meta-analyses of genome scan data from a subset of these studies, but not including our data, provides additional evidence for a chromosome 7q autism locus.<sup>46,47</sup> Autism candidate genes in the vicinity of our peak centered at  $\sim 133$  cM include Reelin ( $\sim 112\text{--}114$  cM),<sup>48</sup> FOXP2 ( $\sim 123$  cM),<sup>49</sup> RAY1 ( $\sim 126$  cM),<sup>50</sup> and WNT2 ( $\sim 126$  cM),<sup>51</sup> but not AUTS2 (82.16 cM)<sup>52</sup> or EN2 (180.84 cM)<sup>53</sup> (Figure 3a and b). Linkage and association studies in non-autism groups suggest that there may be a chromosome 7 gene(s) for specific language impairment<sup>54</sup> and dyslexia<sup>55</sup> at  $\sim 120$  and  $\sim 133$  cM, respectively, potentially related to autism and the signal we observe. Other regions where our results overlap with those of other groups include chromosomes 10p ( $\sim 13\text{--}31$  cM),<sup>14,17</sup> 4,<sup>15,20</sup> 19p<sup>21</sup> and 3.<sup>14</sup> Notably absent were signals at 2q ( $\sim 177\text{--}186$  cM) reported by two groups<sup>13,17</sup> and 17q ( $\sim 61$  cM).<sup>14</sup>

The rationale for our gender stratification is based on the high male:female ratio ( $\sim 3\text{--}4/1$ ) for autism individuals, and on other linkage studies where families were divided into MO and FC groups.<sup>17,20</sup> We replicated the finding of an FC group chromosome 4 signal,<sup>20</sup> and a modest bias from the MO group for the chromosome 7 signal,<sup>17</sup> but no MO derived chromosome 17 signal.<sup>14,20</sup> We also observed in our MO family results a unique signal at 11q13.4 (84 cM) that was not present in the FC group results and only partially overlapped with a small signal at 59 cM from the primary genome scan (Figure 1). The finding of linkage signals dependent on the gender of affected individuals raises the intriguing possibility that there are male- and female-specific genes, or more likely, that penetrance of some susceptibility genes is significantly different in male subjects versus female subjects. Thus, the genes responsible for the putative chromosome 11 signals might not cause autism

susceptibility, or might have greatly reduced penetrance, in female subjects. Conversely, the gene responsible for the chromosome 4 signal from the FC group would presumably act in both sexes, but with a relatively higher penetrance in female than male subjects compared to other loci. Because there are very few female-only families in our sample in the FC group (8/74), it is unlikely that this signal represents a gene that only functions in female subjects. Since autism is a developmental disorder, these genes may affect or regulate the development of sexual dimorphism in the brain and in behavior.<sup>56</sup>

The rationale for examining regression was based on previous work by Molloy *et al.*,<sup>27</sup> where 34 families were identified with at least two affected individuals who showed regression. Signals in this earlier study were observed for chromosome 7 at 170.14 cM and chromosome 21 at 16.16 cM. We did not see linkage evidence at these locations in our regression-positive group. However, because we had very few families in which there were two individuals with regression, we used a more relaxed criterion for our regression-positive group, requiring only a single individual with regression for a family to be included in this group. Thus our work is not a direct replication of the Molloy *et al.* study.<sup>27</sup> Perhaps the most intriguing result from our regression studies comes from the regression-negative strata. When only regression-negative families were considered, evidence for linkage on both chromosomes 4 and 3 increases by one and two orders of magnitude, respectively, when compared to results from using all the families (Tables S2 and S6; Figures 1, 3, 5 and 6). This work suggests differences in the genetics underlying regression versus the genetics of the early developmental delay that leads to autism in non-regression subjects.

The language acquisition work provides evidence for a novel locus affecting age at first word on chromosome 9, with weaker evidence for additional genes elsewhere. The significance of these signals must be interpreted with caution since empirical estimates of *P*-values for chromosome 9 were substantially less significant than were the asymptotic *P*-values (Figure 5c). Practical constraints prevented computation of empirical *P*-values for all suggestive locations. We also found supporting evidence for a language acquisition locus on chromosome 7 initially reported by Alarcon *et al.*<sup>28</sup> at 151.31–178.06 cM, with a fairly strong signal based on age at first phrase (Figure 3a and b). This region does not overlap with the primary chromosome 7 peak from our study and from other studies (Figures 1 and 3a, b), but does include *Engrailed 2* (*EN2*) at 180 cM, an autism candidate gene.<sup>53</sup> We did not find significant supporting evidence for language-related signals on other chromosomes (2, 3, 7q (~103 cM) or 13) that were identified in previous linkage studies.<sup>13,29,30</sup>

In comparing our results to other studies, we provide strong support for a chromosome 7 autism locus, and lesser but still compelling support for loci on chromosomes 3, 4 and 11. However, we were not

able to reproduce a number of substantial signals reported by others. There are a number of potentially confounding factors that may explain these discordant findings. First, not all studies have the large sample sizes needed to detect all true autism loci, and many of the earlier studies are based on fairly small sample sizes for which asymptotic assumptions are less reliable than for larger samples. Second, some of the signals reported, including some here, may be false-positive results. Likewise, some of the smaller signals reported may be true autism loci. Only attempts to replicate results across multiple data sets can resolve these issues.

A third possible source of discordant findings may be differences across studies in trait definition and pedigree ascertainment. Use of a BPH in genetic studies is justified by multiple studies showing that close relatives of individuals with autism have symptoms related to autism,<sup>8</sup> and both twin and family sib studies support using a BPH definition in genetic studies.<sup>57</sup> However, while sample size increases with a BPH, so does the possibility of phenocopies, which have a strong negative effect on power to detect linkage. We attempted to minimize heterogeneity within our sample by using a rigorous diagnostic protocol and by using a data set with affected individuals meeting the strictest criteria for autism. The combined use of the ADR-I, ADOS-G and a clinical diagnosis has the advantage of increased accuracy in distinguishing between autism and closely related phenotypes, therefore reducing the phenocopy rate and consequent heterogeneity. For the chromosome 7 region, which is the most consistently replicated signal across studies, using the strict diagnostic group produced a more significant result when compared to the broad category, even though the latter was a larger sample. This is consistent with the hypothesis that the broader diagnosis includes more heterogeneity with respect to linkage to chromosome 7 than does the strict diagnosis. However, for some of the other signals, results were more significant for the broad diagnosis than for the strict diagnosis. These regions may represent loci for which the phenotype is, in fact, broader than the strict autism diagnosis, and thus restricting the sample size may have reduced the evidence in favor of linkage by eliminating some families or individuals that, in fact, contribute to the evidence for linkage.

A fourth difference among studies may be ascertainment of cohorts with somewhat different characteristics. Our study, with its requirement for two cases with a strict diagnosis of autism in our strict sample, may represent a somewhat different underlying sample of autism mutations than do samples with less stringent diagnostic criteria for inclusion. From the work presented here, clearly differences across studies in gender composition, language and regression can affect linkage results. Other differences such as the IQ distribution, family size and ethnic origins may also be important. For example, the Auranen *et al.* study<sup>12</sup> used Finnish families that had

a predominance of male affected individuals, and replicating the chromosome 3 signal at 186.55 cM in this group reported may not be possible with other samples.

Identification of autism genes may require several different approaches. Larger sample size for linkage identification and replication will be needed, though if heterogeneity is increased, there will be no advantage or a disadvantage to pooling more families. Instead, it may be useful to identify phenotypic subgroups from within a larger pooled sample. In addition, endophenotype analysis may permit mapping of loci contributing to specific autism domains that will be difficult to genetically map using only diagnosis as the trait. While some useful information can be obtained from the diagnostic tools such as the ADI-R and the ADOS-G, other instruments such as the Broader Phenotype Autism Symptom Scale (BPAS),<sup>58,59</sup> designed specifically for quantitative traits related to autism in both affected and normal individuals, are needed so the phenotypic information from all individuals in the pedigrees can be used to increase power to detect linkage.<sup>60</sup>

## Acknowledgments

This research was supported by a program project grant from the National Institute of Child Health and Human Development and the National Institute on Deafness and Communication Disability (PO1HD34565), which is part of the NICHD/NIDCD Collaborative Program of Excellence in Autism. We gratefully acknowledge the contributions of the Diagnostic and Statistical Cores of the UW Autism Program Project, Cathy Brock who assisted in recruitment of participants, Hiep Nguyen for computer support, and the parents and their children who participated in this study. This work was also supported in part by the Veterans Affairs Administration (GDS). Work at the University of Utah was supported by NICHD Grant HD35476, Grant MO1-RR00064 from the National Center for Research Resources, the Devonshire Foundation and the Utah Autism Research Foundation.

## Electronic-Database Information

The URLs for data presented herein are as follows:

Online Mendelian Inheritance in Man (OMIM), <http://www.ncbi.nlm.nih.gov/Omim>.

Rutgers Genetic Map (MAP-O-MAT), <http://comp-gen.rutgers.edu/mapomat>.

## References

- 1 Werner E, Dawson G. Validation of the phenomenon of autistic regression using home videotapes. *Arch Gen Psychiatry* 2005; **62**: 889–895.
- 2 Folstein S, Rutter M. Infantile autism: a genetic study of 21 twin pairs. *J Child Psychol Psychiatry* 1977; **18**: 297–321.
- 3 Ritvo ER, Freeman BJ, Mason-Brothers A, Mo A, Ritvo AM. Concordance for the syndrome of autism in 40 pairs of afflicted twins. *Am J Psychiatry* 1985; **142**: 74–77.

- 4 Folstein S, Rutter M. Genetic influences and infantile autism. *Nature* 1977; **265**: 726–728.
- 5 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.
- 6 Steffenburg S, Gillberg C, Heggren L, Andersson L, Gillberg I, Jakobsson G et al. A twin study of autism in Denmark, Finland, Iceland, Norway, and Sweden. *J Child Psychol Psychiatry* 1989; **30**: 405–416.
- 7 August GJ, Stewart MA, Tsai L. The incidence of cognitive disabilities in the siblings of autistic children. *Br J Psychiatry* 1981; **138**: 416–422.
- 8 Bailey A, Phillips W, Rutter M. Autism: towards an integration of clinical, genetic, neuropsychological, and neurobiological perspectives. *J Child Psychol Psychiatry* 1996; **37**: 89–126.
- 9 Smalley SL, Asarnow RF, Spence A. Autism and genetics. *Arch Gen Psychiatry* 1988; **45**: 953–961.
- 10 Delong GR, Dwyer JT. Correlation of family history with specific autistic subgroups: Aspergers and bipolar affective disease. *J Autism Dev Disord* 1988; **18**: 593–600.
- 11 Vorstman JA, Staal WG, Hochstenbach PF, Franke L, van Daalen E, van Engeland H. Overview of cytogenetic regions of interest (CROIs) associated with the autism phenotype across the human genome. *Mol Psychiatry* 2006; **11**: 18–28.
- 12 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.
- 13 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.
- 14 Cantor RM, Kono N, Duvall JA, AlvarezRetuerto A, Stone JL, Alarcon M et al. Replication of autism linkage: fine-mapping peak at 17q21. *Am J Hum Genet* 2005; **76**: 1050–1056.
- 15 Barrett S, Beck JC, Bernier R, Bisson E, Braun TA, Casavant TL et al. An autosomal genomic screen for autism. *Am J Med Genet* 1999; **88**: 609–615.
- 16 International Molecular Genetic Study of Autism Consortium. A genomewide screen for autism: strong evidence for linkage to chromosomes 2q, 7q, and 16p. *Am J Hum Genet* 2001; **69**: 570–581.
- 17 Lamb JA, Barnby G, Bonora E, Sykes N, Bacchelli E, Blasi F et al. Analysis of IMGSAC autism susceptibility loci: evidence for sex limited and parent of origin specific effects. *J Med Genet* 2005; **42**: 132–137.
- 18 Philippe A, Martinez M, Guilloudbataille M, Gillberg C, Rastam M, Sponheim E et al. Genome-wide scan for autism susceptibility genes. *Hum Mol Genet* 1999; **8**: 805–812.
- 19 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.
- 20 Stone JL, Merriman B, Cantor RM, Yonan AL, Gilliam TC, Geschwind DH et al. Evidence for sex-specific risk alleles in autism spectrum disorder. *Am J Hum Genet* 2004; **75**: 1117–1123.
- 21 McCauley JL, Li C, Jiang L, Olson LM, Crockett G, Gainer K et al. Genome-wide and Ordered-Subset linkage analyses provide support for autism loci on 17q and 19p with evidence of phenotypic and interlocus genetic correlates. *BMC Med Genet* 2005; **6**: 1.
- 22 Liu JJ, Nyholt DR, Magnussen P, Parano E, Pavone P, Geschwind D et al. A genomewide screen for autism susceptibility loci. *Am J Hum Genet* 2001; **69**: 327–340.
- 23 Shao YJ, 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.
- 24 Yonan AL, Alarcon M, Cheng R, Magnusson PKE, Spence SJ, Palmer AA et al. A genomewide screen of 345 families for autism-susceptibility loci. *Am J Hum Genet* 2003; **73**: 886–897.
- 25 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.
- 26 Wassink TH, Brzustowicz LM, Bartlett CW, Szatmari P. The search for autism disease genes. *Mental Retard Dev Disabil Res Rev* 2004; **10**: 272–283.

- 27 Molloy CA, Keddache M, Martin LJ. Evidence for linkage on 21q and 7q in a subset of autism characterized by developmental regression. *Mol Psychiatry* 2005; **10**: 741–746.
- 28 Alarcon M, Cantor RM, Liu J, Gilliam TC, the Autism Genetic Resource Exchange Consortium, Geschwind DH. Evidence for a language quantitative trait locus on chromosome 7q in multiplex families. *Am J Hum Genet* 2002; **70**: 60–71.
- 29 Alarcon 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.
- 30 Bradford Y, Braun T, Cassavant T, Folstein S, Haines J, Hutcheson H *et al*. Incorporating language phenotypes strengthens evidence of linkage to autism. *Am J Med Genet* 2001; **105**: 539–547.
- 31 Buxbaum JD, Silverman J, Keddache M, Smith CJ, Hollander E, Ramoz N *et al*. Linkage analysis for autism in a subset families with obsessive-compulsive behaviors: evidence for an autism susceptibility gene on chromosome 1 and further support for susceptibility genes on chromosome 6 and 19. *Mol Psychiatry* 2004; **9**: 144–150.
- 32 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.
- 33 Lord C, Risi S, Lambrecht L, Cook EH, Leventhal BL, DiLavore P *et al*. The Autism Diagnostic Observation Schedule-Generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J Autism Develop Disord* 2000; **30**: 205–223.
- 34 American Psychiatric Association. *Diagnostic and statistical manual of mental disorders*, 4th edn. American Psychiatric Association: Arlington, VA, 1994, pp 65–78.
- 35 Mullen EM. *Mullen Scales of Early Learning*. Western Psychological Services: Los Angeles, CA, 1997.
- 36 Sattler J. *Assessment of Children: Cognitive Applications*. Jerome Sattler, Publisher, Inc.: La Mesa, CA, 1992.
- 37 Epstein MP, Duren WL, Boehnke M. Improved inference of relationship for pairs of individuals. *Am J Hum Genet* 2000; **67**: 1219–1231.
- 38 Heath SC. Markov chain Monte Carlo segregation and linkage analysis for oligogenic models. *Am J Hum Genet* 1997; **61**: 748–760.
- 39 Abecasis G, Cherny S, Cookson W, Cardon L. Merlin-rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 2002; **30**: 97–101.
- 40 Kong X, Murphy K, Raj T, He C, White PS, Matise TC. A combined linkage-physical map of the human genome. *Am J Hum Genet* 2004; **75**: 1143–1148.
- 41 Whittemore AS, Halpern J. A class of tests for linkage using affected pedigree members. *Biometrics* 1994; **50**: 118–127.
- 42 Kong A, Cox NJ. Allele-sharing models. LOD scores and accurate linkage tests. *Am J Hum Genet* 1997; **61**: 1179–1188.
- 43 Amos C. I. Robust variance-components approach for assessing genetic linkage in pedigrees. *Am J Hum Genet* 1994; **54**: 535–543.
- 44 Daw E, Thompson E, Wijsman E. Bias in multipoint linkage analysis arising from map misspecification. *Genet Epidemiol* 2000; **19**: 336–380.
- 45 Shao YJ, 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.
- 46 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.
- 47 Trikalinos TA, Karvouni A, Zintzaras E, Ylisaukkoja T, Peltonen L, Jarvela I *et al*. A heterogeneity-based genome search meta-analysis for autism-spectrum disorders. *Mol Psychiatry* 2006; **11**: 29–36.
- 48 Persico AM, D'Agruma L, Malorano N, Totaro A, Militerni R, Bravaccio C *et al*. Reelin gene alleles and haplotypes as a factor predisposing to autistic disorder. *Mol Psychiatry* 2001; **6**: 150–159.
- 49 Lai CSL, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP. A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* 2001; **413**: 519–523.
- 50 Vincent JB, Herbrick J-A, Gurling HMD, Bolton PF, Roberts W, Scherer SW. Identification of a novel gene on chromosome 7q31 that is interrupted by a translocation breakpoint in an autistic individual. *Am J Hum Genet* 2000; **67**: 510–514.
- 51 Wassink TH, Piven J, Vieland VJ, Huang J, Swiderski RE, Pietila J *et al*. Evidence supporting WNT2 as an autism susceptibility gene. *Am J Med Genet* 2001; **105**: 406–413.
- 52 Sultana R, Yu C-E, Yu J, Munson J, Chen D, Hua W *et al*. A candidate gene for autism on chromosome 7. *Genomics* 2002; **80**: 129–134.
- 53 Benayed R, Gharani N, Rossman I, Mancuso V, Lazar G, Kamdar S *et al*. Support for the homeobox transcription factor gene ENGRAILED 2 as an autism spectrum disorder susceptibility locus. *Am J Hum Genet* 2005; **77**: 851–868.
- 54 O'Brien EK, Zhang XY, Nishimura C, Tomblin JB, Murray JC. Association of specific language impairment (SLI) to the region of 7q31. *Am J Hum Genet* 2003; **72**: 1536–1543.
- 55 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.
- 56 Baron-Cohen S, Knickmeyer RC, Belmonte MK. Sex differences in the brain: implications for explaining autism. *Science* 2005; **310**: 819–823.
- 57 Le Couteur A, Rutter M, Lord C, Rios P, Robertson S, Holdgrafer M *et al*. Autism diagnostic interview: a standardized investigator-based instrument. *J Autism Dev Disord* 1989; **19**: 363–387.
- 58 Sung YJ, Dawson G, Munson J, Estes A, Schellenberg GD, Wijsman EM. Genetic investigation of quantitative traits related to autism: use of a multivariate polygenic models with ascertainment adjustment. *Am J Hum Genet* 2005; **76**: 68–81.
- 59 Dawson G, Estes A, Munson J, Schellenberg GD, Bernier R, Abbott R *et al*. Quantitative assessment of autism symptoms in autism probands and parents: Broader Phenotype Autism Symptom Scale. *J Autism Dev Disord* 2006 (in press).
- 60 Wijsman E, Amos C. Genetic analysis of simulated oligogenic traits in nuclear and extended pedigrees: summary of GAW10 contributions. *Genet Epidemiol* 1997; **14**: 719–735.

Supplementary Information accompanies the paper on the Molecular Psychiatry website (<http://www.nature.com/mp>)