Am. J. Hum. Genet. 65:493–507, 1999

A Genomic Screen of Autism: Evidence for a Multilocus Etiology

Neil Risch,[1] Donna Spiker,[2] Linda Lotspeich,[2] Nassim Nouri,[1] David Hinds,[1] Joachim Hallmayer,[3] Luba Kalaydjieva,[4] Patty McCague,[2] Sue Dimiceli,[2] Tawna Pitts,[1] Loan Nguyen,[1] Joan Yang,[1] Courtney Harper,[1] Danielle Thorpe,[2] Saritha Vermeer,[2] Helena Young,[2] Joan Hebert,[1] Alice Lin,[1] Joan Ferguson,[2] Carla Chiotti,[2] Susan Wiese-Slater,[2] Tamara Rogers,[4] Boyd Salmon,[4] Peter Nicholas,[5] P. Brent Petersen,[5] Carmen Pingree,[5] William McMahon,[5] Dona L. Wong,[2] L. Luca Cavalli-Sforza,[1] Helena C. Kraemer,[2] and Richard M. Myers[1]

Departments of [1]Genetics and [2]Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford; [3]Centre for Clinical Research in Neuropsychiatry, Graylands Hospital/University of Western Australia, and [5]Centre for Human Genetics, Edith Cowan University, Perth; and [6]Department of Psychiatry, University of Utah, Salt Lake City

Summary

We have conducted a genome screen of autism, by linkage analysis in an initial set of 90 multiplex sibships, with parents, containing 97 independent affected sib pairs (ASPs), with follow-up in 49 additional multiplex sibships, containing 50 ASPs. In total, 519 markers were genotyped, including 362 for the initial screen, and an additional 157 were genotyped in the follow-up. As a control, we also included in the analysis unaffected sibs, which provided 51 discordant sib pairs (DSPs) for the initial screen and 29 for the follow-up. In the initial phase of the work, we observed increased identity by descent (IBD) in the ASPs (sharing of 51.6%) compared with the DSPs (sharing of 50.8%). The excess sharing in the ASPs could not be attributed to the effect of a small number of loci but, rather, was due to the modest increase in the entire distribution of IBD. These results are most compatible with a model specifying a large number of loci (perhaps $\geq$15) and are less compatible with models specifying $\leq$10 loci. The largest LOD score obtained in the initial scan was for a marker on chromosome 1p; this region also showed positive sharing in the replication family set, giving a maximum multipoint LOD score of 2.15 for both sets combined. Thus, there may exist a gene of moderate effect in this region. We had only modestly positive or negative linkage evidence in candidate regions identified in other studies. Our results suggest that positional cloning of susceptibility loci by linkage analysis may be a formidable task and that other approaches may be necessary.

Received December 23, 1998; accepted for publication May 19, 1999; electronically published July 9, 1999.

Address for correspondence and reprints: Dr. Neil Risch, Department of Genetics, M322, Stanford University School of Medicine, Stanford, CA 94305-5120. E-mail: Risch@lahmed.stanford.edu

0002-9297/99/6502-0025$02.00

Introduction

Autism (MIM 209850) is a pervasive neurodevelopmental disorder with symptoms usually apparent during the first 3 years of life. It is characterized by a triad of features, including limited or absent verbal communication, a lack of social reciprocity or responsiveness, and limited, stereotypical, and ritualized interests and behaviors. Autism is the most severe of the disorders now characterized as "pervasive developmental disorders" (PDD) (MIM 209850), which also include Asperger syndrome (MIM 209850) and PPD that is not otherwise specified (PDD-NOS).

Autism was first typified as a distinct clinical entity by Kanner (1943). The etiology of this syndrome has been debated ever since, with various proponents favoring behavioral, environmental, dietary, viral/immunologic, autoimmune, or genetic theories. On the basis of the prevalence studies that have been performed, the indication is that autism is generally rare, with a mean prevalence of 1/2,500 children (Smalley et al. 1988).

The infrequency of familial cases of autism probably led some early theorists to doubt an inherited component in this disorder (Hanson and Gottesman 1976). For example, for a typical rare recessive disease, multiple-incidence sibships are not unusual, since the recurrence risk in sibships is 25%. The recurrence risk of autism in sibships is far below that 25% figure. However, with the completion of several systematic family studies, it is now possible to provide a consistent sibling recurrence figure of 2%–6%, which is substantially higher than the population prevalence of 1/2,500, or 0.04% (Smalley et al. 1988; Bolton et al. 1994). The most recent summary of family studies provided an overall sibling recurrence risk of 2.2% (13/597), with a 95% confidence limit of 1.1%–3.3% (Szatmari et al. 1998). Thus, although familial cases are rare, they are substantially more frequent than would be predicted to occur by chance.

Further supporting a genetic etiology in autism are the results of twin studies, which have documented a dramatically higher concordance rate in MZ twins than in DZ twins (Folstein and Rutter 1977). Summarizing the largest and most systematic of these studies (Steffenburg et al. 1989; Bailey et al. 1995), we can derive MZ-twin and DZ-twin probandwise concordances of 81% (47/58) and 0%, respectively, for autism. The latter figure is likely to be an underestimate as a result of small sample sizes, and the true DZ-twin concordance is likely to be similar to the nontwin-sib rate of 2%–6%. Using a 3% rate gives an MZ-twin concordance:DZ-twin concordance ratio (MZ:DZ concordance ratio) of ~25-fold. We note that, whereas the sib recurrence is based on numerous studies and a sizable total sibling population ($n = 597$), the MZ-twin concordance rate is based on a much smaller sample and thus is more susceptible to statistical fluctuation. Nonetheless, it is likely that the MZ:DZ concordance ratio in autism is quite large.

Autistic features have also been described in individuals with well-characterized genetic disorders, most notably fragile X syndrome (FMR1 [MIM 309550]) (Feinstein and Reiss 1998) and tuberous sclerosis (TSC1 [MIM 191100]) (Smalley 1998). In addition, case studies have been reported that show associations between autism and a variety of isolated chromosomal translocations, inversions, and deletions (Gillberg 1998). These findings provide evidence that genetic abnormalities can, in some cases, produce behaviors typical of autism, but they account for only a small percentage of cases and do not explain the familial aggregation of autism.

Autism does not follow a simple mode of inheritance. Genealogical studies suggested an increased risk primarily in close relatives (sibs) (Jorde et al. 1990), and a reported segregation analysis was most consistent with a polygenic model of inheritance (Jorde et al. 1991). These results suggest that family recurrence is unlikely to be explained by a simple, monogenic mechanism.

The first linkage study in autism was performed by Spence et al. (1985), using 20 classic blood-group and serum-protein markers on 34 multiplex sibships and assuming a recessive model. No significant linkage findings were obtained.

Recent advances in molecular biology have allowed genome surveys for disease-susceptibility loci to be performed on a much larger scale. For example, Hallmayer et al. (1996) examined 35 microsatellite markers on the X chromosome in 38 multiplex autism families. Although they excluded most of the chromosome from carrying a gene of large effect, including the fragile X region, they found modest evidence (maximum LOD score 1.24) for linkage in one region on proximal Xq.

A genome screen of the full genome, using 316 microsatellite markers on 39 multiplex families, has also been reported recently (International Molecular Genetic Study of Autism Consortium 1998). Initial positive findings were followed up in an additional 60 families. No statistically significant findings were obtained. The most positive finding was a LOD score of 2.53, on chromosome 7q, with the next most positive finding being a LOD score of 1.97, near the telomere of chromosome 16p.

In this report we describe the results of what, to date, is the largest genome screen in autism. We have surveyed 90 multiplex sibships with 97 independent affected sib pairs (ASPs) in an initial analysis of 362 microsatellite markers. Regions with positive results, as well as other candidate regions, were then followed up with additional markers, as well as by analysis of a second set of 49 multiplex sibships containing 50 independent ASPs. A total of 519 markers was analyzed in the first set of families, and 149 of these markers were also analyzed in the second set of families.

Families And Methods

Family Recruitment and Diagnostic Assessments

Families were recruited nationwide for this study, by advertisement with local and national parent support groups and by referrals from clinicians. Initial intake criteria required that at least two siblings have a clinical diagnosis of a PDD. If both the initial telephone intake interview of parents and collected medical records were consistent with a presumptive diagnosis of autism in at least two children, then a follow-up visit by a diagnostician was arranged with the family. During these visits, the children were assessed—both by interview, of one or both parents, using the Autism Diagnostic Interview (ADI) (LeCouteur et al. 1989; Lord et al. 1994, 1997), and by observation of the children, using the Autism Diagnosis Observation Schedule (ADOS) (Lord et al. 1989)—to determine a research diagnosis of autism. The ADI is a scored, semistructured interview of parents that is based on ICD-10 and DSM-IV criteria for the diagnosis of autism. The ADOS is a semistructured instrument, which allows assessment of the child through observation of his or her behavior and which is used to corroborate the results of the ADI interview. Diagnostic assessments (by ADI and ADOS) were videotaped and subjected to independent reliability checks by other trained interviewers (typically, the senior clinical investigators). To be considered affected in the linkage analysis, an individual had to satisfy the prespecified cutoff scores in all three symptom areas of the ADI (social impairment, language and communication impairment, and unusual and restricted interests) as well as have an age at onset of <3 years. In addition, review of the ADOS tape by two or more diagnosticians was used to exclude children who did not show significant impairments in

Table 1

Distribution of Family Structures

Group and No. of Affected Sibs	No. of Unaffected Sibs	No. of Available Parents	Families
FS1:			
2	0	2	48
2	0	1	15
2	1	2	13
2	1	1	5
2	2	2	2
2	3	1	2
3	0	2	2
3	1	2	1
4	0	2	1
4	1	1	1
FS2:			
2	0	2	34
2	0	1	1
2	1	2	12
2	2	2	1
3	1	2	1

social and communicative reciprocity, irrespective of the results of the ADI interview (e.g., this would lead to the exclusion of some individuals who would have had a clinical diagnosis of PDD-NOS). Families were excluded from subsequent analysis if, on the basis of all available information, there was no consensus that at least two affected children had deficits consistent with a diagnosis of autism. For more details on the diagnostic protocol and reliability of videotapes of the ADI assessments, see the work of Spiker et al. (1994).

Children's records were reviewed to obtain IQ and mental-age (MA) information. When available, nonverbal IQ scores were obtained from performance subtests of either the Wechsler Scales or the Stanford-Binet Intelligence Scale-4th Edition, from the Leiter International Performance Scale, or from the Merrill-Palmer Scale. When unavailable, scores from such tests as the Stanford-Binet Intelligence Scale-3rd Edition, Slossen Intelligence Scale, or McCarthy Scales were used. For children for whom results of these tests were not available, MAs and ratio IQ scores were derived from nonverbal scales of various developmental instruments (e.g., Daily Living Scale of the Vineland Adaptive Behavior Scale, the Bayley Mental Scale, and Developmental Inventory II).

For inclusion in the linkage analysis, we also required that at least one affected child have both an MA of $\geq$18 mo and a nonverbal IQ $\geq$30. This led to the exclusion of some additional families ($N = 6$) in which all children were severely impaired (MA <18 mo and IQ <30).

Two-Stage Analysis

Families were analyzed in two stages. The initial stage involved genotyping of family set 1 (FS1), which included 90 multiplex families. These families comprise a total of 187 affected and 30 unaffected sibs; 67 families have two parents available, and the remaining 23 have one parent. The precise breakdown of family structures is given in table 1. Among the affected sibs, the male:female ratio is 3.6:1, whereas in the unaffected sibs it is 0.8:1.

Subsequent to the initial genome screen performed on FS1, a follow-up study was conducted on a second family set (FS2), which comprises a total of 49 multiplex families, ascertained in a manner similar to that used with FS1. These families contain a total of 99 affected and 15 unaffected sibs; 48 of the 49 families have both parents available, whereas the single remaining family has one parent. The precise breakdown, by family structure, is given in table 1. The male:female ratio in the affected sibs is 3.0:1, versus 0.7:1 in the unaffected sibs.

Laboratory Procedures

Blood was collected from all affected individuals and, if available, from their parents and unaffected siblings. Blood samples were drawn into Vacutainer sodium heparin (green cap) tubes. The lymphoblasts were isolated and immortalized with fresh or frozen virus stocks, by standard Epstein-Barr–virus transformation protocols (Anderson and Gusella 1984). The DNA was extracted from both whole blood and immortalized lymphoblast cell lines, by standard proteinase K–digestion and salting-out procedures.

Genotyping was performed in two laboratories in-

Table 2

Markers with Highest Sharing in ASPs and DSPs in Initial Genome Screen

Marker	No. (%) Sharing	Z Score
ASPs:		
D1S1631	80/122 (65.6)	3.44
D13S779	64/101 (63.4)	2.69
D9S282	60/95 (63.2)	2.66
D1S1609	79/130 (60.8)	2.46
D3S2418	60/96 (62.5)	2.45
D17S1298	51/81 (63.0)	2.33
D15S652	79/132 (59.8)	2.26
D10S1412	53/86 (61.6)	2.16
D1S534	81/137 (59.1)	2.14
DSPs:		
GATA145D0	34/49 (69.4)	2.71
D16S403	42/63 (66.7)	2.65
D18S844	39/60 (65.0)	2.32
D16S3253	29/43 (67.4)	2.29
D7S1819	35/54 (65.3)	2.18
D4S2361	32/49 (65.3)	2.08
D11S2371	26/39 (66.7)	2.08
D4S2366	34/53 (64.2)	2.06
D20S482	31/48 (64.6)	2.02
D16S769	32/50 (64.0)	1.98

dependently, according to somewhat different procedures. In one lab, PCR was performed, by Perkin-Elmer 9700 thermal cyclers, in a 10-μl volume in 96-well plates and included 30 ng of genomic DNA, 67 mM Tris-HCl pH 8.8, 200 μM of each dNTP, 16 mM $(NH_4)_2SO_4$, 0.01% Tween-20, 1.5–3 mM $MgCl_2$, 0.2 μM of each primer, and 0.2 units of Ampli*Taq* polymerase (Perkin-Elmer). The forward primers for each marker were fluorescently labeled with FAM, HEX, or TET dyes. The reaction was first denatured for 75 s at 95°C; then, 35 cycles of 45 s at 95°C, 45 s at 57° C, and 60 s at 72° C were performed, followed by a final elongation step of 7–12 min at 72° C. PCR products were pooled from ~7–15 markers that were separable on the basis of size and dye color and were run on ABI Prizm 377 DNA sequencers. Tamara-labeled DNA internal-size standards were run in each lane, and the products were resolved by the GeneScan and Genotyper software packages from Applied Biosystems. Three CEPH control DNAs from individuals 1331-01, 1331-02, and 1347-02 were used to verify the sizes of PCR products for each marker.

In the second lab, PCR assays were performed in 8-μl reactions containing 10 ng of genomic DNA; 2.5 mM of each dNTP; 0.2–0.6 mM of each primer pair labeled with the fluorescent dyes FAM, HEX, or TET; 0.8 μl of 10 × buffer (Perkin-Elmer); 1.5–2.5 mM $MgCl_2$, and 0.2 units of Ampli*Taq* Gold Polymerase (Perkin-Elmer). PCR assays were performed in an Applied Biosystems 9600 Thermocycler by denaturation for 10 min at 94°C, 15 s at 60°C, and 15 s at 72°C. Another 20 cycles of 30 s at 94°C, 15 s at 65°C, and 15 s at 72°C were performed, and these were followed by a final extension step of 10 min at 72°C. Fluorescently labeled markers were analyzed on an Applied Biosystems 373 Genetic Analyzer, by Genescan and Genotyper software. In both labs, polymorphic bands were scored, and alleles were assigned to the pedigree members, by researchers who were blind to affection status.

Genetic Markers

Marker set 1 (MS1).—The genetic analysis was performed in two stages. The first stage consisted of a genome screen on the 90 multiplex FS1 families and used a total of 362 microsatellite markers. These were derived from version 8.0 of the Marshfield fluorescently labeled genome screening set (Center for Medical Genetics, Marshfield Medical Research Foundation) and were obtained from Research Genetics. Of these markers, 346 are autosomal, 14 are X linked, and 2 are pseudoautosomal. If a sex-averaged total autosomal map length of 3,500 cM is assumed, the average spacing between markers for the autosomal component is 10 cM; if a length of 180 cM for the X chromosome is assumed, the average spacing on that chromosome is 12 cM. Of the intermarker intervals, 28 have an interval length >15 cM, although only 7 are >20 cM; the largest gap is 23 cM.

Marker set 2 (MS2).—Additional microsatellite markers were obtained from Research Genetics and were analyzed subsequent to the initial genome screen. The choice of markers was based on (1) saturation of regions giving positive linkage evidence in the analysis of FS1 with MS1; (2) follow-up of specific regions, on the basis of either candidate loci (chromosome regions 6p and 15q) or the genome screen of the International Molecular Genetic Study of Autism Consortium (1998) (chromosome region 7q); and (3) addition of markers in regions that had large gaps in the initial screen with MS1 and that did not have significantly negative linkage evidence in the preliminary analysis. In total, 157 additional markers were typed, which constituted MS2. All of these markers were typed in FS1; 89 of these were also typed in FS2. In addition, 60 of the original MS1 markers were genotyped in FS2 in the follow-up, for reasons (1) and (2) above. Thus, in total, 519 markers were typed in FS1, and 149 were typed in FS2.

Statistical Analysis

Linkage analysis.—Because all siblings had at least one parent available and most had both parents available, the statistical analyses were based entirely on identity by descent (IBD), obviating the need to use allele frequencies of the parents. When one parent is missing, a sib pair can be scored for IBD with the included parent, without bias, provided that the typed parent is heterozygous and that each of the two sibs inherits from the untyped parent an allele that is distinct from the alleles of the typed parent. Also, sharing from the untyped parent was scored, when possible, by reconstruction of the genotype of the missing parent, by use of one affected sib and all unaffected sibs (in the ASP analysis). If the genotype of the parent can be so reconstructed, then unbiased scoring can be performed by comparison of the affected sib with each remaining affected sib, who was not used for reconstruction.

Sib-pair analyses were based on fully independent pairs. When a family contained three or more affected sibs, independent pairs were formed between the first sib and each remaining sib (giving $n - 1$ pairs for a family with n affected sibs). In the 90 multiplex FS1 families, there are thus a total of 97 independent ASPs; in the 49 multiplex FS2 families, there are 50 independent ASPs (see table 1).

Initially, the number of alleles shared versus the number of alleles unshared, at each marker across sib pairs, was evaluated and compared to a χ^2 distribution. Under the null hypothesis of no linkage, the expectation is that sharing will be 50%. Subsequently, multipoint sib-pair

analysis was performed to extract the maximum information from the genetic-marker data. This analysis is parameterized in terms of the value λ_S, the sibling recurrence ratio, which translates into specific values for ASP IBD. For a genetic model with no dominance variance, the probability z_0 of an ASP sharing no alleles IBD is $.25/\lambda_S$, the probability z_1 of an ASP sharing one allele IBD is .50, and the probability z_2 of an ASP sharing two alleles IBD is $.25(2 - \frac{1}{\lambda_S})$; the overall sharing is $y = .25(3 - \frac{1}{\lambda_S})$. For a multiplicative model which allows dominance variance, again $z_0 = \frac{.25}{\lambda_S}$, $z_1 = (\sqrt{\lambda_S} - \frac{1}{2})/\lambda_S$, and $z_2 = (1 - \frac{1}{2}\sqrt{\lambda_S})^2$; the overall sharing is $y = 1 - \frac{1}{2}\sqrt{\lambda_S}$. We use the latter model in all the multipoint analyses that we present, although, for modest gene effects, results are nearly identical in either model.

Multipoint LOD scores can be calculated for a fixed model (i.e., fixed value of λ_S); this approach allows for exclusion mapping. For a fixed value of λ_S, negative LOD scores can be obtained if the model gives a poorer representation of the marker data than does no gene effect. We use the LOD-score criterion of -2 for exclusion of a locus with a given λ_S value; this provides a conservative exclusion level (Hauser et al. 1996). Alternatively, the likelihood of the marker data can be maximized at each point, as a function of the genetic model (λ_S value), to derive a maximum LOD score (MLS) curve. This curve never drops to <0, the value obtained when $\lambda_S = 1$.

Linkage-disequilibrium analysis (LD).—We also examined LD, with all 519 autosomal and X-linked loci, using a transmission/disequilibrium test (TDT) (Spielman et al. 1993). For families with both parents, we scored directly transmissions to all affected children from heterozygous parents. As in the linkage analysis, we can still score children from single parents, provided that the parent is heterozygous and that the child is heterozygous for alleles different than those in the typed parent. We can also score transmissions from the missing parent when the genotype of that parent can be reconstructed on the basis of the unaffected sibs.

We examined two statistical tests—a "global χ^2" (GCS) test, which sums X = (observed−expected)^2/expected across each allele tested, and a "maximum χ^2" (MCS) test, which takes the maximum value of the statistic X across all tested alleles. In this case, neither statistic has a formal χ^2 distribution when multiple sibs from the same family are tested, if linkage is present (Spielman et al. 1993). To make our test of LD independent of any linkage present, we calculated the level of significance empirically, using Monte Carlo simulation; we simply randomly inverted or not, with probability 50%, the allele label for the two alleles of heterozygous parents and recalculated the two statistics for each permutation step. The proportion of simulated statistics that exceed the observed value is the derived empirical significance level. This procedure allows us to assess LD, conditional on the degree of sharing of alleles, at a given locus, by sibs from multiplex families (Martin et al. 1997; Lazzeroni and Lange 1998).

Computer program.—All statistical analysis, including the pointwise and multipoint sib-pair analyses and the TDT analyses, were performed with the ASPEX program package. We also used this package to confirm relationships among study subjects and to search for and eliminate unlikely double crossovers in the multipoint analyses, in regions that had dense markers.

Controls

In the course of this study, a number of controls were employed, to determine the degree to which potential biases might be confounding our results. After ~60 markers were run in FS1, we examined the distribution of IBD for all sib pairs, to confirm the presumed relationships, identify any mixup of samples, and identify any potential half-sibs. All families segregated appropriately according to presumed relationships (no half-sibs were identified), with the exception of two sib pairs, in which IBD was nearly 100%. These pairs were reported to us to be DZ twins, whereas our genotyping determined them to be MZ. We left these twin pairs in the analysis, both as a positive control and to determine the genotyping-error rate. The lab remained blind to the identity of these pairs.

As another control for our genetic analyses, we also included unaffected sibs in our molecular and statistical analyses. FS1 contains 30 unaffected sibs, which allowed us to construct 51 independent affected-unaffected sib pairs, or DSPs. These were derived by pairing, in each family, one unaffected sib with each affected sib. Remaining unaffected sibs were used for reconstruction of the genotypes of untyped parents. This selection scheme allowed us to tally IBD for these pairs as fully independent. Because of the typically modest recurrence risk of autism in sibships (Smalley et al. 1988), we assume that the DSPs should deviate little from the null expectation of 50% sharing—and that they thus can serve as a useful negative control.

Results

Family Recruitment

The first stage of analysis included the 90 multiplex FS1 families comprising 187 affected offspring, 30 unaffected sibs, and 157 parents. Two additional families with affected MZ twins (originally thought to be DZ) were also included. At the time of ADI administration, the age range of the affected siblings was 2.9–40.9 years (mean 9.8 years, SD 7.3 years, 25th–75th percentile 4.9–11.3 years). MA and IQ estimates were determined

on the basis of available diagnostic evaluations and school records. The mean nonverbal IQ was 66 (SD 28, range 16–160). One hundred eight subjects (57%) had IQ scores <70. The mean MA was 68 mo (SD 55 mo, range 13–373 mo). There were no families in which all affected children had IQ scores <30. In one family, one affected child had an MA of <18 mo.

The second set of study subjects, obtained with similar recruitment and inclusion/exclusion criteria, comprised a total of 99 affected children, 15 unaffected sibs, and 97 parents, from 49 families. For this group, at the time of ADI administration, the age range of affected subjects was 2.8–25.3 years (mean 6.6 years, SD 3.5 years, 25th–75th percentile 4.3–7.6 years). The mean nonverbal IQ was 61 (SD 23, range 15–124). Sixty-two subjects (63%) had IQ scores <70. Mean MA was 47 mo (SD 23 mo, range 13–150 mo). Again, there were no families in which all children had IQ scores <30, and there were three children with MA <18 mo.

The unaffected children were presumed to be developing normally, on the basis of parental report; all of these children were in regular educational programs and schools. None had any clinical diagnoses, according to parental report, nor, during telephone intake interviews, did parents report any developmental concerns. Children with a possible PDD diagnosis based on the ADI interview but who did not meet the ADI criteria for autism were categorized as being of uncertain status and were not included in this analysis. Eighty-eight percent of our sample was white (of varied European and Middle Eastern origin), 5% was black, 3% was Hispanic, and 4% was Asian. The vast majority of affected subjects were living with parents, and a small number were in residential living facilities.

FS1: Pointwise IBD Scores

A total of 346 autosomal, 14 X-linked, and 2 pseudoautosomal microsatellite loci were studied in FS1, which contained 97 independent ASPs and 51 independent DSPs. Tallying the total number of alleles shared across all 346 autosomal markers gives 19,902/38,572, or 51.6%, for the ASPs and 9,344/18,386, or 50.8%, for the DSPs. The slight increase in IBD, to >50%, in the DSPs suggests the possibility of some modest scoring bias toward allele sharing in sibships.

Of greatest interest is the right-hand tail of the observed IBD distribution in the ASPs and DSPs. For each marker, we calculated a Z score, $Z = (s - u)/\sqrt{s + u}$, where s is the number of shared alleles and u is the number of unshared alleles. Under the null hypothesis of no linkage, Z should have a normal distribution with mean 0 and variance 1. When a threshold of 1.96 ($P = .025$ for a one-sided Z test) is assumed, there are 9 loci in the ASP group that exceed this threshold and 10 loci in the DSP group that exceed this threshold, which are listed in table 2. These numbers are not different from the number expected ($.025 \times 360 = 9.0$) under the null hypothesis of no linkage. The most significant locus, in either group, was D1S1631, for the ASPs, with a sharing of 65.6% and a Z score of 3.44. The remaining loci in the ASP group were found on chromosomes 1 (two additional loci), 3, 9, 10, 13, 15, and 17. The locus on chromosome 15, D15S652, is telomeric of and not linked to the candidate region, on proximal 15q, associated with the inverted duplication found in some autistic subjects (Gillberg et al. 1991; Cook et al. 1997). Aside from locus D1S1631, there appears to be no substantial difference between the tail Z scores for the ASP group and those for the DSP group.

We also examined the full distribution of Z scores, for both the ASPs and the DSPs. These distributions are given in figure 1. For convenience, we have plotted each Z score as a function of $\Phi^{-1}[(R - \frac{1}{2})/S]$, where R is the ordered rank of that Z score, $S = 346$ is the total number of autosomal markers, and Φ is the cumulative normal-distribution function. Under the null hypothesis (no linkage), this curve should fit the line $y = x$. In figure 1, the line $y = x$ is depicted as the thin unbroken line; the thick dotted line represents DSPs, and the thick dashed line represents ASPs. As can be seen in the figure, both the ASP and DSP lines run parallel but above the line $y = x$, but more so in the case of the ASP line. This indicates that the entire distribution of Z scores for the ASPs is shifted upward, compared with that for the DSPs, and this shift is most noticeable at the left end of the distri-

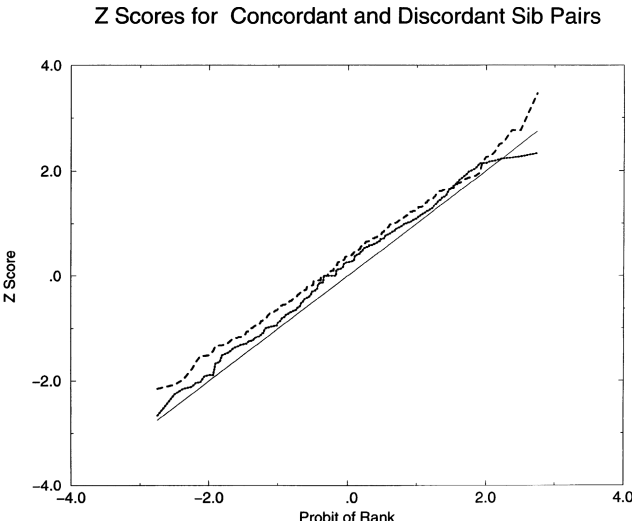

Figure 1 Observed distribution of Z scores for sharing at 346 autosomal marker loci in ASPs and DSPs, compared with the expected distribution under no linkage. The thin unbroken line represents no linkage; the thick dashed line represents sharing for ASPs; the thick dotted line represents sharing for DSPs.

bution (i.e., deficiency of negative sharing). The final "blip" at the right end of the ASP line is due to marker D1S1631.

Observed versus Expected Distribution Of IBD Scores

As discussed above, family and twin studies suggest that autism is a highly heritable disorder. However, these data alone cannot reveal the precise mode of inheritance. The very high MZ:DZ concordance ratio, taken at face value, is strongly suggestive of multiple interacting loci (epistasis), with no single locus having a substantial effect. The results of our initial genome screen would seem to support this conclusion. However, given the high recurrence-risk ratio for sibs (λ_S = 75–100) and the high (15–25) MZ:DZ concordance ratio, one may reasonably ask what the distribution of IBD scores would look like for various genetic models consistent with those values. For the purpose of computational simplification, we assume a multiplicative epistatic model (Risch 1990) with K loci of equal effect. Given the 15–25 MZ:DZ concordance ratio, models with as few as two loci can be excluded. For a model with K loci, the ith locus will have an individual value of $\lambda_{Si} = (\lambda_S)^{\frac{1}{K}}$ for $i = 1, \ldots, K$ (Risch 1990). The ASP sharing y_i at each of these loci will be $y_i = 1 - .5(\lambda_{Si})^{-\frac{1}{2}} = 1 - .5(\lambda_S)^{-\frac{1}{2K}}$ according to a multiplicative model. For example, for $K = 3$, y_i = 75.7%; for $K = 5$, $y_i = 67.5\%$; for $K = 10$, y_i = 59.7%; and, for $K = 20$, $y_i = 55.1\%$. If $\lambda_S = 150$, then the corresponding y_i values for $K = 3$, 5, 10, and 20 are only slightly higher—namely, 78.3%, 69.7%, 61.1%, and 55.9%, respectively. Similar numbers can be obtained by parameterization in terms of the MZ:DZ concordance ratio, which we here denote as "λ_{MD}." According to a multiplicative model, for K contributing loci of equal effect, the total λ_{MD} is $(\lambda_{MDi})^K$, where λ_{MDi} is the ratio contributed by the ith locus. Furthermore, the expected ASP IBD sharing for a locus with value λ_{MDi} is given simply by $y_i = \frac{1}{2}\sqrt{\lambda_{MDi}}$. Since, in a model with K loci of equal effect, $\lambda_{MDi} = (\lambda_{MD})^{\frac{1}{K}}$ for $i = 1, \ldots, K$, we can calculate the expected sharing, under the assumption that $\lambda_{MD} = 25$, for $K = 3$, 5, 10, and 20 as being 85.5%, 69.0%, 58.7%, and 54.2%, respectively. It can be seen that these numbers are larger than those based on the sib risk ratio for smaller K values (i.e., <5) but that they begin to converge at $K > 5$.

Certainly our data are not consistent with a three-locus model, since we do not have three regions with sharing approaching 70%. Recombination will tend to decrease sharing away from these high values, but, given our map density, it is unlikely that the sharing would be significantly diminished. In fact, we can calculate systematically, making some simplifying assumptions, what the distribution of sharing (and associated Z scores) should be under these alternative genetic models.

Consider the distribution of Z scores derived from a genome screen. For a given marker, Z is defined, as above, as $(s - u)/\sqrt{s + u}$. Under the null hypothesis of no linkage, Z is approximately normally distributed, with mean 0 and variance 1. Now consider a genetic model of K disease loci, each of which has sharing y_i. We assume that the loci are unlinked. Assume a genetic map with a marker every w cM. On average, the distance of a disease locus to the closest marker will be $\frac{w}{4}$; the distances to remaining markers, in order of distance, are $\frac{3w}{4}$, $\frac{5w}{4}$, $\frac{7w}{4}$, etc. For a marker at recombination distance r from a disease locus, the sharing will be $y(r) = [r^2 + (1 - r)^2]y + 2r(1 - r)(1 - y)$. Thus, assuming a map function relating recombination fraction r to map distance w—for example, $r = f(w)$—we can calculate the expected values of sharing y for markers surrounding a disease locus.

In the present analysis, we assume the Kosambi mapping function $r = f(w) = (1 - e^{-4w})/2(1 + e^{-4w})$. As in our screen, we assume a map density of 10 cM, with a total map length of 3,500 cM. For each of K putative disease loci, we estimate sharing $y(r)$ at neighboring loci at distances 2.5 cM, 7.5 cM, 12.5 cM,..., up to 47.5 cM, using the Kosambi map function and the $y(r)$ formula given above.

For the model with K loci of equal effect, there are K loci each with value $y(r)$, where $y(r)$ ranges over the 10 values based on the 10 different marker distances. In our genome screen of ASPs, the average total number of transmissions scored at a locus was 110 (i.e., $s + u$, in the notation above). Therefore, for the K markers with expected sharing $y(r)$, we used sampling from a normal distribution with mean $y(r)$ and variance $v(r) = y(r)[1 - y(r)]/110$ to obtain the expected distribution of their observed sharing. We selected the K points from this distribution to be equally spaced to preserve the mean $y(r)$ and variance $v(r)$. For example, for $K = 5$ we chose the five points $y_1(r) = y(r) - \sqrt{2v(r)}$, $y_2(r) = y(r) - \sqrt{v(r)/2}$, $y_3(r) = y(r)$, $y_4(r) = y(r) + \sqrt{v(r)/2}$, and $y_5(r) = y(r) + \sqrt{2v(r)}$. These five points have sample mean $y(r)$ and variance $v(r)$. For the jth such point, the expected Z score was then derived by taking $Z_j(r) = 2\sqrt{110}[y_j(r) - \frac{1}{2}] = 10.49[2y_j(r) - 1]$. As an example, consider the genetic model with $\lambda_S = 75$, $K = 5$, and the third-closest marker (12.5 cM). For this case, y = .675, r = .122, $y(r)$ = .600, $y_1(r)$ = .534, $y_2(r)$ = .567, $y_3(r)$ = .600, $y_4(r)$ = .633, and $y_5(r)$ = .666; $Z_1(r)$ = .71, $Z_2(r) = 1.41$, $Z_3(r) = 2.10$, $Z_4(r) = 2.79$, and $Z_5(r) = 3.48$.

In this fashion, Z scores for $10K$ markers were derived. For the remaining $346 - 10K$ markers, we distributed their Z scores as above but according to a normal distribution with mean 0 and variance 1. We have calculated these expected distributions, assuming that $\lambda_S = 75$ and K = 3–20. The results for K = 3, 5, 10,

and 20 are given in figure 2. As in figure 1, we have plotted the Z scores as a function of $\Phi^{-1}[(R - \frac{1}{2})/S]$, where R is the rank of the observed Z score and S is the total number of markers (346). A normal distribution appears as a straight line ($y = x$) on this plot and, in figure 2, is depicted by the thick unbroken line. The expected curve corresponding to $K = 3$ is given by the thin dashed lined, for $K = 5$ by the thin dotted line, for $K = 10$ by the thin short-dashed line, and for $K = 20$ by the thin unbroken line. The curve for the actual data is given by the thick dashed line. In calculating the Z scores for the actual data, we used 50.8% for expected sharing under the null hypothesis, instead of 50%, to allow for some potential genotyping bias, as observed in the DSPs. As can be seen in figure 2, the actual data most closely approximate the expectations for the model with $K = 20$. The fact that the thick dashed line lies above the $y = x$ line at the left end of the plot indicates that the excess sharing in ASPs compared with DSPs is due to the entire distribution being shifted to the right, rather than to a small subset of marker loci showing increased sharing. The right end of the plot, although reasonably consistent with $K = 20$, is quite deviant from the distribution expected for $K = 5$ or 10, with the exception of the final marker (D1S1631). For other values of K, the distributions lie between the curves in figure 2. For example, for $K = 15$ the distribution lies approximately midway between the line for $K = 10$ and that for $K = 20$; for $K = 13$ the distribution is somewhat closer to the line for $K = 10$.

To assess goodness of fit of each predicted distribution to the observed distribution, we calculated the statistic $W = \Sigma_{i=1}^{346}(Z_{i1} - Z_{i0})^2$, where Z_{i1} is the ith (rank-ordered) Z score for the expected distribution and Z_{i0} is the corresponding Z score for the observed distribution. A small value of W indicates a good fit of the observed distribution to the predicted distribution. The corresponding value of W for values of $K = 3$, 5, 10, 13, 15, 17, and 20 is 53.1, 16.7, 5.1, 3.6, 3.1, 2.8, and 2.7, respectively. Thus, models with $\leq$10 loci appear unlikely, whereas those with $\geq$15 seem more plausible. The best fit is given by the model with $K = 20$ loci, although the value of W does not increase much up to $K = 15$. Models specifying a small number of loci (i.e., one or two) with moderate effect ($\lambda_S = 2$) also give a reasonable fit (although not as good as is given by the model of 20 loci with equal effect), provided that the number of residual loci remains large (e.g., $\geq$15). Thus, our results, although consistent with a genetic basis of autism, suggest that the most likely model is one with many ($\geq$15) contributing loci, with a possible gene of larger effect on chromosome 1.

In calculating the expected distributions, we have made a number of simplifying assumptions, such as equal spacing and equal polymorphism of the markers.

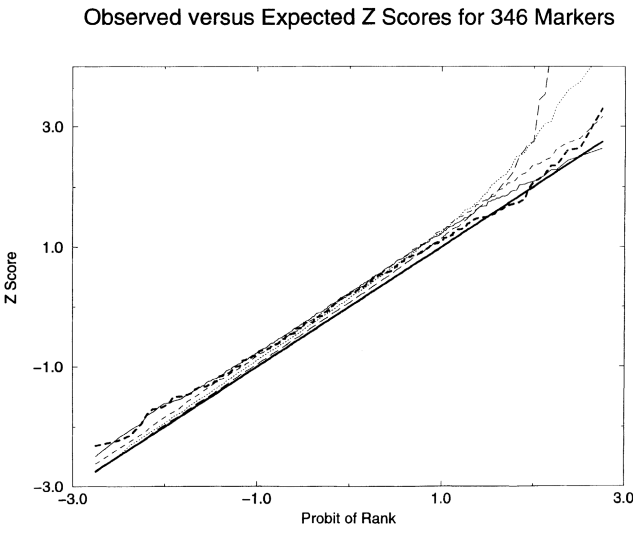

Figure 2 Observed distribution of Z scores for sharing at 346 autosomal marker loci in ASPs, compared with the expected distribution, when a model of $K = 3$, 5, 10, or 20 loci of equal effect is assumed. The thick unbroken line represents no linkage; the thick dashed line represents the observed data for ASPs; the thin long-dashed line represents $K = 3$; the thin dotted line represents $K = 5$; the thin short-dashed line represents $K = 10$; the thin unbroken line represents $K = 20$.

Most significant among these assumptions, however, is that the disease-susceptibility loci are unlinked, which will be unlikely for models with $K \geq 20$. The effect of having linked disease loci is to increase sharing locally in the region of the linked loci—and to decrease it, on average, elsewhere. Thus, the impact will be to create a distribution more similar to the case of fewer disease loci with larger effect. Therefore, our modeling is actually conservative in concluding that the observed distribution is most compatible with a large number of disease loci.

Sharing by Gender of Parent

The observation of genomewide excess sharing and a putative model of many genes of small effect accounting for the total familial effect in autism allow us to posit a further prediction. If, in fact, there are many genes scattered around the genome, then the excess of sharing that we have observed should differ when tallied separately for paternal versus maternal meioses. The reason is that the female linkage map, at 4,200 cM, is ~1.5 times the genetic length of the male map, at 2,800 cM. Because there is globally less recombination in males, the expected genetic distance (recombination fraction [θ]) to a disease-susceptibility locus in males is expected to be less than that in females, and thus the expected allele sharing at a marker linked to the disease locus is expected to be greater when one examines fathers than when one examines mothers.

Table 3

Global Allele Sharing in ASPs and DSPs, by Sex of Parent and Genetic Map–Interval Male:Female Ratio

Sex Ratio (Male:Female)	No. (%) Sharing	
	Paternal	Maternal
ASPs:		
>1 (85 markers)	2,336/4,560 (51.2)	2,413/4,712 (51.2)
<1 (261 markers)	7,459/14,275 (52.3)	7,694/15,025 (51.2)
Total	9,795/18,835 (52.0)	10,107/19,737 (51.2)
Total (maternal + paternal)	19,902/38,572 (51.6)	
DSPs:		
>1 (85 markers)	1,110/2,160 (51.4)	1,167/2,304 (50.7)
<1 (261 markers)	3,401/6,722 (50.6)	3,666/7,200 (50.9)
Total	4,511/8,882 (50.8)	4,833/9,504 (50.9)
Total (maternal + paternal)	9,344/18,386 (50.8)	

To quantitate this prediction, define $d(0) = y - \frac{1}{2}$, where y is the sharing at a disease locus. Then, at θ distance, the value of y is $y(\theta) = [\theta^2 + (1-\theta)^2]y + 2\theta(1-\theta)(1-y)$, so that $d(\theta) = y(\theta) - \frac{1}{2} = (1-2\theta)^2 d(0)$. If we assume a uniform distribution for the map distance, w, from a random marker locus to the nearest disease locus, then, over some total map length L, the expected value of d is

$$\mathrm{E}(d) = \frac{1}{L}\int_0^L d(0)[1-2\theta(w)]^2 dw \ ,$$

where w is some mapping function relating map distance to θ. For the Haldane mapping function $\theta(w) = (1 - e^{-2w})/2$, $\mathrm{E}(d) = d(0)(1-e^{-4L})/L \cong d(0)/L$, for $L > 50$ cM. Similarly, for the Kosambi mapping function $\theta(w) = (e^{4w}-1)/2(e^{4w}+1)$, $\mathrm{E}(d) = d(0)[(1+e^{2L})^{-1} - \log(1+e^{-2L}) + \log 2 - \frac{1}{2}]/L \cong .193d(0)/L$ for $L > 50$ cM. In either case, $\mathrm{E}(d) \propto \frac{1}{L}$. Therefore, if female map distances are 1.5 times male map distances and if we let d_m represent the d value for male (paternal) meioses and let d_f represent that for female meioses, then we have $\mathrm{E}(d_m)/\mathrm{E}(d_f) \propto 1.5$. Therefore, when the observed sharing rate of .508 (from the discordant pairs) is used as the null value, the observed sharing for concordant pairs, .516, gives a sex-averaged value of $\mathrm{E}(d) = .008$. Therefore, the prediction discussed above would give sex- specific $\mathrm{E}(d)$ values of $\mathrm{E}(d_m) = .010$ and $\mathrm{E}(d_f) = .006$, or sex-specific sharing values of $y_m = 51.8\%$ and $y_f = 51.4\%$. If we use 50% for the null-hypothesis sharing value, then the sex-specific sharing becomes $y_m = 51.9\%$ and $y_f = 51.3\%$.

The observed total sharing for ASPs, stratified by sex of parent, is $y_m = 9{,}795/18{,}835 = 52.0\%$ and $y_f = 10{,}107/19{,}737 = 51.2\%$ (table 3), close to the prediction. For comparison, we also examined the parent-specific allele sharing for the DSPs. There was no difference in overall allele sharing, by sex of parent, for this group (50.8% paternal, 50.9% maternal).

We can make this prediction more precise by noting that the male:female ratio of map distances is not uniform throughout the genome but, rather, is region specific. Some regions (e.g., telomeres) tend to show an excess of recombination in males. Thus, we would expect the observed excess of paternal allele sharing to be restricted to those regions where the female map distances exceed those of the male. To test this hypothesis, we identified the location of each of the 346 marker loci on the Marshfield genetic maps (Center for Medical Genetics, Marshfield Medical Research Foundation). By examining, for each locus, flanking markers up to 10 cM away (sex averaged), we determined whether the genetic length of the flanking interval was greater in males (male:female ratio >1) or in females (male:female <1). As expected, the number of markers with a male:female ratio >1 ($n = 85$) was approximately one-third the number with a male:female ratio <1 ($n = 261$). As predicted, the paternal excess of sharing in ASPs was restricted to markers lying in regions with a male:female ratio <1 (52.3% paternal sharing vs. 51.2% maternal sharing) (table 3). Again, the DSPs showed no such trend, since the paternal and maternal sharing were nearly equal in both groups (table 3).

Multipoint Sib-Pair Analysis in FS1

We performed multipoint sib-pair analysis on the data from FS1, for the 360 autosomal and X-linked markers of MS1. The results are shown in figure 3. In the figure, the MLS curve is represented by the dashed line. There were nine locations that gave an MLS ⩾1.0: chromosomes 1p (MLS 1.87), 1q (MLS 1.19), 7p (MLS 1.00), 11p (MLS 1.25), 13q (MLS 1.49), 15q (MLS 1.75), 17p (MLS 1.30), 18q (MLS 1.00), and 20p (MLS 1.09). The MLS on chromosome 1p was diminished compared with

the significance obtained with marker D1S1631 (table 2), because of less sharing at neighboring markers.

These results are consistent with those shown in figures 1 and 2, in that no regions revealed linkage evidence that was near significance, even when a liberal criterion of MLS 3.0 was used. Thus, it is unlikely that any genes of moderate to large effect underlie this disorder.

Follow-up in FS2

The 49 families of FS2, comprising 50 ASPs, were analyzed for a total of 149 markers, 60 of which were original (i.e., MS1) markers typed in FS1. A total of 157 additional markers (MS2) were typed in FS1; 89 of these were also typed in FS2. Many new markers were typed only in FS1, either to fill in larger gaps in the original screen or to increase marker density in certain candidate regions, in particular on chromosomes 6p and 15q. On the basis of preliminary positive results in the first stage of analysis, markers were added on chromosomes 1 (two regions), 3, 9, 10, 13, 15, and 17.

Results of multipoint analysis of the combined data from both FS1 and FS2, for all 517 autosomal and X-linked loci, are given in figure 3 (the MLS curve for the total data is represented by the unbroken line; the exclusion curve fixing $\lambda_{Si} = 3.0$ is represented by the dotted line). The linkage evidence, in FS2, for markers on proximal chromosome 1p was positive, with allele sharing of 50%–66% (average 56%); in FS1, maximal sharing in multipoint analysis was observed near marker D1S1631. With the inclusion of FS2 and additional markers in this region, the MLS of 2.15 now occurs at a slightly more proximal location (near marker D1S1675). This location gives, by far, the strongest linkage evidence in our study. The next most significant result is on chromosome 17p, where the MLS peaked at 1.21, near marker D17S1876. Only two other regions had an MLS $\geq$1.0, one each on chromosome 7p (MLS 1.01, near marker D7S2564) and 18q (MLS 1.00, near marker D18S878). All other regions with MLS $\geq$1.0 in the initial scan were less positive in FS2, leading to a reduction in the total MLS for FS1 and FS2 combined. Notably, only chromosome 1p had substantially positive allele sharing in FS2, leading to a modest increase in the MLS for FS1 and FS2 combined.

Proximal chromosome 15q is a candidate region for an autism-susceptibility locus, because of reported inverted duplications in a number of patients (Gillberg et al. 1991; Cook et al. 1997). Furthermore, some positive linkage results were reported in this region in a small collection of multiplex families (Pericak-Vance et al. 1997). We previously have reported that, in this region of chromosome 15, there is a lack of linkage evidence in the 97 FS1 ASPs (Salmon et al., in press). The addition of 50 ASPs from FS2 has not altered this conclusion, since our linkage evidence is still uniformly negative across the entire region (fig. 3). The positive linkage results that we obtained on this chromosome were near marker D15S1050, which is 62 cM away.

We have also previously reported a lack of linkage evidence with markers in the HLA region on chromosome 6p in FS1 (Rogers et al., in press). The data in this region remain negative with the addition of FS2.

In a previous analysis of 38 of the ASPs included in the present study, we examined markers on the X chromosome. Although we excluded most of the chromosome, including the fragile X region, from harboring a gene of large effect, a region of Xq did show modest evidence of linkage, with a LOD score of 1.24. In the present study, based on a much larger number of families, there is no evidence of linkage anywhere on the X chromosome.

A genome screen of 39 multiplex families with autism, with follow-up in an additional 60 families, found suggestive evidence of linkage on chromosomes 7q and 16p (International Molecular Genetic Study of Autism Consortium 1998). The MLS in the same region of 7q is 0.62 (for marker D7S684), on the basis of a total of 139 families and 147 ASPs. We had a slightly more positive result at marker D7S1804 (MLS 0.93), located 10.3 cM more proximal. The original study had an MLS of 2.53. Thus, if a susceptibility locus resides in this region, its effect is likely to be small, given the modest LOD scores observed.

In our initial scan (i.e., with FS1), a region of chromosome 13q had an MLS of 1.65, near marker D13S779. In our follow-up, the linkage evidence at this location was negative (46.3% sharing); however, a region of modestly increased sharing appeared somewhat distal in the same area, with an MLS of 0.68, near marker D13S800 (fig. 3). This region has also been reported in another recent genome screen (Vieland et al. 1998). Again, the modest level of linkage evidence in our collection of 147 ASPs suggests that a gene of large effect in this region is unlikely.

Using a strict exclusion criterion of -2 (Hauser et al. 1996), we were able to exclude ~95% of the genome, for a value of $\lambda_{Si} = 3.0$ (fig. 3). When the less strict exclusion criterion of -1 was used, 99% of the genome could be excluded.

Positive Controls

In our study, we had initially included two twin pairs originally thought to be DZ but subsequently, through our genotyping, shown to be MZ. These subjects were left in the study as a positive control and to estimate the rate of genotyping error. The individuals performing ge-

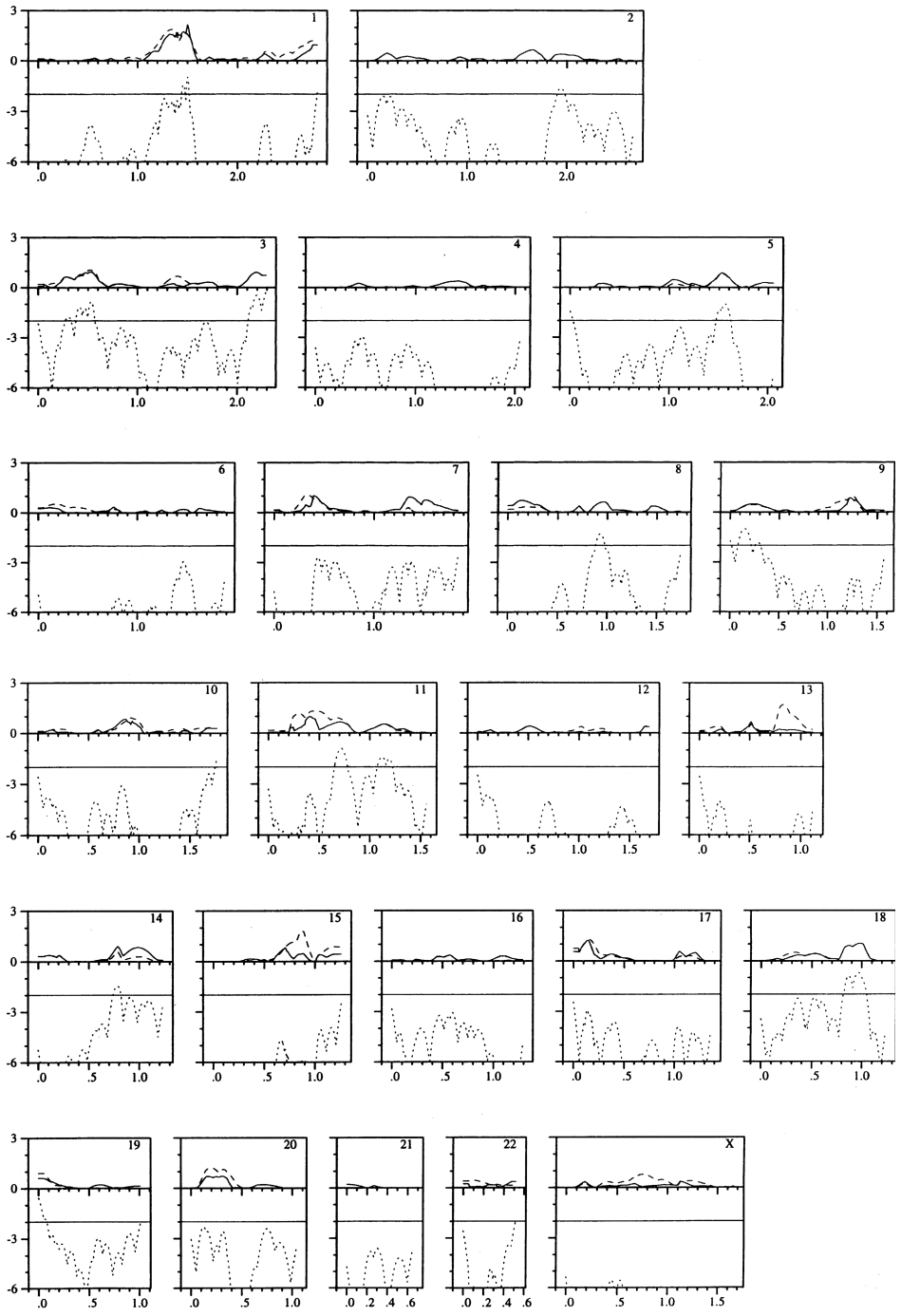

Figure 3 Multipoint sib-pair analysis of ASPs. The dashed line represents the MLS for FS1 with MS1; the unbroken line represents the MLS for FS1 + FS2, with all markers; the dotted line represents the exclusion plot for $\lambda_S = 3.0$, with all families and markers.

notyping were blinded to the status of these subjects. Considering a total of 946 loci scored for these two pairs, we found them to be genotypically discordant at 10 loci, or a rate of 1.0%. It should be noted that, in our study, families were analyzed intact, so that some genotyping errors (Mendelian inconsistencies) could be identified. These Mendelian inconsistencies were identified and corrected (by additional genotyping, if necessary) and were not included in the error rate given above.

As a second positive control, we included the pseu-

doautosomal marker DXYS154. This marker is completely linked to sex, since it does not recombine between the X and Y chromosomes. In our study, we had a large predominance of affected males (male:female ratio 3.6:1 in FS1 and 2.8:1 in FS2). Thus, a majority of our ASPs are brother pairs, whereas a minority are sister pairs or of mixed sex. Overall, 73% of our ASPs are concordant for sex (63% males and 10% females). Therefore, we expect, at locus DXYS154, excess sharing of alleles inherited from the fathers, but not of those inherited from the mothers. In FS1 and FS2 combined, there were 58 paternal alleles shared and 22 paternal alleles not shared, giving a χ^2 of 16.2 (the LOD score equivalent of which is 3.52). By contrast, there was no difference in the sharing of maternal alleles: 42 alleles were shared and 45 alleles were not shared. The significantly increased sharing of paternal alleles at this marker demonstrates that our methodology was sufficiently sensitive to detect a real excess of sharing of this magnitude (i.e., 73%) when such an excess was present.

LD

We performed TDTs for all 517 autosomal and X-linked loci analyzed in this study. Here we report the results of the GCS test (see the Linkage-Disequilibrium analysis [LD] subsection, above). In figure 4 we have plotted the distribution of observed (empirically derived) *P* values, as a function of their rank. Under the null hypothesis of no LD, the distribution of *P* values should be uniform, and thus this plot should approximate the line $y = x$, which, in figure 4, is represented by the straight unbroken solid line; the values for affected offspring are represented by the dashed line, and those for unaffected offspring are represented as the jagged unbroken line. As can be seen in the figure, for $P < .05$ the observed and expected distributions are the same for the two types of offspring and fit the line $y = x$ well. These results are not unexpected, in that it is a priori unlikely that a disease locus would lie very close to and in LD

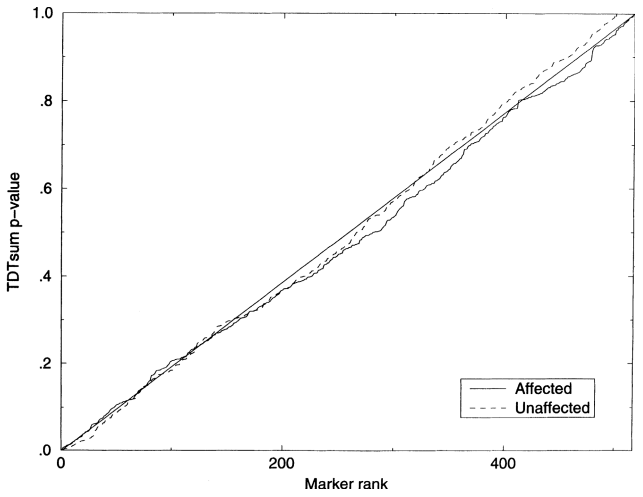

Figure 4 Observed distribution of *P* values from TDT tests for 519 markers for affected and unaffected offspring. The straight unbroken line represents expectation under no LD; the dashed line represents unaffected offspring; the jagged unbroken line represents affected offspring.

Discussion

We have performed a genomic screen by linkage analysis of 90 multiplex sibships with autism, including 97 independent ASPs, with 360 autosomal and X-linked markers. Regions that, in this screen, initially either (*a*) had positive results, (*b*) had gaps, or (*c*) were identified as specific candidate regions were subjected to follow-up with additional markers. We also genotyped a second set of 49 multiplex sibships (50 ASPs), with markers in specific target regions. In total, 519 markers were run on FS1, and 149 were run on FS2. More than 160,000 genotypes were generated.

Analysis of allelic sharing for the initial 360 markers in FS1 was most consistent with a multigenic inheritance pattern, in that a global excess of allele sharing was observed in the ASPs versus the DSPs, but not specifically for a small number of loci. The overall distribution of allele sharing was most consistent with a model of $\geqslant$15 susceptibility loci.

These results are not inconsistent with the observed recurrence risks in nuclear families and twins. The very high (25-fold) MZ:DZ concordance ratio is indicative of at least several interacting loci and, potentially, of many such loci. Multilocus inheritance does not preclude the possibility of one or a few loci with larger effects; there is a limit, however, on how large these effects can be, given our modest linkage findings. A previous study of family recurrence risks in autism found a model with 3 interactive loci as providing the best fit to the data, with a range of 2–10 loci (Pickles et al. 1995). Our results clearly exclude a model with as few as 3 loci and are also inconsistent with a model with as few as 10 loci.

Our results also do not preclude the possibility of etiologic heterogeneity, with a subset of cases having a simpler (e.g., Mendelian) etiology. However, as explained above, such a group would need to be relatively minor, given both our results and the large MZ:DZ concordance ratio.

Our most significant findings were for proximal chromosome 1p. The most significant increase in sharing in our initial screen was for marker D1S1631, with 66% sharing and a Z score of 3.44 (table 2). The multipoint

analysis of FS1 brought this score down to an MLS of 1.88; however, this region was also the most positive in our follow-up and, in 147 ASPs, gave a total MLS of 2.15 and sharing of 60%. We note, however, that this result falls short of formal statistical significance for a genome screen (Lander and Kruglyak 1995). The MLS location after follow-up was somewhat more proximal than that in the initial screen, near marker D1S1675.

No other region had an MLS >1.3 in our follow-up. Our next most significant finding was for chromosome 17p, where the MLS was 1.21. We had modestly positive LOD scores in regions identified in other genome screens—chromosomes 7q and 13q. Our results for chromosome 15q, in the Prader-Willi region, which, on the basis of cytogenetic evidence, is a candidate region for an autism–susceptibility locus, were uniformly negative, consistent with our earlier results (Salmon et al., in press). However, in our initial screen, we did identify another region, more distal on chromosome 15q, that had an MLS of 1.75; but, in the follow-up, the MLS in this region decreased to 0.81.

We employed several controls in this study, to identify potential biases. First, we analyzed unaffected siblings in our multiplex sibships, to create a comparison group for our linkage and LD results. Second, we kept two pairs of MZ twins, originally thought to be DZ, in the genotyping set, as a positive control for our genotyping, as well as to estimate the error rate in genotyping. Finally, we included a pseudoautosomal marker completely linked to sex, as a positive control, reflecting the excess of sex-concordant (primarily brother) pairs in our study. As predicted, this marker showed a clear, statistically significant paternal excess of sharing (MLS 3.52). These controls provide confidence in the validity of our conclusions.

We also took a conservative approach to phenotyping. All study subjects were evaluated by both ADI assessment and ADOS assessment, and all questionable cases were excluded from analysis. Furthermore, to reduce potential heterogeneity, we excluded families in which all affected children had extremely low IQ, since this group may have distinct etiologies related to their severe-to-profound mental retardation. In total, we excluded 45 families from further analysis. The primary reason for the family exclusions was that not all affected sibs had a strict diagnosis of autism, as defined by agreement of all diagnosticians. As a result, the families in our study included few broadly defined, questionable, or mild cases (e.g., clinical diagnoses of either PDD-NOS or Asperger syndrome), even though most of the children in these excluded families did exhibit symptoms and behaviors indicative of the autism spectrum, some at the very high end of the spectrum and some at the very low end.

We employed these conservative diagnostic rules for inclusion within the study, which were applied equally to all affected subjects in a family, for two reasons. First, we wanted to enhance our chances of detecting any linkage, by creating a more homogeneous set of families. Second, we wanted to gain confidence that any negative results that we might obtain would not be due to the use of a too-broad diagnostic spectrum whose genetic basis is still uncertain (even though a number of recent family studies of autism probands suggest the presence of a broad range of mild behavioral symptoms and possibly related disorders in these families [Bailey et al. 1998]). Although our exclusion of "mild" or autism "spectrum" cases might be questioned by some who argue strongly for their inclusion within genetic studies, the only negative effect that this exclusion would have on our analysis would be to reduce the sample size. A homogeneous group of strictly defined affected subjects should maximize the chances of detection of linkage, for any plausible genetic model.

Although our genomic-screen results are largely negative, in that we did not identify a chromosomal region with significant linkage evidence, we have not formally excluded the possibility of one or a few disease-predisposing loci of moderate effect. On the basis of our results, the most likely location for such a locus or loci is on chromosome 1p—and, possibly, on chromosome 17p. We are currently examining these regions more closely. However, given the lack of strong linkage evidence in this large collection of families, positional cloning of any susceptibility genes in this disorder may be a difficult undertaking, and other approaches, such as LD studies of candidate genes and/or genomewide association studies may be required.

Acknowledgments

This research has been supported by National Institute of Mental Health grants MH52708, MH39437, MH00219, and MH00980; National Institutes of Health HG00348; National Health Medical Research Council grant 0034328; and by grants from the Scottish Rite, the Spunk Fund, Inc., the Rebecca and Solomon Baker Fund, the APEX Foundation, the National Alliance for Research in Schizophrenia and Affective Disorders, the endowment fund of the Nancy Pritzker Laboratory (Stanford); and by gifts from the Autism Society of America, the Janet M. Grace Pervasive Developmental Disorders Fund, and families and friends of individuals with autism. We are indebted to Drs. Magda Campbell, Isabelle Rapin, Peter Szatmari, and Luke Tsai for their expert clinical advice as study advisors, and we wish to thank the parent support groups and our clinician colleagues who referred families to this project. Our gratitude goes to the families with autism who have been our research partners in this endeavor. We also

wish to acknowledge the pioneering foresight of Dr. Roland Ciaranello, who led this study before his untimely death.

Electronic-Database Information

Accession numbers and URLs for data in this article are as follows:

ASPEX program package, ftp://lahmed.stanford.edu/pub/aspex

Center for Medical Genetics, Marshfield Medical Research Foundation, http://www.marshmed.org/genetics (for genetic markers)

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/omim (for TSC1 [MIM 191100]; autism, PDD, and Asperger syndrome [MIM 209850]; and FMR1 [MIM 309550])

References

Anderson MA, Gusella JF (1984) Use of cyclosporin A in establishing Epstein-Barr virus–transformed lymphoblastoid cell lines. In Vitro 11:856–858

Bailey A, LeCouteur A, Gottesman I, Bolton P, Simonoff E, Yuzda E, Rutter M (1995) Autism as a strongly genetic disorder: evidence from a British twin study. Psychol Med 25:63–77

Bailey A, Palferman S, Heavey L, LeCouteur A (1998) Autism: the phenotype in relatives. J Autism Dev Disord 28:369–392

Bolton P, Macdonald H, Pickles A, Rios P, Goode S, Crowson M, Bailey A, et al (1994) A case-control family history study of autism. J Child Psychol Psychiatry 35:877–900

Cook EH Jr, Lindgren V, Leventhal BL, Courchesne R, Lincoln A, Shulman C, Lord C, et al (1997) Autism or atypical autism in maternally but not paternally derived proximal 15q duplication. Am J Hum Genet 60:928–934

Feinstein C, Reiss AL (1998) Autism: the point of view from fragile X studies. J Autism Dev Disord 28:393–405

Folstein S, Rutter M (1977) Infantile autism: a genetic study of 21 twin pairs. J Child Psychol Psychiatry 18:297–321

Gillberg C (1998) Chromosomal disorders and autism. J Autism Dev Disord 28:415–425

Gillberg C, Steffenburg S, Wahlstrom J, Gillberg I, Sjostedt A, Martinsson T, Liedgren S, et al (1991) Autism associated with marker chromosome. J Am Acad Child Adolesc Psychiatry 30:489–494

Hallmayer J, Hebert JM, Spiker D, Lotspeich L, McMahon WM, Petersen PB, Nicholas P, et al (1996) Autism and the X chromosome. Arch Gen Psychiatry 53:985–989

Hanson DR, Gottesman I (1976) The genetics, if any, of infantile autism and childhood schizophrenia. J Autism Child Schizophr 6:209–234

Hauser ER, Boehnke M, Guo SW, Risch N (1996) Affected sib-pair interval mapping and exclusion for complex genetic traits: sampling considerations. Genet Epidemiol 13: 117–137

International Molecular Genetic Study of Autism Consortium (1998) A full genome screen for autism with evidence of linkage to a region of chromosome 7q. Hum Mol Genet 7: 571–578

Jorde LB, Hasstedt SJ, Ritvo ER, Mason-Brothers A, Freeman BJ, Pingree C, McMahon WM, et al (1991) Complex segregation analysis of autism. Am J Hum Genet 49:932–938

Jorde LB, Mason-Brothers A, Wasdmann R, Ritvo ER, Freeman BJ, Pingree C, McMahon WM, et al (1990) The UCLA-University of Utah epidemiologic survey of autism: genealogical analysis of familial aggregation. Am J Med Genet 36:85–88

Kanner L (1943) Autistic disturbances of affective contact. Nerv Child 2:217–250

Lander E, Kruglyak L (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet 11:241–247

Lazzeroni LC, Lange K (1998) A conditional inference framework for extending the transmission/disequilibrium test. Hum Hered 48:67–81

Le Couteur A, Rutter M, Lord C, Rios P, Robertson S, Holdgrafer M, McLennan J (1989) Autism Diagnostic Interview: a standarized investigator-based instrument. J Autism Dev Disord 19:363–387

Lord C, Pickles A, McLennan J, Rutter M, Bregman J, Folstein S, Fombonne E, et al (1997) Diagnosing autism: analyses of data from the Autism Diagnostic Interview. J Autism Dev Disord 27:501–517

Lord C, Rutter M, Good S, Heemsbergen J, Jordan H, Mawhood L, Schopler E (1989) Autism diagnostic observation schedule: a standardized observation of communicative and social behavior. J Autism Dev Disord 19:185–212

Lord C, Rutter M, LeCouteur A (1994) Autism Diagnostic Interview—Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. J Autism Dev Disord 24:659–685

Martin ER, Kaplan NL, Weir BS (1997) Tests for linkage and association in nuclear families. Am J Hum Genet 61: 439–448

Pericak-Vance MA, Wolpert CM, Menold MM, Bass MP, Delong GR, Beaty LM, Zimmerman A, et al (1997) Linkage evidence supports the involvement of chromosome 15 in autistic disorder (AUT). Am J Hum Genet Suppl 61:A40

Pickles A, Bolton P, Macdonald H, Bailey A, LeCouteur A, Sim C-H, Rutter M (1995) Latent-class analysis of recurrence risks for complex phenotypes with selection and measurement error: a twin and family history study of autism. Am J Hum Genet 57:717–726

Risch N (1990) Linkage strategies for genetically complex diseases. I. Multilocus models. Am J Hum Genet 46:222–228

Rogers T, Kalaydjieva L, Hallmayer J, Petersen PB, Nicholas P, Pingree C, McMahon W, et al. Exclusion of linkage to the HLA region in ninety multiplex sibships with autism. J Autism Dev Disord (in press)

Salmon B, Rogers T, Kalaydjieva L, Hallmayer J, Petersen PB, Nicholas P, Pingree C, et al. Absence of linkage to chromosome 15q11-q13 markers in ninety multiplex families with autism. Am J Med Genet (in press)

Smalley SL (1998) Autism and tuberous sclerosis. J Autism Dev Disord 28:407–414

Smalley SL, Asarnow RF, Spence MA (1988) Autism and genetics: a decade of research. Arch Gen Psychiatry 45: 953–961

Spence MA, Ritvo ER, Marazita ML, Funderburk SJ, Sparkes RS, Freeman BJ (1985). Gene mapping studies with the syndrome of autism. Behav Genet 15:1–13

Spielman RS, McGinnis RE, Ewens WJ (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 52:506–516

Spiker DK, Lotspeich L, Kraemer HC, Hallmayer J, McMahon W, Peterson PB, Nicholas P, et al (1994) The genetics of autism: characteristics of affected and unaffected children from 37 multiplex families. Am J Med Genet 54:27–35

Steffenburg S, Gillberg C, Hellgren L, Andersson L, Gillberg L, Jakobsson G, Bohman M (1989) A twin study of autism in Denmark, Finland, Iceland, Norway and Sweden. J Child Psychol Psychiatry 30:405–416

Szatmari P, Jones MB, Zwaigenbaum L, MacLean JE (1998) Genetics of autism: overview and new directions. J Autism Dev Disord 28:351–368

Vieland V, for the Collaborative Linkage Study of Autism (CLSA) (1998) Results of a genomic screen for autism include strong evidence of linkage to chromosome 13. Am J Hum Genet Suppl 63:A16