Am. J. Hum. Genet. 72:539–548, 2003

Fine Mapping of Autistic Disorder to Chromosome 15q11-q13 by Use of Phenotypic Subtypes

Yujun Shao,[1] M. L. Cuccaro,[1] E. R. Hauser,[1] K. L. Raiford,[1] M. M. Menold,[1] C. M. Wolpert,[1] S. A. Ravan,[2] L. Elston,[1] K. Decena,[1] S. L. Donnelly,[1] R. K. Abramson,[2] H. H. Wright,[2] G. R. DeLong,[1] J. R. Gilbert,[1] and M. A. Pericak-Vance[1]

[1]Duke University Medical Center, Durham, NC; and [2]W. S. Hall Psychiatric Institute, University of South Carolina, Columbia

Autistic disorder (AutD) is a complex genetic disease. Available evidence suggests that several genes contribute to the underlying genetic risk for the development of AutD. However, both etiologic heterogeneity and genetic heterogeneity confound the discovery of AutD-susceptibility genes. Chromosome 15q11-q13 has been identified as a strong candidate region on the basis of both the frequent occurrence of chromosomal abnormalities in that region and numerous suggestive linkage and association findings. Ordered-subset analysis (OSA) is a novel statistical method to identify a homogeneous subset of families that contribute to overall linkage at a given chromosomal location and thus to potentially help in the fine mapping and localization of the susceptibility gene within a chromosomal area. For the present analysis, a factor that represents insistence on sameness (IS)—derived from a principal-component factor analysis using data on 221 patients with AutD from the repetitive behaviors/stereotyped patterns domain in the Autism Diagnostic Interview–Revised—was used as a covariate in OSA. Analysis of families sharing high scores on the IS factor increased linkage evidence for the 15q11-q13 region, at the *GABRB3* locus, from a LOD score of 1.45 to a LOD score of 4.71. These results narrow our region of interest on chromosome 15 to an area surrounding the γ-aminobutyric acid–receptor subunit genes, in AutD, and support the hypothesis that the analysis of phenotypic homogeneous subtypes may be a powerful tool for the mapping of disease-susceptibility genes in complex traits.

Introduction

Autistic disorder (AutD [MIM *209850; MIM *607373]) is a complex neurodevelopmental disorder characterized by significant disturbances in social, communicative, and behavioral functioning. Initially described by Leo Kanner (1943), the core features of AutD have remained consistent since this early formulation. Onset of AutD occurs before age 3 years; symptoms continue throughout life. AutD is the most studied disorder among a larger clinical group known as "pervasive developmental disorders" (PDDs). Other PDDs include Asperger disorder, Rett syndrome (MIM #312750), childhood disintegrative disorder, and pervasive developmental disorder not otherwise specified. Rett disorder can now clearly be distinguished by the mutations in the *MECP2* gene. The most recent review of multiple epidemiological surveys estimates the prevalence of AutD at ~1 per 1,000 children, with the prevalence for all PDDs at 6 or 7 per 1,000 (Fombonne 2002).

Several twin and family studies have shown strong evidence for genetic factors in the etiology of AutD (Ritvo et al. 1985; Folstein and Piven 1991; Lotspeich and Ciaranello 1993; Szatmari et al. 1998). The sibling-recurrence-risk ratio (λ_s), estimated from epidemiological studies, ranges from 50 to 150 (Smalley et al. 1988; Folstein and Piven 1991; Bailey et al. 1995; International Molecular Genetic Study of Autism Consortium 1998) and is considerably higher than the λ_s estimates for other complex disorders, such as multiple sclerosis (20–40) (Ebers et al. 1995), Alzheimer disease (4–5) (Farrer et al. 1997), and schizophrenia (9–10) (Risch 1990; Kendler et al. 1993). It is estimated that several chromosomal loci contribute to genetic susceptibility in AutD (Pickles et al. 1995; Maestrini et al. 1998). In an effort to identify these genes, several genome screens have been performed (International Molecular Genetic Study of Autism Consortium 1998; Barrett et al. 1999; Philippe et al. 1999; Risch et al. 1999; Collaborative Linkage Study of Autism 2001; Shao et al. 2002*b*). By use of the criteria of LOD score >1 (parametric or sib-pair method) and/or nominal *P* value <.05 between a marker and AutD, a number of regions have been identified as interesting. However, despite the high estimate

Received October 18, 2002; accepted for publication November 26, 2002; electronically published February 3, 2003.

Address for correspondence and reprints: Dr. Margaret A. Pericak-Vance, Center for Human Genetics, Duke University Medical Center, Box 3445, Durham, NC 27710. E-mail: mpv@chg.duhs.duke.edu

0002-9297/2003/7203-0005$15.00

of λ_s, most of these studies have failed to identify loci with the classical threshold for significant evidence of linkage (LOD $\geq$ 3.6) (Lander and Kruglyak 1995). This divergence between epidemiological studies and linkage results is common in complex traits (e.g., AutD) in which both multiple genes and environmental factors are likely to contribute to the underlying disease etiology. In fact, similar findings have been observed in other common complex diseases, such as hypertension (Krushkal et al. 1999; Xu et al. 1999; Levy et al. 2000; Allayee et al. 2001), cardiovascular disease (Rubattu et al. 1996; Broeckel et al. 2002; Harrap et al. 2002), and multiple sclerosis (Haines et al. 1996). Both etiologic heterogeneity and genetic heterogeneity are hypothesized as contributing to the failure to obtain significant LOD scores that has been observed in AutD.

Heterogeneity also confounds efforts to localize and identify the genes underlying AutD. One way to address this complication is to identify homogeneous subsets of families that have the same underlying AutD genetic susceptibility. Several analyses of familial data on AutD have supported the hypothesis that examination of homogeneous subsets of data is a valuable approach in the mapping of AutD loci. Subsets of patients with AutD who were selected on the basis of delayed speech substantially increased evidence for linkage at loci on chromosomes 2 and 7 (Folstein and Mankoski 2000; Bradford et al. 2001; Buxbaum et al. 2001; Shao et al. 2002*a*).

The limitation of stratification analysis, however, is that the subset of families must be defined prior to linkage analyses. Thus, the identification of appropriate trait-related–covariate cutoff values that define genetically homogeneous subsets is crucial to the application of such a stratification approach. New approaches to incorporate covariates into linkage analysis have recently been proposed. Ordered-subset analysis (OSA) is one way to circumvent the definition of arbitrary or a priori cutoffs (Hauser and Boehnke 1998; Ghosh et al. 2000; Hauser et al. 2001). OSA can identify family subsets defined by the level of a trait-related covariate that provides maximal evidence for linkage, without requiring a priori definition of the subset. In this method, families are ranked according to their family-specific covariate values, and family-specific LOD scores are added one by one, in rank order. After each family is added, the maximum LOD scores for the subset of families are determined. The maximum subset LOD (OSA-LOD) score and the families in the subset leading to that score are then identified. The significance of the increase in the OSA-LOD score relative to the overall LOD score is estimated using a permutation test.

The class of behaviors described generally as "repetitive behaviors and stereotyped patterns" (RB) is an important trait-related covariate worthy of investigation in AutD for several reasons: (*a*) Despite that RB is one of the three core domains required for the diagnosis of AutD and has been well defined by the Autism Diagnostic Interview–Revised (ADI-R) (Lord et al. 1994), it is the least studied of the core domains. (*b*) The ADI-R has thoroughly characterized the essential diagnostic features of AutD and provides coverage of a variety of constitutive features within the RB domain; however, the ADI-R RB domain algorithm is designed to determine a cutoff point for diagnosis and lacks the sensitivity to capture the full extent of the severity of either the RB domain or underlying subgroups. This global characterization of RB appears to be inconsistent with the observed clinical patterns and suggests the presence of possible factors within the RB class (a factor analysis aiming to refine this domain is described below, in the "Statistical Analysis" subsection). Cuccaro et al. (in press) used factor analysis on the RB items from the ADI-R and identified as least two distinct factors (see the "Families and Methods" section). (*c*) The high-sibling-concordance pattern in RB makes it a good candidate feature for the identification of homogeneous subgroups. Familiality among sibships with AutD has been observed by Silverman et al. (2002) for select RB items within the ADI-R subarea scores in a sample of 212 sibships. This was the only domain score that showed significant familiality in their data set. (*d*) There is support for the presence of repetitive phenomena in children with Prader-Willi syndrome (PWS [MIM #176270]), which is caused by the loss of the paternal contribution of the proximal part (15q11-q13) of the long arm of chromosome 15 (Ohta et al. 1999). Cytogenetic abnormalities on chromosomal region 15q11-q13 are also among the most frequent cytogenetic abnormalities reported in patients with AutD (Martin et al. 2000). Of individuals with AutD or other PDDs, ~1% have a duplication of the 15q11-q13 region (Wolpert et al. 2000; Boyar et al. 2001). This evidence also leads us to reinvestigate chromosome 15.

In addition to the existence of patients with AutD who have cytogenetic abnormalities in the region of 15q11-q13, there are several other lines of evidence supporting this region as a candidate region that potentially harbors AutD-risk genes (Cook et al. 1998; Rineer et al. 1998; Schroer et al. 1998). That this region contains the disease loci for both PWS and Angelman syndrome (AS [MIM #105830]) raises interest (Ozcelik et al. 1992; Knoll et al. 1993; Wevrick et al. 1994; Jay et al. 1997), since a subset of individuals with PWS or AS have been reported to exhibit behavior similar to that seen in AutD (Arrieta et al. 1994; Demb and Papola 1995; Summers et al. 1995; Steffenburg et al. 1996; Dykens and Kasari 1997). Thus, the 15q11-q13 region is one of the most complex regions of the genome in terms of genome stability, gene expression, and gene imprinting.

Several groups have also identified this region as interesting through linkage studies in autism (Philippe

et al. 1999; Risch et al. 1999; Liu et al. 2001). In addition to the linkage findings, significant linkage disequilibrium (LD) has been reported with AutD and both the γ-aminobutyric acid (GABA)–receptor β3-subunit gene (*GABRB3* [MIM *137192]) and the GABAγ genes (*GABRG1* [MIM *137166], *GABRG2* [MIM *137164], and *GABRG3* [MIM *600233]), at the 15q11-q13 region (Menold et al. 2001). The convergence of the linkage and association data, together with cytogenetic evidence, supports 15q11-q13 as a chromosomal region harboring a susceptibility gene for AutD. However, research results for this region are not consistent. Some studies have not replicated these results (Salmon et al. 1999). This region is so broad that more candidate genes for AutD have been proposed, such as the *APBA2* and *SLP1* genes (Maddox et al. 1999; Smith et al. 2000). Hence, new approaches are needed to further investigate this region. OSA is a powerful method to identify homogeneous subgroups of families with distinct underlying etiologies, to potentially help in the fine mapping and localization of the susceptibility gene for AutD.

Families and Methods

Families

Families were ascertained using clinical referrals and active recruitment through lay organizations providing services to families with AutD. There was no known overlap between the families included in the present study and those in other AutD genetic research studies. Detailed diagnostic evaluations of the family data are as described elsewhere (Ashley-Koch et al. 1999). In brief, the ADI-R (Lord et al. 1994) was used to confirm the clinical diagnosis of AutD. Patients with AutD were included in the study if they were between 3 and 21 years of age. The Vineland Adaptive Behavior Scales interview (survey edition) was administered, to assess developmental level (Sparrow et al. 1984). Discrepancies between ADI-R and clinical findings were resolved using additional clinical measures, including the Autism Diagnostic Observation Schedule–Generic (DiLavore et al. 1995). Only individuals diagnosed with idiopathic AutD were included in the study. Individuals with symptomatology of disorders associated with AutD—such as fragile X syndrome, tuberous sclerosis complex, AS, and PWS—were excluded from the data set. Blood was obtained from patients with AutD and their parents and siblings, for DNA extraction. All aspects of the research study were approved by the institutional review boards at the participating institutions, and informed consent was obtained from all participants or their parents or legal guardians.

Clinical Assessment Instrument

The ADI-R (Lord et al. 1994) was used to confirm the clinical diagnosis of AutD on the basis of *DSM-IV/ICD-10* criteria. The ADI-R is the most widely used research measure for the identification of AutD. This instrument is a standardized semistructured interview used in the differential diagnosis of the PDDs. It consists of a hierarchical series of probes and questions that allow caretakers to provide detailed descriptions of the individual's behaviors. The ADI-R contains items in three content areas, or domains: (1) social interaction, (2) nonverbal or verbal communication, and (3) restricted, repetitive behaviors and interests. The majority of the items yield scores ranging from 0 (absence of feature) to 3 (clear display of feature, with accompanying impairment or disruption). Select groups of items are summed into three domain scores that form the basis for the diagnostic algorithm. The classification of AutD requires that an individual exceed cutoff scores in each of the three content areas (Lord et al. 1997). The ADI-R yields a diagnostic algorithm that accurately and reliably discriminates AutD from other developmental disorders, such as mental retardation or language disorder (Lord et al. 1997).

Statistical Analysis

Two-point parametric linkage analysis of all markers was performed using the Fastlink program of the Linkage software package (Lathrop et al. 1984). To minimize the effect of incorrect penetrance functions, we examined both low-penetrance (affected individuals only) autosomal dominant and autosomal recessive models, with allele frequencies of 1/1,000 and 1/100, respectively (Shao et al. 2002*b*). Marker-allele frequencies were estimated from a set of 50–100 unrelated white individuals unaffected by AutD. In addition to standard likelihood methods, genetic-trait model-independent methods were used to assess linkage. Affected-sib-pair (ASP) analysis was performed by Aspex (Hinds and Risch 1998), which calculates a maximum LOD score.

To reduce the phenotypic heterogeneity underlying AutD, we performed OSA (Hauser and Boehnke 1998; Ghosh et al. 2000) in an effort to identify more homogeneous groups of families on the basis of RB, since RB add to behavioral heterogeneity in AutD. An exploratory factor analysis was conducted to refine the RB domain. Thirteen ADI-R items that represent repetitive interests and behaviors were selected for use in the factor-extraction procedure. The items were drawn from those ADI-R items that cover a range of repetitive motor, sensory, and behavioral phenomena in AutD. Cuccaro et al. (in press) hypothesized that there were at least two distinct groups of RB factors and that the factors would respectively index developmental disability and autism. Given the overlap in autism and mental retardation, it was pro-

posed that at least one factor would be associated with developmental disability. However, a more subtle pattern of autism-specific deficits would also be identified, with an emphasis on compulsions and difficulties with change. The factor analysis was run using the ADI-R protocols of 221 independent probands from our data set. Principal-component factor analysis using a correlation matrix with varimax rotation (the most commonly used orthogonal rotation that maximizes the variance of the loadings) yielded a two-factor solution. Item loadings with absolute values >0.35 were used to describe the factors. The first factor consisted of seven items collectively referred to as "repetitive sensory and motor behaviors and interests" (RSMB) (ADI-R items 81, 72, 77, 71, 84, 85, and 76). The second factor consisted of three items collectively referred to as "insistence on sameness" (IS) (ADI-R items 73, 74, and 75). This second factor is consistent with the early conceptualizations of Kanner (1943), who identified maintenance of or insistence on sameness as a core feature of autism. Table 1 lists the factors and loadings.

The ADI-R RB domain score and the RSMB and IS factors were then analyzed separately as covariates in OSA. The OSA method proceeds as follows: First, the families are ranked according to their mean sibship value for the factor score (e.g., the IS score), from highest to lowest. Second, starting with the family with the highest IS score, we add family-specific LOD scores (LOD scores computed separately for each family) one by one, in rank order, until all families are included. Third, after each family is added, the maximum LOD score for the current subset of families is determined and is denoted as the "subset LOD score"; the highest of the subset LOD scores is determined, along with the families in the subset leading to that OSA-LOD score. Finally, the family-specific multipoint LOD scores are calculated using the Genehunter-Plus software package, on the basis of the information from all available affected relative pairs (ARPs) and with the incorporation of Kong and Cox's model (Kruglyak et al. 1996; Kong and Cox 1997).

The *P* value for the significance of the increase in the OSA-LOD score from the baseline LOD score was estimated by using a permutation test of the hypothesis that the covariate-defined subset yields a significant increase in the LOD score as compared with that for randomly ordered families. Simulation shows that this procedure preserves the appropriate type I error rate and can increase power to detect a locus in the presence of heterogeneity (Hauser et al. 2001; E. R. Hauser, R. M. Watanabe, W. L. Duren, M. P. Bass, C. D. Langefeld, and M. Boehnke, unpublished data).

Multivariate analysis of variance (MANOVA) was used to compare the two subsets (i.e., the subset identified by OSA and the other subset comprising the rest of the data) on the area scores of communication, social interaction, and restricted- or repetitive-behavior ADI-R domain scores. Since these scores are highly correlated and the observations in siblings are not independent, the univariate *t* test is not appropriate here. The three domain scores served as the set of response variables in MANOVA, in looking for clinical feature difference between these two groups. The dependent observations in siblings were also adjusted. PROC GLM (SAS) was used.

Table 1

Factor Loadings for RSMB and IS

ADI-R Item and Variable	Loading[a] Factor 1	Factor 2
RSMB:		
81. Hand and finger mannerisms	***.72***	.06
72. Repetitive use of objects	***.69***	.06
77. Unusual sensory interests	***.58***	.20
71. Unusual preoccupations	***.52***	.07
84. Complex mannerisms	***.49***	.29
85. Rocking	***.43***	−.03
76. Unusual attachment to objects	***.42***	.13
IS:		
73. Difficulties with minor changes in personal routine or environment	−.01	***.80***
74. Resistance to trivial changes in environment	.13	***.69***
75. Compulsions/rituals	.20	***.59***
Other:		
70. Circumscribed interests	.29	.12
36. Sensitivity to noise	.27	.04
78. Idiosyncratic negative response to specific sensory stimuli	.14	.30

[a] The values shown in boldface italics are those item loadings that had absolute values >.35, the level considered significant to describe the factors.

Genotyping

Genomic DNA was extracted from whole blood by using the Pure Gene method and standard protocols (Vance 1998). PCR was performed in 96-well microtiter plates by using 30 ng of total genomic DNA. The individual 10-μl PCR mixtures contained 1 × PCR buffer, 0.2 mM of dNTPs, 1.5 mM of $MgCl_2$, 0.5 μM of each primer, and 0.5 U of Platinum *Taq* (Life Technologies). PCRs were run on Hybaid Touchdown thermal cyclers (ThermoHybaid US) by using the following touchdown program: 1 cycle of 3 min at 94°C; 2 cycles each of 5 s at 94°C, 30 s at 65°C, 63°C, 61°C, 59°C, or 57°C, and 30 s at 72°C; 30 cycles of 5 s at 94°C, 30 s at 55°C, and 45 s at 72°C; and 1 cycle of 3 min at 72°C. DNA fragments were separated on 54-well 6% denaturing acrylamide gels, were subsequently visualized by staining with SYBR gold (Molecular Probes), and were scanned on a Hitachi FMBIO II fluorimager (Hitachi Instruments). Gel images were analyzed with Bioimage software (Genomic Solutions), and the resulting data were sent through the CHG Data Coordinating Center

(Rimmler et al. 1999), for quality-control (QC) analysis. To minimize systematic errors due to sample switches, gel loading, or reading errors, we added QC samples to each gel analyzed in the laboratory. Before the merging of genotypic data, agreement between the genotypes for the QC and its matching sample was required.

Results

In the 15q11-q13 region (extending from 9 cM to 23 cM), eight markers were genotyped: GATA143C02, GABRB3, D15S97, GABRA5, D15S822, D15S975, D15S156, and D15S217. Eighty-one multiplex families with completed genotypes and with completed covariate information were included in the present analysis. Table 2 shows the baseline LOD score for the entire AutD data set. Neither the ADI-R domain score nor the RSMB component showed any significant (empirical $P < .05$) increase in OSA-LOD score over the baseline LOD score. However, OSA using the IS component as a covariate yielded a significant increase in LOD scores.

The peak OSA-LOD score (based on the nonparametric ARP method) for this region is 3.19 at marker GABRB3, for a subset of 23 families with the highest mean IS scores. These families are those with affected individuals who have the highest scores on the IS factor. Thus, the mean IS score can be construed as a potential index of severity in that specific domain: higher scores reflect greater impairment associated with IS. Figure 1 compares the multipoint LOD scores for the baseline group and for the IS subset. The nonparametric ARP-method–based LOD score for the entire data set under the same maximization parameters at marker GABRB3 is 0.18. The increase in the LOD score is 3.01. By use of a permutation-test framework (Hauser and Boehnke 1998; Hauser et al. 2001), the observed OSA-LOD score was compared with the OSA-LOD score obtained at any point along the chromosome when families were added in random order. An empirical P value of .0095 was obtained as the proportion of 5,000 random orderings of the families that gave an OSA-LOD score greater than the observed score.

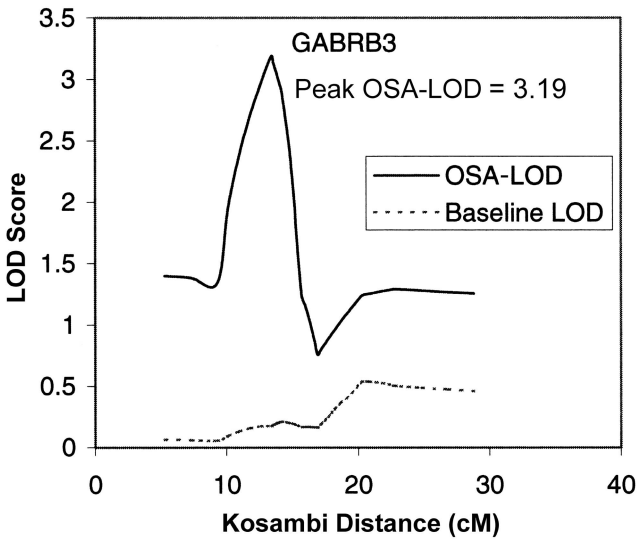

Figure 1 Nonparametric multipoint OSA results for the chromosome 15 region.

LOD scores for the IS subset (23 families with the highest IS scores) were compared with the non-IS subset (the rest of the 81-family subset), applying traditional parametric two-point linkage analysis using the Fastlink software package (Lathrop et al. 1984). The parametric LOD scores for these two subsets are shown in table 2. At marker GABRB3, the IS subset yields LOD scores of 4.71, under the dominant affecteds-only model, and 3.83, under the recessive affecteds-only model. The non-IS subset yields LOD scores of 0.16, under the dominant

Table 2

LOD Scores for the Entire Data Set and the IS and Non-IS Subsets

	Entire Data Set ($N = 81$)			IS Subset ($N = 23$)			Non-IS Subset ($N = 58$)		
	LOD Score[a]			LOD Score[a]			LOD Score[a]		
Marker	Dominant[c]	Recessive[c]	MLS[b]	Dominant[c]	Recessive[c]	MLS[b]	Dominant[c]	Recessive[c]	MLS[b]
GATA143C02	−.107	−.07	0	.7	.72	.25	−.29	−.28	0
GABRB3	***1.40***	***1.07***	***.21***	***4.71***	***3.83***	***2.72***	***.16***	***.04***	***0***
D15S97	.04	−.04	0	1.41	1.12	1.31	−.08	−.21	0
GABRA5	−.04	−.09	0	.4	.31	.26	−.11	−.14	0
D15S822	.35	.16	.01	1.63	1.10	1	−.05	−.07	0
D15S975	−.00	.17	.22	1.26	1.09	.33	−.20	−.07	.21
D15S156	.56	.56	.01	1.28	1.35	.5	.11	.12	0
D15S217	−.05	.12	.22	1.03	.80	.86	−.30	−.18	.04

[a] Two-point maximum LOD score based on parametric linkage analysis method.
[b] Two-point maximum LOD score based on nonparametric ASP method.
[c] Low-penetrance model.

affecteds-only model, and 0.04, under the recessive affecteds-only model.

The clinical features for the IS subset were compared with those for the non-IS subset. Table 3 shows the comparisons between the two groups for ADI-R domain scores for social interaction, verbal and nonverbal communication, and repetitive behaviors. Since these domain scores are highly correlated, MANOVA was performed to adjust for correlation between dependent variables and correlation within families. As expected, there was a significant overall difference between these two groups ($P = .0043$), since there was a significant difference with respect to the repetitive-behaviors domain score ($P = .0006$) when families were ranked by IS score. Meanwhile, there were also significant differences between the two groups with respect to the social-interaction domain score ($P = .007$). Although the nonverbal-communication score did not show statistical significance between two groups ($P = .08$), the verbal-communication score has shown difference ($P = .02$). The verbal comparison involved fewer participants (65% of patients), because some affected individuals remain nonverbal and therefore are not assigned a verbal-communication domain score under the ADI-R. The mean values for each group, after adjustment for the other factors, are shown in table 3.

Table 3

Means and Multivariate and Univariate Results: Comparison between the IS and Non-IS Subsets

		Adjusted Mean for	
	P	IS Subset ($N = 46$)	Non-IS Subset ($N = 122$)
Multivariate test	.0043		
Univariate test:			
Social interaction	.007	20.43	16.65
Repetitive behaviors	.0006	7.78	5.84
Nonverbal communication	.08	9.43	8.14
Verbal communication	.02	16.18	13.94

Discussion

The purpose of the OSA approach is to reduce the impact that genetic heterogeneity has on linkage findings in AutD by the identification of a subset of families with a more homogeneous phenotype, resulting in increased overall evidence for linkage to a chromosomal region with a more distinct LOD-score peak (Hauser and Boehnke 1998; Ghosh et al. 2000; Hauser et al. 2001). Identification of a subset of families with enhanced linkage evidence is an efficient method to begin the search for trait-associated polymorphisms or mutation analysis, since these families are more likely to contain the susceptibility gene in question than are families not included in the subset. In addition, identification of a trait-related covariate by using the OSA method helps to refine the linkage region and the list of candidate genes in this region on the basis of biological pathways that connect the covariate and the trait. By use of OSA, a phenotypically homogeneous subset of families that most likely contribute to the linkage on chromosome 15 in our data set has been identified, and the linkage region is narrowed. This information will be invaluable in our ongoing molecular analysis of this region.

The overall behavior of OSA has been shown in simulation studies (Hauser et al. 2001). Results show that the OSA method preserves the appropriate type I error rate and can increase the power to detect a locus in the presence of heterogeneity. In addition to the simulation results, OSA has been applied and verified in various studies. OSA replicated the linkage results for breast cancer and chromosome 17q that have been reported by Hall et al. (1990), in their localization of *BRCA1*. The performance of the OSA method has also been examined in a genome screen performed on 719 ASPs from the Finland–United States Investigation of Non–Insulin-Dependent Diabetes Mellitus Genetics study (Valle et al. 1998; Ghosh et al. 2000), an international effort to identify genes for type 2 diabetes. Results for specific traits and chromosomal regions have been presented by Ghosh et al. (2000). Given the variation in observed overall LOD score across a genome screen, the *P* value obtained using the permutation procedure is of crucial importance in evaluating the increase in the linkage evidence when subsetting families. The successful performance of OSA in our data set illustrated that OSA is a very helpful method in follow-up analysis.

Identification of the IS subset in AutD is noteworthy for several reasons. To date, the ADI-R has been used primarily as a diagnostic algorithm. Phenotype refinement using the ADI-R has so far been limited to investigation of language subtypes. The finding of at least two distinct factors within the RB domain conforms to clinical evidence of two phenomena in that domain: (1) RSMB, a global developmental-disability factor in which repetitive motor and sensory phenomena are common, and (2) IS, a factor thought to be more specific to AutD. The IS factor is consistent with the earliest conceptualizations of AutD, in which maintenance of sameness was a core feature (Kanner 1943). In addition, our findings extend the work of Silverman et al. (2002), who have found evidence of familiality for repetitive behaviors in ASPs with AutD. Moderate familiality of the RSMB and IS factors among the ASPs with AutD was also observed in the present data set. Identification of a subset of individuals with AutD who display greater difficulties with compulsions and resistance to change may provide clues to specific brain regions involved in such difficulties. Finally, the IS factor may be associated with greater social deficits,

as indicated by a higher score on the ADI-R socialization domain for this subset. This establishes a preliminary relationship between one core feature of AutD (social impairment) and the repetitive behaviors (e.g., the IS factor) within this group of families.

That no significant increase in OSA-LOD score was obtained by using as a covariate the RB domain score from the ADI-R algorithm also supports the observation that the ADI-R and its various domain scores are constructed primarily for the purpose of diagnosis. The refinement of the RB domain by analysis tools such as principal-component factor analysis and by using the derived factor scores is the key to rendering the domain differential enough to catch the severity gradient of the feature. With the aid of new statistical methods such as OSA, the continuous covariate can be successfully utilized to narrow the susceptibility-gene region. That the IS factor but not the RB factor increased the linkage evidence for AutD in the 15q11-q13 region may be explained by the IS factor's subsuming the phenomenon, often observed in patients with PWS, of obsessive-compulsive behavior, which therefore may be caused by a deletion or disruption of a gene or several genes in the 15q11-q13 region, where the PWS disease locus resides. OSAs using the IS factor score, the RB factor score, and the RSMB domain score to study the other candidate regions of interest, on chromosomes 2 and 7, that have previously been reported (Ashley-Koch et al. 1999; Shao et al. 2002*b*) did not improve the LOD scores, but there is a possibility that the RB factor score would be useful as a covariate in the OSA of the other chromosomal regions.

Our results also strengthen the evidence for linkage between AutD and $GABA_A$. $GABA_A$ is the chief inhibitory neurotransmitter in the human brain, binding and exerting its effects through a complex series of $GABA_A$ receptors. Genes for at least three of the $GABA_A$-receptor subunits ($\beta 3$, $\alpha 5$, and $\gamma 3$) lie in the 15q11-q13 region and are excellent candidate genes for AutD. The peak linkage marker in our data set, GABRB3, is ~60 kb beyond the 3′ end of $\beta 3$-subunit gene. By family-based association analysis, significant linkage disequilibrium has been found between an AutD-susceptibility locus and a marker, GABRB3 155CA-2, in the GABA $\beta 3$-subunit gene, on chromosome 15q11-q13 (Cook et al. 1998; Silverman et al. 2002). Additional evidence for linkage disequilibrium between AutD and marker GABRB3 has been found by Bass et al. (2000) and Menold et al. (2001).

Previously reported duplications of the 15q11-q13 region in AutD have been exclusively of maternal origin. Several reports have suggested that duplications of maternal but not paternal origin increase the risk for developmental disorders (Browne et al. 1997; Cook et al. 1997; Ashley-Koch et al. 1999). Increased maternal allele sharing was found in the IS subset. Maternal sharing versus paternal sharing was respectively 58.5% versus 50% in the entire data set. In the IS subset, maternal sharing versus paternal sharing was respectively 93.5% versus 66.7%, with respective maximum LOD scores of 2.97 versus −0.06. The evidence for a significant increase in maternal sharing not only suggests the involvement of imprinting genes in the region but also serves as a clue to the underlying biological mechanisms. Meanwhile, our analysis also consistently showed that using the IS factor to identify a distinct clinical subgroup is meaningful in candidate-gene study in that chromosomal region. Therefore, behavioral phenotypes can potentially be highly valuable clues to the identification of genes responsible for behavioral abnormalities.

In summary, these data support previous reports implicating the role that one or more genes in the chromosome 15q11-q13 region play in AutD susceptibility and potentially narrow the candidate-gene region. The GABA $\beta 3$-subunit gene is a leading candidate gene for AutD susceptibility. These findings also suggest that empirically derived behavioral factors are valuable in the characterization of AutD data sets for linkage analysis and, thus, that phenotypic subgrouping is a powerful tool in the mapping and identification of genes for complex traits.

Acknowledgments

We thank the autistic patients and their families for their participation, without which this research would have been impossible. This work was supported, in part, by National Institutes of Health (NIH) Program Project Grant NS26630, NIH R01 Grants HD36701 and NS36768, and the National Alliance of Autism Research, through a gift from Audrey Flack and H. Robert Marcus.

Electronic-Database Information

Accession numbers and the URL for data presented herein are as follows:

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for AutD [MIM *209850; MIM *607373], Rett syndrome [MIM #312750], PWS [MIM #176270], AS [MIM #105830], *GABRB3* [MIM *137192], *GABRG1* [MIM *137166], *GABRG2* [MIM *137164], and *GABRG3* [MIM *600233])

References

Allayee H, de Bruin TW, Michelle DK, Cheng LS, Ipp E, Cantor RM, Krass KL, Keulen ET, Aouizerat BE, Lusis AJ, Rotter JI (2001) Genome scan for blood pressure in Dutch dyslipidemic families reveals linkage to a locus on chromosome 4p. Hypertension 38:773–778

Arrieta I, Lobato MN, Martinez B, Criado B (1994) Parental origin of Robertsonian translocation (15q22q) and Prader-

Willi syndrome associated with autism. Psychiatr Genet 4: 63–65

Ashley-Koch A, Wolpert CM, Menold MM, Zaeem L, Basu S, Donnelly SL, Ravan SA, Powell CM, Qumsiyeh MB, Aylsworth AS, Vance JM, Gilbert JR, Wright HH, Abramson RK, DeLong GR, Cuccaro ML, Pericak-Vance MA (1999) Genetic studies of autistic disorder and chromosome 7. Genomics 61: 227–236

Bailey A, Le Couteur 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

Barrett S, Beck JC, Bernier R, Bisson E, Braun TA, Casavant TL, Childress D, et al (1999) An autosomal genomic screen for autism: Collaborative Linkage Study of Autism. Am J Med Genet 88:609–615

Bass MP, Menold MM, Wolpert CM, Donnelly SL, Ravan SA, Hauser ER, Maddox LO, Vance JM, Abramson RK, Wright HH, Gilbert JR, Cuccaro ML, DeLong GR, Pericak-Vance MA (2000) Genetic studies in autistic disorder and chromosome 15. Neurogenetics 2:219–226

Boyar FZ, Whitney MM, Lossie AC, Gray BA, Keller KL, Stalker HJ, Zori RT, Geffken G, Mutch J, Edge PJ, Voeller KS, Williams CA, Driscoll DJ (2001) A family with a grandmaternally derived interstitial duplication of proximal 15q. Clin Genet 60:421–430

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

Broeckel U, Hengstenberg C, Mayer B, Holmer S, Martin LJ, Comuzzie AG, Blangero J, Nurnberg P, Reis A, Riegger GA, Jacob HJ, Schunkert H (2002) A comprehensive linkage analysis for myocardial infarction and its related risk factors. Nat Genet 30:210–214

Browne CE, Dennis NR, Maher E, Long FL, Nicholson JC, Sillibourne J, Barber JC (1997) Inherited interstitial duplications of proximal 15q: genotype-phenotype correlations. Am J Hum Genet 61:1342–1352

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

Collaborative Linkage Study of Autism (2001) An autosomal genomic screen for autism. Am J Med Genet 105:609–615

Cook EH Jr, Courchesne RY, Cox NJ, Lord C, Gonen D, Guter SJ, Lincoln A, Nix K, Haas R, Leventhal BL, Courchesne E (1998) Linkage-disequilibrium mapping of autistic disorder, with 15q11-13 markers. Am J Hum Genet 62:1077–1083

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

Cuccaro ML, Bass MP, Abramson RK, Ravan SA, Wright HH, Wolpert CM, Donnelly SL, Pericak-Vance MA. Behavioral comparisons in autistic individuals from multiplex and singleton families. J Autism Dev Disord (in press)

Demb HB, Papola P (1995) PDD and Prader-Willi syndrome. J Am Acad Child Adolesc Psychiatry 34:539–540

DiLavore PC, Lord C, Rutter M (1995) The pre-linguistic autism diagnostic observation schedule. J Autism Dev Disord 25:355–379

Dykens EM, Kasari C (1997) Maladaptive behavior in children with Prader-Willi syndrome, Down syndrome, and nonspecific mental retardation. Am J Ment Retard 102:228–237

Ebers GC, Sadovnick AD, Risch NJ (1995) A genetic basis for familial aggregation in multiple sclerosis. Nature 377: 150–151

Farrer LA, Cupples LA, Haines JL, Hyman B, Kukull WA, Mayeux R, Myers RH, Pericak-Vance MA, Risch N, van Duijn CM, APOE and Alzheimer Disease Meta Analysis Consortium (1997) Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease: a meta-analysis. JAMA 278:1349–1356

Folstein SE, Mankoski RE (2000) Chromosome 7q: where autism meets language disorder? Am J Hum Genet 67:278–281

Folstein SE, Piven J (1991) Etiology of autism: genetic influences. Pediatrics 87:767–773

Fombonne E (2002) Epidemiological trends in rates of autism. Mol Psychiatry 7 Suppl 2:S4–S6

Ghosh S, Watanabe RM, Valle TT, Hauser ER, Magnuson VL, Langefeld CD, Ally DS, et al (2000) The Finland–United States Investigation of Non–Insulin-Dependent Diabetes Mellitus Genetics (FUSION) study. I. An autosomal genome scan for genes that predispose to type 2 diabetes. Am J Hum Genet 67:1174–1185

Haines JL, Ter-Minassian M, Bazyk A, Gusella JF, Kim DJ, Terwedow H, Pericak-Vance MA, et al (1996) A complete genomic screen for multiple sclerosis underscores a role for the major histocompatability complex. Nat Genet 13: 469–471

Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B, King MC (1990) Linkage of early-onset familial breast cancer to chromosome 17q21. Science 250:1684–1689

Harrap SB, Zammit KS, Wong ZY, Williams FM, Bahlo M, Tonkin AM, Anderson ST (2002) Genome-wide linkage analysis of the acute coronary syndrome suggests a locus on chromosome 2. Arterioscler Thromb Vasc Biol 22:874–878

Hauser ER, Bass MP, Martin ER, Watanabe RM, Duren WL, Boehnke M (2001) Power of the ordered subset method for detection and localization of genes in linkage analysis of complex traits. Am J Hum Genet Suppl 69:529

Hauser ER, Boehnke M (1998) Genetic linkage analysis of complex genetic traits by using affected sibling pairs. Biometrics 54:1238–1246

Hinds D, Risch N (1998) The Aspex package: affected sib pair exclusion mapping. Stanford University, Stanford, CA

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

Jay P, Rougeulle C, Massacrier A, Moncla A, Mattei MG, Malzac P, Roeckel N, Taviaux S, Lefranc JL, Cau P, Berta P, Lalande M, Muscatelli F (1997) The human necdin gene, NDN, is maternally imprinted and located in the Prader-Willi syndrome chromosomal region. Nat Genet 17:357–361

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

Kendler KS, McGuire M, Gruenberg AM, O'Hare A, Spellman M, Walsh D (1993) The Roscommon Family Study. I. Methods, diagnosis of probands, and risk of schizophrenia in relatives. Arch Gen Psychiatry 50:527–540

Knoll JHM, Wagstaff J, Lalande M (1993) Cytogenetic and molecular studies in the Prader-Willi and Angelman syndromes: an overview. Am J Med Genet 46:2–6

Kong A, Cox NJ (1997) Allele-sharing models: LOD scores and accurate linkage tests. Am J Hum Genet 61:1179–1188

Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet 58:1347–1363

Krushkal J, Ferrell R, Mockrin SC, Turner ST, Sing CF, Boerwinkle E (1999) Genome-wide linkage analyses of systolic blood pressure using highly discordant siblings. Circulation 99:1407–1410

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

Lathrop GM, Lalouel JM, Julier C, Ott J (1984) Strategies for multilocus linkage analysis in humans. Proc Natl Acad Sci USA 81:3443–3446

Levy D, Destefano AL, Larson MG, O'Donnell CJ, Lifton RP, Gavras H, Cupples LA, Myers RH (2000) Evidence for a gene influencing blood pressure on chromosome 17: genome scan linkage results for longitudinal blood pressure phenotypes in subjects from the Framingham Heart Study. Hypertension 36:477–483

Liu J, Nyholt DR, Magnussen P, Parano E, Pavone P, Geschwind D, Lord C, Iversen P, Hoh J, Ott J, Gilliam TC (2001) A genomewide screen for autism susceptibility loci. Am J Hum Genet 69:327–340

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

Lord C, Rutter M, Le Couteur 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

Lotspeich LJ, Ciaranello RD (1993) The neurobiology and genetics of infantile autism. Int Rev Neurobiol 35:87–129

Maddox LO, Menold MM, Bass MP, Rogala AR, Pericak-Vance MA, Vance JM, Gilbert JR (1999) Autistic disorder and chromosome 15q11-q13: construction and analysis of a BAC/PAC contig. Genomics 62:325–331

Maestrini E, Marlow AJ, Weeks DE, Monaco AP (1998) Molecular genetic investigations of autism. J Autism Dev Disord 28:427–437

Martin ER, Menold MM, Wolpert CM, Bass MP, Donnelly SL, Ravan SA, Zimmerman A, Gilbert JR, Vance JM, Maddox LO, Wright HH, Abramson RK, DeLong GR, Cuccaro ML, Pericak-Vance MA (2000) Analysis of linkage disequilibrium in γ-aminobutyric acid receptor subunit genes in autistic disorder. Am J Med Genet 96:43–48

Menold MM, Shao YJ, Wolpert CM, Donnelly SL, Raiford KL, Martin ER, Ravan SA, Abramson RK, Wright HH, DeLong GR, Cuccaro ML, Pericak-Vance MA, Gilbert JR (2001) Association analysis of chromosome 15 $GABA_A$ receptor subunit genes in autistic disorder. J Neurogenet 15: 245–259

Ohta T, Gray TA, Rogan PK, Buiting K, Gabriel JM, Saitoh S, Muralidhar B, Bilienska B, Krajewska-Walasek M, Driscoll DJ, Horsthemke B, Butler MG, Nicholls RD (1999) Imprinting-mutation mechanisms in Prader-Willi syndrome. Am J Hum Genet 64:397–413

Ozcelik T, Leff S, Robinson W, Donlon T, Lalande M, Sanjines E, Schinzel A, Francke U (1992) Small nuclear ribonucleoprotein polypeptide N (SNRPN), an expressed gene in the Prader-Willi syndrome critical region. Nat Genet 2:265–269

Philippe A, Martinez M, Guilloud-Bataille M, Gillberg C, Råstam M, Sponheim E, Coleman M, Zappella M, Aschauer H, Van Maldergem L, Penet C, Feingold J, Brice A, Leboyer M, Paris Autism Research International Sibpair Study (1999) Genome-wide scan for autism susceptibility genes. Hum Mol Genet 8:805–812

Pickles A, Bolton P, Macdonald H, Bailey A, Le Couteur A, Sim CH, 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

Rimmler JB, Haynes CS, McDowell JG, Stajich JE, Adams CS, Slotterbeck BD, Rogala AR, West SG, Gilbert JR, Hauser ER, Vance JM, Pericak-Vance MA (1999) DataTracker: comprehensive software for data quality control protocols in complex disease studies. Am J Hum Genet Suppl 65:A442

Rineer S, Finucane B, Simon EW (1998) Autistic symptoms among children and young adults with isodicentric chromosome 15. Am J Med Genet 81:428–433

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

Risch N, Spiker D, Lotspeich L, Nouri N, Hinds D, Hallmayer J, Kalaydjieva L, et al (1999) A genomic screen of autism: evidence for a multilocus etiology. Am J Hum Genet 65: 493–507

Ritvo ER, Freeman BJ, Mason-Brothers A, Mo A, Ritvo AM (1985) Concordance for the syndrome of autism in 40 pairs of afflicted twins. Am J Psychiatry 142:74–77

Rubattu S, Volpe M, Kreutz R, Ganten U, Ganten D, Lindpaintner K (1996) Chromosomal mapping of quantitative trait loci contributing to stroke in a rat model of complex human disease. Nat Genet 13:429–434

Salmon B, Hallmayer J, Rogers T, Kalaydjieva L, Petersen PB, Nicholas P, Pingree C, McMahon W, Spiker D, Lotspeich L, Kraemer H, McCague P, Dimiceli S, Nouri N, Pitts T, Yang J, Hinds D, Myers RM, Risch N (1999) Absence of linkage and linkage disequilibrium to chromosome 15q11-q13 markers in 139 multiplex families with autism. Am J Med Genet 88:551–556

Schroer RJ, Phelan MC, Michaelis RC, Crawford EC, Skinner SA, Cuccaro M, Simensen RJ, Bishop J, Skinner C, Fender D, Stevenson RE (1998) Autism and maternally derived aberrations of chromosome 15q. Am J Med Genet 76:327–336

Shao Y, Raiford KL, Wolpert CM, Cope HA, Ravan SA, Ashley-Koch AA, Abramson RK, Wright HH, DeLong RG, Gilbert JR, Cuccaro ML, Pericak-Vance MA (2002*a*) Pheno-

typic homogeneity provides increased support for linkage on chromosome 2 in autistic disorder. Am J Hum Genet 70: 1058–1061

Shao Y, Wolpert CM, Raiford KL, Menold MM, Donnelly SL, Ravan SA, Bass MP, McClain C, von Wendt L, Vance JM, Abramson RH, Wright HH, Ashley-Koch A, Gilbert JR, DeLong RG, Cuccaro ML, Pericak-Vance MA (2002*b*) Genomic screen and follow-up analysis for autistic disorder. Am J Med Genet 114:99–105

Silverman JM, Smith CJ, Schmeidler J, Hollander E, Lawlor BA, Fitzgerald M, Buxbaum JD, Delaney K, Galvin P (2002) Symptom domains in autism and related conditions: evidence for familiality. Am J Med Genet 114:64–73

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

Smith M, Filipek PA, Wu C, Bocian M, Hakim S, Modahl C, Spence MA (2000) Analysis of a 1-megabase deletion in 15q22-q23 in an autistic patient: identification of candidate genes for autism and of homologous DNA segments in 15q22-q23 and 15q11-q13. Am J Med Genet 96:765–770

Sparrow SS, Balla D, Cicchetti D (1984) Vineland Adaptive Behavior Scales, interview edition. American Guidance Service, Circle Pines, MN

Steffenburg S, Gillberg CL, Steffenburg U, Kyllerman M (1996) Autism in Angelman syndrome: a population-based study. Pediatr Neurol 14:131–136

Summers JA, Allison DB, Lynch PS, Sandler L (1995) Behaviour problems in Angelman syndrome. J Intellect Disabil Res 39:97–106

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

Valle T, Tuomilehto J, Bergman RN, Ghosh S, Hauser ER, Eriksson J, Nylund S, Kohtamaki K, Tuomilehto-Wolf E, Toivanen L, Vidgren G, Ehnholm C, Blaschak J, Langefeld CD, Watanabe RM, Magnuson V, Ally DS, Hagopian W, Ross E, Buchanan TA, Collins F, Boehnke M (1998) Mapping genes for NIDDM: design of the Finland–United States Investigation of NIDDM Genetics (FUSION) study. Diabetes Care 21:949–958

Vance JM (1998) The collection of biological samples for DNA analysis. In: Haines JL, Pericak-Vance MA (eds) Approaches to gene mapping in complex human diseases. Wiley-Liss, New York, pp 201–211

Wevrick R, Kerns JA, Francke U (1994) Identification of a novel paternally expressed gene in the Prader-Willi syndrome region. Hum Mol Genet 3:1877–1882

Wolpert CM, Menold MM, Bass MP, Qumsiyeh MB, Donnelly SL, Ravan SA, Vance JM, Gilbert JR, Abramson RK, Wright HH, Cuccaro ML, Pericak-Vance MA (2000) Three probands with autistic disorder and isodicentric chromosome 15. Am J Med Genet 96:365–372

Xu X, Rogus JJ, Terwedow HA, Yang J, Wang Z, Chen C, Niu T, Wang B, Xu H, Weiss S, Schork NJ, Fang Z (1999) An extreme-sib-pair genome scan for genes regulating blood pressure. Am J Hum Genet 64:1694–1701