Published in final edited form as:
J Am Acad Child Adolesc Psychiatry. 2011 July ; 50(7): 687–696.e13. doi:10.1016/j.jaac.2011.05.002.

Identification of genetic loci underlying the phenotypic constructs of autism spectrum disorders Running head: Genetic loci for latent factors in ASD

Xiao-Qing Liu, M.D., **Stelios Georgiades, M.A.**, **Eric Duku, M.Sc.**, **Ann Thompson, M.S.**, **Bernie Devlin, Ph.D.**, **Edwin H. Cook, M.D.**, **Ellen M. Wijsman, Ph.D.**, **Andrew D. Paterson, M.D.**, and **Peter Szatmari, M.D.**
Dr. Liu is with the University of Manitoba, Canada. Dr. Szatmari, Mr. Georgiades, Mr. Duku, and Ms. Thompson are with The Offord Centre for Child Studies, McMaster University, Canada. Dr. Devlin is with the University of Pittsburgh School of Medicine. Dr. Cook is with the Institute for Juvenile Research, University of Illinois at Chicago. Dr. Wijsman is with the University of Washington. Dr. Paterson is with the Program in Genetics and Genome Biology, The Hospital for Sick Children, Canada.

Abstract

Objective—To investigate the underlying phenotypic constructs in autism spectrum disorders (ASD) and to identify genetic loci that are linked to these empirically derived factors.

Method—Exploratory factor analysis was applied to two datasets with 28 selected Autism Diagnostic Interview-Revised (ADI-R) algorithm items. The first dataset was from the Autism Genome Project (AGP) phase I (1,236 ASD subjects from 618 families); the second was from the AGP phase II (804 unrelated ASD subjects). Variables derived from the factor analysis were then used as quantitative traits in genome-wide variance components linkage analyses.

Results—Six factors, joint attention, social interaction and communication, non-verbal communication, repetitive sensory-motor behaviour, peer interaction, and compulsion/restricted interests, were retained for both datasets. There was good agreement between the factor loading patterns from the two datasets. All factors showed familial aggregation. Suggestive evidence for linkage was obtained for the joint attention factor on 11q23. Genome-wide significant evidence for linkage was obtained for the repetitive sensory-motor behaviour factor on 19q13.3.

Conclusions—This study demonstrates that the underlying phenotypic constructs based on the ADI-R algorithm items are replicable in independent datasets; and the empirically derived factors are suitable and informative in genetic studies of ASD.

Correspondence to: Peter Szatmari, M.D. Offord Centre for Child Studies Hamilton, Ontario Canada L8N 3Z5 Phone: (905) 521-2100 ext. 77367 Fax: (905) 574-6665 szatmar@mcmaster.ca.

Disclosure: Dr. Cook has served as a consultant for Seaside Therapeutics, and is a site principal investigator for a clinical trial sponsored by Seaside Therapeutics. Drs. Liu, Devlin, Wijsman, Paterson, and Szatmari, and Mr. Georgiades, Mr. Duku, and Ms. Thomspon report no biomedical financial interests or potential conflicts of interest.

Supplemental material cited in this article is available online.

Keywords

autism; ADI-R; factor analysis; linkage analysis; quantitative trait

Introduction

Autism spectrum disorders (ASD) are a group of neurodevelopmental disorders, including autistic disorder, Asperger syndrome, and pervasive developmental disorder not otherwise specified, that are characterized by different degrees of 1) deficits in social interaction; 2) deficits in verbal and non-verbal communication; and 3) repetitive and stereotyped behaviours and interests.[1,2] This three-domain conceptualization of ASD is primarily based on clinical acumen rather than empirical evidence.[3]

To date, many studies have examined the underlying phenotypic dimensions of ASD.[4-13] However, possibly due to differences in statistical methods, item selection, sample ascertainment, sample sizes, and other aspects, the results appear to be inconsistent with empirically derived factors ranging from one to six.[5] Four studies have incorporated such factors in genetic analyses of ASD.[14-17] Cannon et al.[14] used two variables, insistence on sameness (IS) and repetitive sensory-motor actions (RSMA), in linkage analyses. However, no factor analysis was performed. Instead the variables were obtained by dichotomizing the sum of IS or RSMA items from the Autism Diagnostic Interview-Revised (ADI-R) 'repetitive behaviours/stereotyped patterns' domain. The other three studies[15-17] did not use the derived factors as primary traits in their genetic analyses, rather the ASD diagnosis was the primary trait and factors were used to define subsets of families to increase genetic homogeneity.

Compared to binary traits (e.g. presence or absence of ASD), well-defined quantitative traits may be more informative for genetic studies because they correspond better to ASD as syndromes with a spectrum of severities. A number of studies[18-23] have used quantitative phenotypes, especially, the composite domain total scores from ADI-R, for genetic analysis of ASD. Although suggestive evidence of linkage was observed for these total scores, in general, these linkage signals were not as strong as the results from studies using the ASD diagnosis as the primary trait.[18-20,22] One explanation is that the total scores of the separate ADI-R domains as sums of numerous items do not correspond to the true phenotypic constructs of ASD.

The present study investigated the underlying phenotypic constructs of ASD using two large, independent datasets from the Autism Genome Project. The empirically derived factors were then used as quantitative traits for linkage analysis to identify their genetic loci.

Method

Study samples

The Autism Genome Project (AGP) is a consortium of scientists from North America and Europe.[24] Two independent AGP datasets were selected: one included families with two affected relatives (1,236 ASD patients from 618 families, AGP phase I (AGPI)); the other included one ASD subject from each family (804 ASD patients, AGP phase II (AGPII)). For genetic analysis, the parents of the AGPI cases were also included. Details of the sample inclusion criteria can be found in the supplementary data. Informed consent was obtained from all participants in the study and institutional review boards approved our procedures.

ADI-R items and covariates

The ADI-R is a semi-structured interview conducted with the primary caregiver about a child's symptoms both currently and during early development.[25] There are 37 'ever/most abnormal' algorithm items. Two items, 'Friendship at 10-15 years old' and 'Inappropriate facial expressions', were excluded due to high missing rates (because of the age criterion for the former, and for the latter, missing if the item 'Range of facial expressions used to communicate' was coded as 3, i.e. little or no indication of emotion). Seven items which were designed only for verbal individuals were also excluded. The remaining 28 items, available for both verbal and non-verbal individuals, were used in the factor analysis. For each item, the original values 0, 1, 2, and 3 were used with the assumption that 3 was sufficiently different from 2 in our data. Individuals with values 7, 8 and 9 on any item were excluded. All ADI-R assessors were research trained.

Five variables were used as potential covariates for the derived factors: gender, AGP site, age at ADI-R assessment (in months), verbal/non-verbal status, and ethnicity. Ethnicities, defined as Caucasian or non-Caucasian, were estimated using tagSNPs from the Affymetrix 10K array.[22] Covariate effects were tested by including all covariates in a mixed linear model for the related ASD subjects in AGPI.

Exploratory factor analysis

Unweighted least squares factor analysis (ULSFA) was applied to identify the number of factors and pattern of factor loadings in AGPI and AGPII using the SAS FACTOR procedure (v9.1.3, Cary, NC).[26] Factors were retained if 1) a specific aspect of autistic symptoms could be assigned, 2) there was a minimum of three ADI-R items loading on each factor, and 3) the combined variance of all retained factors accounted for as much of the common variance among the 28 ADI-R items as possible so the subsequent genetic analysis could be performed for all different aspects of ASD. Because the ADI-R items are ordinal, polychoric correlations were calculated instead of Pearson correlations. The polychoric correlation matrix was then used in the factor analysis. Orthogonal transformation (varimax rotation) was chosen to maximize independence among factors. The loading threshold was set to 0.35 for factor interpretation.

Since AGPI contains related ASD cases, ULSFA was applied to samples which contained one randomly selected affected individual from each family (n=618). This process was repeated 100 times and the factor loading pattern was summarized. After confirming that the loading pattern from the randomly selected samples was very similar to the factor pattern from AGPI with related ASD individuals, the latter was used to further compare with the loading pattern from AGPII (n=804) using coefficients of congruence.[27]

Genetic markers

The genotypes for AGPI were obtained using Affymetrix (Santa Clara, California) 10K SNP arrays at the Translational Genomics Research Institute.[24] Quality control has been described.[22] A total of 5,371 tagSNPs were selected for linkage analyses so that they were not in strong linkage disequilibrium with each other (maximum D'=0.6).[28] Details of the marker information can be found in the supplementary data.

Genetic analysis

The derived factors from AGPI were used as quantitative traits for heritability estimation using SOLAR (v4.1.0)[29] and for genome-wide multipoint variance components linkage analysis using Merlin --vc option (v1.1.2).[30] Complex ascertainment criteria, as employed in AGPI (families were recruited only if they had ≥2 ASDs), may have a large impact on heritability estimates.[31] No ascertainment correction was performed because normative data

for the ADI-R items from a general population sample does not exist. Due to non-normality of the non-verbal communication factor, rank transformation was performed to achieve normality. Analyses were performed for all 618 families as well as for the subset of 517 Caucasian families. In addition, for each set of families, two models (with and without adjustment for covariates – age at ADI-R assessment, age squared, verbal/non-verbal status, gender) were employed. The significance of the top two linkage results were evaluated using a simulation approach in Merlin.[30] Details of the simulation method can be found in Supplement 1, available online.

In addition to linkage analysis using the factors based on the weights (the coefficients in the linear equation relating factors to the original item values) from AGPI, the linkage analysis was also performed using the factors calculated based on the weights from AGPII. These analyses could provide further evaluation of the similarities of factor analysis results between AGPI and AGPII. If the factor results were truly comparable, then their linkage signals should be similar.

Results

Sample description

Details of the ASD subjects from AGPI and AGPII are in Table 1. The major differences between AGPI and AGPII were the AGP site (e.g. some sites in AGPI were not included in AGPII, and vice versa), and family type (all AGPI samples were from multiplex families while only 28% of the AGPII samples were from multiplex families). For four AGP sites – Autism Genetics Resource Exchange (AGRE), Canadian Autism Genetics (CANAGEN), Collaborative Programs of Excellence in Autism (CPEA), and International Molecular Genetic Study of Autism Consortium (IMGSAC) – which had >100 samples from either AGPI or AGPII, we compared the samples from each AGP site for diagnosis (autism vs. ASD), gender, verbal/non-verbal status, and age at ADI-R assessment. There were significant differences in diagnosis from two sites: IMGSAC had more samples with an ASD diagnosis in AGPII than in AGPI ($p=0.01$), and CANAGEN had more ASD subjects in AGPI than in AGPII ($p=0.001$). All other comparisons were not statistically significant.

Exploratory factor analysis

For both AGPI and AGPII, six factors were retained. According to the common characteristics of the items that were correlated with a particular factor (correlation coefficient $(r)\geq 0.35$), the factors represent themes related to joint attention, social interaction and communication, peer interaction, non-verbal communication, repetitive sensory-motor behaviour, and compulsion/restricted interests. These six factors accounted for most of the common variance among the 28 ADI-R items, with the first factor accounting for >70% of the total common variance and the remaining factors accounting for much lower proportions of the total common variance (3 to 11%) before the varimax rotation (Table 2 and Table S1 (available online)). For comparison to the final factor loading patterns with six factors, Tables S2 and S3, available online, present the loading patterns when two to five factors were retained for AGPI and AGPII, respectively.

Table S4, available online, provides a summary of the factor loading patterns for the 100 randomly selected samples from AGPI. The mean loading values from the 100 samples were very similar to the values from AGPI with the coefficients of congruence >0.999 for all factors. This indicates that the factor analysis results using AGPI with related individuals were very similar to the results from the samples with no related individuals. The factor loading patterns were also quite comparable for AGPI and AGPII with the coefficients of congruence ranging from 0.84 to 0.99 (Table S5, available online). However, the order of

factors in AGPII was different from the order in AGPI. This may reflect statistical variation or sample differences between AGPI and AGPII as described above. For statistical variation, in the factor analyses using 100 random samples from AGPI, we found that even though the proportions of common variance explained by the first two factors were very different before the varimax rotation (70±1% vs. 11±0.6%), they were very similar after the rotation (24.2±5% vs. 23.9±4%). As a result, the order of the first two factors could be different in different samples.

Factors and covariates

All the factors except the non-verbal communication factor from AGPI were normally distributed. The non-verbal communication factor was bimodally distributed with the low-value distribution mainly from the verbal samples and the high-value distribution from both the verbal and non-verbal samples. This factor was rank transformed and the transformed factor was used in the following analyses.

All six covariates (gender, age at ADI-R assessment, age squared, verbal/non-verbal status, AGP site, and estimated ethnicity) were associated with at least three of the six factors ($p<0.05$) (Table S6, available online). Due to the phenotypic as well as possible genetic differences by ethnicity (e.g. different allele frequencies in samples from different ancestral backgrounds), the Caucasian families from AGPI were used after the initial analyses using all the AGPI families. Gender, age at ADI-R assessment, age squared, and verbal/non-verbal status were selected as covariates in genetic analyses. AGP site was not included because we suspected that the differences in the factor scores across sites might be caused by true differences in severity among individuals rather than by measurement error, since the ADI-R was administrated by trained clinicians or assessors who had demonstrated >80% reliability compared to the trainers across all scoring items. In addition, our previous study[22] has shown that the effect of the AGP site as a covariate on the linkage results of domain total scores as quantitative traits were relatively small, with genome-wide changes in LOD scores ≤0.5.

Genetic analysis

All the factors showed significant familial aggregation (Table 2 and Table S7 (available online). Most of the heritability estimates ranged from 0.46 to 0.70 (depending on whether all or Caucasian families were used and whether covariates were included), while the peer interaction factor had a heritability estimate of 0.27-0.33 possibly due to the effect of contextual opportunities to observe and learn on the items of this factor within a social interaction setting. Compared to twin studies, heritability estimates from siblings cannot separate shared genetic from shared environmental influences, therefore these results should be interpreted with caution.

Figure 1 illustrates the genome-wide linkage analysis results for the factors (chromosome-specific results are in Figure S1, available online). Two chromosomal regions, 11q23.1-q23.3 for the joint attention factor (LOD score=4.0) and 19q13.32-q13.33 for the repetitive sensory-motor behaviour factor (LOD score=4.92), presented strong evidence for linkage when Caucasian families were used and when covariate effects were adjusted (Figure 2, Table 3).

Of the 1,000 simulated genome-wide datasets, 76 had a LOD score ≥4.0 for the joint attention factor, indicating that the linkage result at the 11q23.1-q23.3 region clearly met the genome-wide suggestive linkage criterion but was just short of ‘significant’. Only 7 simulations had a LOD score ≥4.92 for the repetitive sensory-motor behaviour factor

demonstrating that the linkage result at the 19q13.32-q13.33 region was genome-wide significant.

For the linkage analyses using the factors based on the weights from AGPII, the overall linkage results were similar to the results from AGPI. At the two highlighted regions, the peak LOD scores were 3.54 (p=0.00003) at 19q13.32-q13.33 for the repetitive sensory-motor behaviour factor, and 3.27 (p=0.00005) at 11q23.1-q23.3 for the joint attention factor. This demonstrates that the linkage results were reasonably consistent when the factors were generated from independent ASD datasets. This also indirectly testifies that the factor analysis results from AGPI and AGPII were similar.

Discussion

This is the first study which applied empirically derived factors as quantitative traits in genetic analysis of ASD. The factor loading patterns of the ADI-R algorithm items were replicated using two independent AGP datasets. In contrast to conventional factor analysis, which retains only the top factors that account for a large proportion of the total common variance, we retained factors which explained as little as 3% of the total common variance. As a result, most of the common variance among the 28 ADI-R algorithm items was accounted for in this study.

Two previous factor analysis studies employed the ADI-R algorithm items to determine the factor structure of ASD.[6,7] Of them, the study by Snow et al. is similar to our study (i.e. it used a relatively large sample size, polychoric correlations, and the unweighted least squares factor analysis method). In addition, some of the cases from the Snow et al. study likely overlap with those from our AGRE sample (Table 1: AGRE samples made up 34% of AGPI and 6% of AGPII). They found a two-domain model with a combined social-communication factor, and a restricted and repetitive behaviour factor. The major difference between these two studies is the rotation method: the quartimin rotation (a type of oblique rotation) in Snow et al.; the orthogonal rotation in our study. When the oblique rotation was applied to our data, the factor loading pattern was similar to the pattern using the orthogonal rotation (data not shown). However, since the oblique rotation allows factors to be correlated with each other, all factors, except the compulsion/restricted interests factor, were highly correlated with the correlation coefficients ranging from 0.23 to 0.57 in AGPII. We decided to apply the orthogonal rotation so that the factors would not be highly correlated and unique genetic loci could be identified for each factor. Another disparity between these two studies is that verbal and non-verbal individuals were analyzed separately in Snow et al. while we analyzed them together but adjusted for verbal/non-verbal status as a covariate for all the factors prior to linkage analysis. Despite these differences, both studies found that the items from the reciprocal social interaction and communication domains tended to load on the same factor/factors, i.e. the social-communication factor in Snow et al. and the joint attention and social interaction and communication factors in this study.

For the ADI-R items from the restricted, repetitive, and stereotyped behaviour domain, many factor analysis studies have consistently found two factors: repetitive sensory-motor actions (RMSA) or lower-order repetitive behaviours, and insistence on sameness (IS) or higher-order repetitive behaviours.[13,32] There were also two factors for this ADI-R domain in our study (Table 2). Our repetitive sensory-motor behaviour factor was similar to the RSMA factor in the previous studies. However, the compulsion/restricted interests factor was different from the IS factor. It was also different from the third factor (circumscribed interests) in Lam et al.[33] This might be due to the fact that non-algorithm items were used in previous studies while we only included the algorithm items for which we had the most complete data. Using AGPII, we found that the compulsion/restricted interests factor was

positively correlated with verbal IQ after adjusting for age at ADI-R assessment, verbal/nonverbal status, and gender (p=0.003). In contrast, the repetitive sensory-motor behaviour factor was negatively correlated with verbal (p=0.01) and performance IQ (p=0.05). These results demonstrate that these two factors are distinct from each other and correspond to the lower/higher-order repetitive behaviours defined in previous studies.

All the empirically derived factors in this study have shown familial aggregation. However, the proportion of the total common variance that a factor could explain was not a reliable predictor of heritability estimates. For example, the compulsion/restricted interests factor accounted for the smallest proportion of the total common variance (3%) but had the highest heritability (65%) (Table 2). In addition, as has been reported before,[34] heritability estimates are also not reliable predictors of the results of linkage analysis, e.g. the most significant linkage signal in this study was found for the repetitive sensory-motor behaviour factor (with a heritability estimate of 0.54) rather than the compulsion/restricted interests factor which had the highest heritability estimate (0.65).

Two regions with strong evidence of linkage were highlighted in this study: (1) at 11q23.1-q23.3 for the joint attention factor with the 1-LOD score range (peak LOD score minus 1) of 5.8Mb (from 110.9 to 116.7Mb based on NCBI build 35) containing 83 genes including neural cell adhesion molecule 1 gene (NCAM1), dopamine receptor D2 (DRD2), and 5-hydroxytryptamine receptor 3A and 3B (HTR3A and HTR3B); and (2) at 19q13.32-q13.33 for the repetitive sensory-motor behaviour factor with the 1-LOD score range of 3.3Mb (from 51.3 to 54.6Mb) containing 122 genes including the solute carrier family 8 (sodium-calcium exchanger), member 2 gene (SLC8A2). The candidate genes at these two regions have been associated with ASD and other psychiatric and neurodevelopmental diseases.[35-40] However, neither of these two regions overlaps with the apparent linkage regions in previous AGP linkage studies.[22,24,41] This is not surprising since two of these studies used the ASD diagnosis as a primary outcome[24,41] and the third study used either subsets of families or ADI-R domain total scores as primary traits.[22] These two regions also did not overlap with the reported linkage regions from other linkage studies of ASD.[42,43] On the other hand, in a meta-analysis of 20 linkage studies for schizophrenia,[44] the 11q22.3-q24.1 region was ranked as the 4th most significant linkage region. The 19q13.32-q13.33 region (51.3-54.6Mb) in our study also overlaps with a linkage region for schizophrenia using cases with positive family history from Aberdeen, Scotland.[45] Recent studies have reported common genes involved in both ASD and schizophrenia.[46,47] If our linkage analysis results at these loci are replicated, further study will be needed to determine if the derived factors and schizophrenia share common variants at these regions.

There are several limitations to this study. First, because of the lack of IQ measures in AGPI, we were not able to use IQ as a covariate in the model. Second, for the factor analysis, only the ADI-R algorithm items were used instead of all the items. Third, the AGPI families were used in several linkage studies before.[22,24,41] Because the primary traits and analysis methods used in the previous studies were very different from those applied in this study, no multiple testing correction was made. However, the analyses presented here do represent secondary analyses and need to be seen in that context. Finally, the linkage analyses in this study involve six quantitative traits, two sets of families (all vs. Caucasian families), and two statistical models (with and without covariates). It has been shown that ‘maximizing’ the LOD score over different model parameter values inflates the LOD scores.[48] When all 6 factors were considered with a total of 6,000 simulated genome-wide scans, 221 had a LOD score ≥4.0 and 37 had a LOD score ≥4.92. Therefore, the p value for the linkage locus at 11q23 for the joint attention factor was 0.04, and was 0.006 for the repetitive sensory-motor factor at 19q13. Because of the overlap between the two sets of families and the similarity of the two statistical models, the final number of tests was

equivalent to 1.53 independent tests.[49] As a result, the final significant level was 0.06 for the linkage signal at 11q23 for the joint attention factor, and was 0.009 at 19q13 for the repetitive sensory-motor factor which was genome-wide significant.

In addition to their application in genetic studies of ASD, the empirical factors in this study could also be useful in clinical practice. Of the six factors, peer-interaction, non-verbal communication, and compulsion/restricted interests were highly correlated ($r \geq 0.35$) with ADI-R items from 'reciprocal social interaction', 'communication', and 'repetitive and stereotyped behaviour' domains, respectively; while the remaining three factors were highly correlated with the items from more than one of the three domains. More studies are needed to determine if these factors correspond with the true areas of deficits in the ASD cases and if the factor scores can be used as a 'proxy' of severity in ASD.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors gratefully acknowledge the families participating in this study, the Autism Genome Project (AGP) Consortium, and the main funders of AGP: Autism Speaks (USA), the Health Research Board (HRB; Ireland), the Medical Research Council (MRC; UK), Genome Canada/Ontario Genomics Institute, and the Hilibrand Foundation (USA). Additional support for the authors was provided by the US National Institutes of Health (HD055782, HD055751), the Canada Research Chair in Genetics of Complex Diseases, and the Canadian Institutes for Health Research (CIHR).

References

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. American Psychiatric Press; Washington DC, USA: 1994.
2. World Health Organization. International statistical classification of diseases and related health problems. World Health Organization; Geneva, Switzerland: 1992.
3. Szatmari P, Georgiades S, Bryson S, et al. Investigating the structure of the restricted, repetitive behaviours and interests domain of autism. J Child Psychol Psychiatry. 2006; 47:582–90. [PubMed: 16712635]
4. Constantino JN, Gruber CP, Davis S, Hayes S, Passanante N, Przybeck T. The factor structure of autistic traits. J Child Psychol Psychiatry. 2004; 45:719–26. [PubMed: 15056304]
5. Kamp-Becker I, Ghahreman M, Smidt J, Remschmidt H. Dimensional structure of the autism phenotype: relations between early development and current presentation. J Autism Dev Disord. 2009; 39:557–71. [PubMed: 18941880]
6. Lecavalier L, Aman MG, Scahill L, et al. Validity of the autism diagnostic interview-revised. Am J Ment Retard. 2006; 111:199–215. [PubMed: 16597187]
7. Snow AV, Lecavalier L, Houts C. The structure of the Autism Diagnostic Interview-Revised: diagnostic and phenotypic implications. J Child Psychol Psychiatry. 2009; 50:734–42. [PubMed: 19207624]
8. Tadevosyan-Leyfer O, Dowd M, Mankoski R, et al. A principal components analysis of the Autism Diagnostic Interview-Revised. J Am Acad Child Adolesc Psychiatry. 2003; 42:864–72. [PubMed: 12819447]
9. Boomsma A, Van Lang ND, De Jonge MV, De Bildt AA, Van Engeland H, Minderaa RB. A new symptom model for autism cross-validated in an independent sample. J Child Psychol Psychiatry. 2008; 49:809–16. [PubMed: 18492042]
10. Frazier TW, Youngstrom EA, Kubu CS, Sinclair L, Rezai A. Exploratory and confirmatory factor analysis of the autism diagnostic interview-revised. J Autism Dev Disord. 2008; 38:474–80. [PubMed: 17619129]

11. Georgiades S, Szatmari P, Zwaigenbaum L, et al. Structure of the autism symptom phenotype: A proposed multidimensional model. J Am Acad Child Adolesc Psychiatry. 2007; 46:188–96. [PubMed: 17242622]
12. Eaves RC, Williams TO Jr. Exploratory and confirmatory factor analyses of the pervasive developmental disorders rating scale for young children with autistic disorder. J Genet Psychol. 2006; 167:65–92. [PubMed: 16629404]
13. Hus V, Pickles A, Cook EH Jr. Risi S, Lord C. Using the autism diagnostic interview--revised to increase phenotypic homogeneity in genetic studies of autism. Biol Psychiatry. 2007; 61:438–48. [PubMed: 17276746]
14. Cannon DS, Miller JS, Robison RJ, et al. Genome-wide linkage analyses of two repetitive behavior phenotypes in Utah pedigrees with autism spectrum disorders. Mol Autism. 2010; 1:3. [PubMed: 20678246]
15. Ma DQ, Jaworski J, Menold MM, et al. Ordered-subset analysis of savant skills in autism for 15q11-q13. Am J Med Genet B Neuropsychiatr Genet. 2005; 135B:38–41. [PubMed: 15756693]
16. Nurmi EL, Dowd M, Tadevosyan-Leyfer O, Haines JL, Folstein SE, Sutcliffe JS. Exploratory subsetting of autism families based on savant skills improves evidence of genetic linkage to 15q11-q13. J Am Acad Child Adolesc Psychiatry. 2003; 42:856–63. [PubMed: 12819446]
17. Shao Y, Cuccaro ML, Hauser ER, et al. Fine mapping of autistic disorder to chromosome 15q11-q13 by use of phenotypic subtypes. Am J Hum Genet. 2003; 72:539–48. [PubMed: 12567325]
18. Alarcon M, Cantor RM, Liu J, Gilliam TC, Geschwind DH. Evidence for a language quantitative trait locus on chromosome 7q in multiplex autism families. Am J Hum Genet. 2002; 70:60–71. [PubMed: 11741194]
19. Alarcon M, Yonan AL, Gilliam TC, Cantor RM, Geschwind DH. Quantitative genome scan and Ordered-Subsets Analysis of autism endophenotypes support language QTLs. Mol Psychiatry. 2005; 10:747–57. [PubMed: 15824743]
20. Chen GK, Kono N, Geschwind DH, Cantor RM. Quantitative trait locus analysis of nonverbal communication in autism spectrum disorder. Mol Psychiatry. 2006; 11:214–20. [PubMed: 16189504]
21. Duvall JA, Lu A, Cantor RM, Todd RD, Constantino JN, Geschwind DH. A quantitative trait locus analysis of social responsiveness in multiplex autism families. Am J Psychiatry. 2007; 164:656–62. [PubMed: 17403980]
22. Liu XQ, Paterson AD, Szatmari P, The Autism Genome Project Consortium. Genome-wide linkage analyses of quantitative and categorical autism subphenotypes. Biol Psychiatry. 2008; 64:561–70. [PubMed: 18632090]
23. Schellenberg GD, Dawson G, Sung YJ, et al. Evidence for multiple loci from a genome scan of autism kindreds. Mol Psychiatry. 2006; 11:1049–60. [PubMed: 16880825]
24. AGP. Mapping autism risk loci using genetic linkage and chromosomal rearrangements. Nat Genet. 2007; 39:319–28. [PubMed: 17322880]
25. Rutter, M.; Le Couteur, A.; Lord, C. Autism Diagnostic Interview-Revised. Western Psychological Services; Los Angeles, USA: 2003.
26. Hatcher, L. A step-by-step approach to using the SAS system for factor analysis and structural equation modeling. SAS Institute Inc; Cary, NC, USA: 1994.
27. Levine, MS. Canonical analysis and factor comparison. Sage Publications; Thousand Oaks, CA, USA: 1977.
28. Huang Q, Shete S, Amos CI. Ignoring linkage disequilibrium among tightly linked markers induces false-positive evidence of linkage for affected sib pair analysis. Am J Hum Genet. 2004; 75:1106–12. [PubMed: 15492927]
29. Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. Am J Hum Genet. 1998; 62:1198–211. [PubMed: 9545414]
30. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet. 2002; 30:97–101. [PubMed: 11731797]
31. Rao DC, Wette R, Ewens WJ. Multifactorial analysis of family data ascertained through truncation: a comparative evaluation of two methods of statistical inference. Am J Hum Genet. 1988; 42:506–15. [PubMed: 3348215]

32. Turner M. Annotation: Repetitive behaviour in autism: a review of psychological research. J Child Psychol Psychiatry. 1999; 40:839–49. [PubMed: 10509879]
33. Lam KS, Bodfish JW, Piven J. Evidence for three subtypes of repetitive behavior in autism that differ in familiality and association with other symptoms. J Child Psychol Psychiatry. 2008; 49:1193–200. [PubMed: 19017031]
34. Wijsman EM, Sung YJ, Buil A, et al. Summary of Genetic Analysis Workshop 15: Group 9 linkage analysis of the CEPH expression data. Genet Epidemiol. 2007; 31(Suppl 1):S75–85. [PubMed: 18046772]
35. Anderson BM, Schnetz-Boutaud NC, Bartlett J, et al. Examination of association of genes in the serotonin system to autism. Neurogenetics. 2009; 10:209–16. [PubMed: 19184136]
36. Glatt SJ, Jonsson EG. The Cys allele of the DRD2 Ser311Cys polymorphism has a dominant effect on risk for schizophrenia: evidence from fixed- and random-effects meta-analyses. Am J Med Genet B Neuropsychiatr Genet. 2006; 141B:149–54. [PubMed: 16402354]
37. Atz ME, Rollins B, Vawter MP. NCAM1 association study of bipolar disorder and schizophrenia: polymorphisms and alternatively spliced isoforms lead to similarities and differences. Psychiatr Genet. 2007; 17:55–67. [PubMed: 17413444]
38. Jeon D, Yang YM, Jeong MJ, Philipson KD, Rhim H, Shin HS. Enhanced learning and memory in mice lacking Na+/Ca2+ exchanger 2. Neuron. 2003; 38:965–76. [PubMed: 12818181]
39. O'Rourke JA, Scharf JM, Yu D, Pauls DL. The genetics of Tourette syndrome: a review. J Psychosom Res. 2009; 67:533–45. [PubMed: 19913658]
40. Sharp SI, McQuillin A, Gurling HM. Genetics of attention-deficit hyperactivity disorder (ADHD). Neuropharmacology. 2009; 57:590–600. [PubMed: 19715710]
41. Vieland VJ, Hallmeyer J, Huang Y, et al. Novel method for combined linkage and genomewide association analysis finds evidence of distinct genetic architecture for two subtypes of autism. J Neurodev Disord. (in press).
42. Freitag CM, Staal W, Klauck SM, Duketis E, Waltes R. Genetics of autistic disorders: review and clinical implications. Eur Child Adolesc Psychiatry. 2010; 19:169–78. [PubMed: 19941018]
43. Trikalinos TA, Karvouni A, Zintzaras E, et al. A heterogeneity-based genome search meta-analysis for autism-spectrum disorders. Mol Psychiatry. 2006; 11:29–36. [PubMed: 16189507]
44. Lewis CM, Levinson DF, Wise LH, et al. Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: Schizophrenia. Am J Hum Genet. 2003; 73:34–48. [PubMed: 12802786]
45. Francks C, Tozzi F, Farmer A, et al. Population-based linkage analysis of schizophrenia and bipolar case-control cohorts identifies a potential susceptibility locus on 19q13. Mol Psychiatry. 2010; 15:319–25. [PubMed: 18794890]
46. Kilpinen H, Ylisaukko-Oja T, Hennah W, et al. Association of DISC1 with autism and Asperger syndrome. Mol Psychiatry. 2008; 13:187–196. [PubMed: 17579608]
47. Kirov G, Gumus D, Chen W, et al. Comparative genome hybridization suggests a role for NRXN1 and APBA2 in schizophrenia. Hum Mol Genet. 2008; 17:458–65. [PubMed: 17989066]
48. Weeks DE, Lehner T, Squires-Wheeler E, Kaufmann C, Ott J. Measuring the inflation of the lod score due to its maximization over model parameter values in human linkage analysis. Genet Epidemiol. 1990; 7:237–43. [PubMed: 2227370]
49. Camp NJ, Farnham JM. Correcting for multiple analyses in genomewide linkage studies. Ann Hum Genet. 2001; 65:577–82. [PubMed: 11851987]

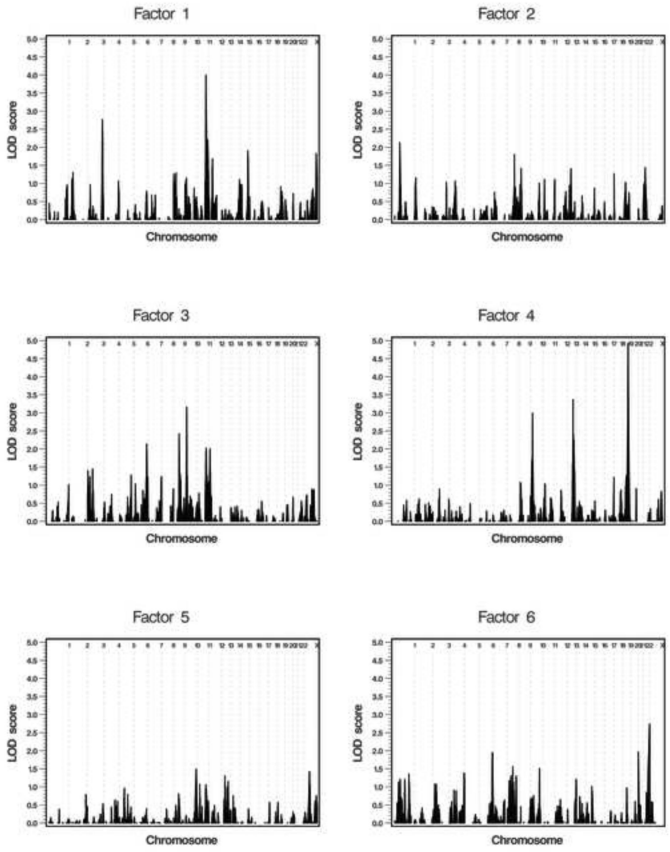

Figure 1.

Genome-wide linkage analysis results for the derived factors after adjustment for covariates using the Caucasian families. Note: Factor 1: joint attention; factor 2: social interaction and communication; factor 3: non-verbal communication (rank transformed); factor 4: repetitive sensory-motor behaviour; factor 5: peer interaction; and factor 6: compulsion/restricted interests. The vertical dashed lines separate the chromosomes. LOD = logarithm of odds.

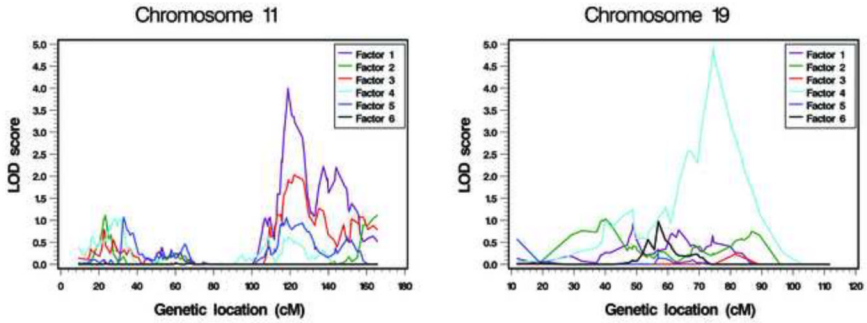

Figure 2.

Highlighted linkage analysis results for the derived factors on chromosomes 11 and 19. Note: The factors were adjusted for covariates and the Caucasian families were used. Factor 1: joint attention; factor 2: social interaction and communication; factor 3: non-verbal communication (rank transformed); factor 4: repetitive sensory-motor behaviour; factor 5: peer interaction; and factor 6: compulsion/restricted interests. LOD = logarithm of odds.

Table 1

Sample characteristics by dataset[a]

	AGPI (n=1,236)	AGPII (n=804)
Diagnosis		
Autism	1,134 (91.7)	760 (94.5)
ASD	102 (8.3)	44 (5.5)
Gender		
Male	990 (80.1)	677 (84.2)
Female	246 (19.9)	127 (15.8)
Verbal/nonverbal status		
Verbal	898 (72.6)	553 (68.8)
Nonverbal	338 (27.4)	251 (31.2)
AGP site		
AGRE	420 (34.0)	46 (5.7)
VANDERBILT	16 (1.3)	22 (2.7)
IMGSAC	330 (26.7)	61 (7.6)
DUKE	50 (4.0)	72 (9.0)
CANAGEN	80 (6.5)	190 (23.6)
INSERM	22 (1.8)	78 (9.7)
STANFORD	64 (5.2)	0 (0)
CPEA	182 (14.7)	22 (2.7)
UNC	58 (4.7)	0 (0)
MT. SINAI	14 (1.1)	0 (0)
IRELAND	0 (0)	113 (14.1)
PORTUGAL	0 (0)	200 (24.9)
Family type[b]		
Simplex	0 (0)	436 (54.2)
Multiplex	1,236 (100)	223 (27.8)
Unknown	0 (0)	145 (18.0)
Ethnicity		
Caucasian	1,036 (83.8)	716 (89.1)
Non-Caucasian	200 (16.2)	88 (10.9)
Age at ADI-R assessment (month)	101±39	96±44

Note: ADI-R = Autism Diagnostic Interview-Revised; AGP = Autism Genome Project; AGPI = patients from AGP phase I; AGPII = patients from AGP phase II; AGRE = Autism Genetics Resource Exchange; CANAGEN = Canadian Autism Genetics; CPEA = Collaborative Programs of Excellence in Autism; IMGSAC = International Molecular Genetic Study of Autism Consortium; INSERM = Institute National de la Santé et de la Recherche Médicale; UNC, University of North Carolina.

[a] Values are count (percentage) or mean ± standard deviation.

[b]For family type, multiplex Autism Genome Project (AGP) families were defined as having at least two individuals receiving autism spectrum disorder (ASD) diagnoses who were first to third degree relatives (for third degree, only considering cousins); simplex families as having only one known ASD individual with no family history of ASD in first to third (cousin) degree relatives.

Table 2

Factor loading patterns (correlations between the items from the Autism Diagnostic Interview-Revised (ADI-R) and the derived factors) for the patients from the Autism Genome Project phase I

	ADI-R items	Factor1: Joint attention	Factor2: Social interaction & communication	Factor3: Non-verbal communication	Factor4: Repetitive sensory-motor behaviour	Factor5: Peer interaction	Factor6: Compulsion/ restricted interests
Social domain	Direct gaze	**0.55**	0.11	0.05	0.20	0.17	0.23
	Social smiling	**0.65**	0.26	0.17	0.10	0.18	0.18
	Range of facial expressions used to communicate	**0.53**	0.22	0.15	0.12	0.13	0.21
	Imaginative play with peers	0.24	**0.66**	0.12	0.15	0.32	−0.09
	Interest in children	**0.36**	0.27	0.09	0.12	**0.65**	0.08
	Response to approaches of other children	0.33	0.23	0.12	0.16	**0.65**	0.10
	Group play with peers	0.23	**0.42**	0.15	0.03	**0.52**	0.06
	Showing and directing attention	**0.45**	**0.46**	0.25	0.20	0.16	−0.06
	Offering to share	0.31	**0.47**	0.23	0.18	0.19	−0.04
	Seeking to share enjoyment with others	**0.48**	0.31	0.26	0.14	0.22	0.01
	Use of other's body to communicate	0.12	0.17	0.19	**0.44**	0.13	−0.14
	Offering comfort	**0.35**	**0.47**	0.15	0.16	0.15	0.00
	Quality of social overtures	**0.61**	0.30	0.16	0.17	0.18	0.04
	Appropriateness of social responses	**0.56**	0.31	0.06	0.15	0.20	0.02
Communication	Pointing to express interest	**0.35**	0.26	0.32	0.27	0.12	−0.02
	Nodding	0.19	0.12	**0.94**	0.24	0.14	0.00
	Head shaking	0.22	0.21	**0.81**	0.15	0.06	0.06
	Conventional/instrumental gestures	**0.39**	**0.46**	**0.40**	0.15	0.19	0.06
	Spontaneous imitation of actions	0.25	**0.59**	0.09	0.17	0.06	0.07
	Imaginative play	0.20	**0.69**	0.10	0.23	0.18	0.05
	Imitative social play	0.33	**0.43**	0.03	0.15	0.32	0.10
Behaviour	Unusual preoccupations	0.03	0.10	0.13	0.15	0.04	**0.40**
	Circumscribed interests	0.09	−0.04	−0.10	−0.08	0.03	**0.52**
	Compulsions/rituals	0.09	−0.02	0.02	0.16	0.02	**0.40**
	Hand and finger mannerisms	0.15	0.10	0.09	**0.55**	0.09	0.10
	Other complex mannerisms or stereotyped body movements	0.10	0.11	0.07	**0.48**	0.04	0.12

	ADI-R items	Factor1: Joint attention	Factor2: Social interaction & communication	Factor3: Non-verbal communication	Factor4: Repetitive sensory-motor behaviour	Factor5: Peer interaction	Factor6: Compulsion/ restricted interests
	Repetitive use of objects or interest in parts of objects	0.07	0.34	0.08	**0.56**	0.04	0.16
	Unusual sensory interests	0.11	0.05	0.08	**0.60**	0.03	0.07
Common variance explained (%) – before rotation		74.0	10.9	8.3	7.0	3.8	3.4
Common variance explained (%) – after rotation		24.7	24.6	16.6	15.0	12.8	6.3
Heritability estimate[a]		0.50	0.49	0.47	0.54	0.29	0.65

Note: Bold if loading ≥0.35.

[a]Heritability estimates from the Caucasian families and after adjustment for four covariates (details in Table S7, available online).

Table 3

Highlighted linkage analysis results (logarithm of odds (LOD) scores with nominal p values)[a]

Factor	Peak SNP	All (618 families)		Caucasian (517 families)	
		No covariate	4 covariates[c]	No covariate	4 covariates[c]
Factor 1: Joint attention	Chr11q23: rs723599 (112,377,515 bp)[b]	2.93 (p=0.0001)	3.92 (p=0.00001)	3.47 (p=0.00003)	4.00 (p=0.00001)
Factor 4: Repetitive sensory-motor behaviour	Chr19q13: rs895355 (52,822,703 bp)[b]	0.98 (p=0.02)	1.87 (p=0.002)	2.66 (p=0.0002)	4.92 (p<0.00001)

Note: The highest LOD scores are in bold. Bp = base pair; SNP = single nucleotide polymorphism.

[a] For both regions, the linkage analysis results improved after adjustment for covariates (e.g. from 3.47 to 4.00 for the joint attention factor, and from 2.66 to 4.92 for the repetitive sensory-motor behaviour factor using the Caucasian families). Restricting the analysis to the Caucasian families resulted in stronger evidence for linkage at 19q13.32-q13.33 (from 1.87 to 4.92 with the adjustment for covariates), while the results at 11q23.1-q23.3 changed little (from 3.92 to 4.00).

[b] National Center for Biotechnology Information (NCBI) build 35.

[c] After adjustment for 4 covariates: gender, age at Autism Diagnostic Interview-Revised (ADI-R) assessment, age squared, and verbal/nonverbal status.