Title Page

Gene ontology enrichment analysis in two independent family-based samples highlights biologically plausible processes for autism spectrum disorders

Running Title

Gene pathways analysis in ASD

Word Count

158 (Abstract)

3491 (Manuscript Body)

Author List

Richard J.L. Anney [1]†PhD, Elizabeth A. Heron [1] PhD, Ricardo Segurado [1] PhD, Elaine M. Kenny [1] PhD, Colm O'Dushlaine [1,2] PhD, Brian L. Yaspan[3] PhD, Elena Parkhomenka[4] PhD, **The Autism Genome Project**‡, Joseph D. Buxbaum[4] PhD, James Sutcliffe[3] PhD, Michael Gill MRCPsych MD [1], Louise Gallagher MRCPsych MD [1].

‡Full list of Autism Genome Project Members is given below

Corresponding Author

Richard JL Anney

Email: anneyr@tcd.ie

Postal Address: Autism Genetics Group, Department of Psychiatry, Institute of Molecular Medicine

Trinity Centre for Health Sciences, St. James' Hospital, James' Street, Dublin 8, Ireland

Tel: ++353 1 896-8461

Fax: ++353 1 896-3405

Abstract

Recent genome-wide association studies (GWAS) have implicated a range of genes from discrete biological pathways in the aetiology of autism. However, despite the strong influence of genetic factors, association studies have yet to identify statistically robust, replicated major effect genes or SNPs. We apply the principle of the SNP ratio test methodology described by O'Dushlaine et al., (2009) to over 2100 families from the Autism Genome Project (AGP). Using a two-stage design we examine association enrichment in 5955 unique gene-ontology classifications across four groupings based on two phenotypic and two ancestral classifications. Based on estimates from simulation we identify excess of association enrichment across all analyses. We observe enrichment in association for sets of genes involved in diverse biological processes including pyruvate metabolism, transcription factor activation, cell-signalling and cell-cycle regulation. Both genes and processes that show enrichment have previously been examined in autistic disorders and offer biologically plausibility to these findings.

Keywords

autism, genome-wide association analysis, pathway analysis, family-based association test, gene-ontology

Introduction

Autism is a complex neurodevelopmental disorder characterized by impairments of varying severity in the three core areas of communication, social interaction and repetitive behaviour. Population prevalence of autism is approximately 15-20 per 10 000 with all autism spectrum disorders (ASD) estimated at 60 in 10 000 children [1,2]. The role of genetic factors in the development of autism is undisputed. Heritability has been estimated as high as 91-93% using a multi-threshold liability model [3]. However, despite the strong influence of genetic factors, autism linkage studies and association studies of common SNPs have not identified any genes of major effect. Recent genome-wide association studies (GWAS), have implicated a number of genes from discrete biological pathways in the aetiology of autism [4-6]. In a recent study by the AGP using these data, we identified genome-wide significant association with *MACROD2* [7]. However, we did not observe strong marker-wise associations within the cadherin gene region (*CDH9*, *CDH10*) or the *TAS2R1*, *SEMA5A* region that were highlighted in the work of Wang and colleagues [4], Ma and colleagues [5] and Weiss and colleagues [6]. In addition to identifying genome-wide significant association it can be hypothesised that additional true vulnerability loci may exist within the nominal to modest range of statistical significance and confer risk to the disorder [8]. A milieu of nominal to modestly associated risk variation fits with a polygenic model of disease and presents additional challenges for the identification of patterns of association within expected experimental noise [9].

One promising approach is to examine association enrichment within "pathways" or groups of genes. The underlying hypothesis of association enrichment analysis is that functional polymorphisms that exist within a group of biologically interrelated genes are in essence

"disrupting" the normal functioning of the biological process of the pathway. Consequently, one can consider the biological process, rather than the individual gene or SNP, in the development of the disease/disorder. By examining the ratio of association signals within a group of genes we can determine whether there is enrichment of the signal above that expected by chance. This strategy also decreases the multiple-testing burden that accompanies GWAS, and can have increased power.

A number of pathway-based methodologies have been developed to examine gene enrichment in association data (reviewed in [10]). These include gene ranking algorithms [11], gene-enrichment algorithms e.g. ALIGATOR (Association LIst Go AnnoTatOR) [9] and SNP-enrichment approaches such as the SRT (SNP ratio test) [12]. The SRT provides a formal test of whether markers within pre-defined pathways show enrichment in association signal over that expected by chance alone. For case-control data, the basic algorithm underpinning the SRT is to first calculate the ratio of the number of nominally associated SNP markers within a pathway to the total number of markers within the pathway. Significance is assigned through a case-randomisation permutation routine, which takes accounts of the linkage disequilibrium between markers.

To apply the SRT to family-based data we are unable to perform standard case-randomisation therefore a pseudo-sibling model is generated from the alleles that are not transmitted to the proband. A proband-randomisation procedure is performed within the family, whereby the affection status of the offspring (case and pseudo-sibling) is permuted. This method allows retention of the linkage disequilibrium structure within the families and retains the advantages of the Transmission Disequilibrium Test design for the family-based association. In the current study, we chose the SRT over other approaches for a number of

reasons. Firstly, as the SRT retains all of the markers from the association analysis it is sensitive to more than one true association signal per gene and therefore gains information in the presence of allelic heterogeneity. Secondly, the SRT's use of multiple association signals across a gene as opposed to a single maximum signal limits potential genotyping artefact effects. Genotyping error at a single point may highlight a gene erroneously in a maximum signal design where this becomes the only observation. However, taking the ratio of all signals across a gene restricts the impact of single points of error as they are more likely to be diluted across the gene. Thirdly, the SRT also controls for gene size and linkage disequilibrium effects by permuting case-ness independently of genotype, consequently maintaining the same recombination patterns. Approaches that do not apply a gene-wise correction to GWAS data can show inflated signals for pathways containing larger genes. This is often the case in brain expressed pathways that are enriched for larger genes such as cell-surface receptors and can lead to misinterpretation of any association enrichment. Finally, as the SRT uses a SNP-wise association statistic over a gene-wise association statistic, we have sufficient observations to examine pathways which may contain fewer genes. Thereby we are able to examine discrete "niche" pathways as well as larger, more diverse gene-sets for enrichment in the GWAS.

For this study we use gene-set lists derived from the gene-ontology (www.geneontology.org) database to examine whether association enrichment is present in a cohort of individuals from the Autism Genome Project (AGP) with a diagnosis of autistic disorder.

Materials and Methods

Subjects

The individuals examined in this study were collected as part of the Autism Genome

Project (AGP) Consortium genome analysis project. The AGP represents more than 50 centres in North America and Europe. Subjects with known karyotypic abnormalities, fragile X mutations or other known genetic disorders were excluded. Diagnostic and ancestral definitions were as previously reported by this group [7]. Briefly, families are grouped into two nested diagnostic classes (*Strict*, and *Spectrum*) based on proband diagnostic measures. To qualify for the *Strict* class, affected individuals met criteria for autism on both primary diagnostic instruments; the Autism Diagnostic Interview-Revised (ADI-R [13]) and Autism Diagnostic Observation Schedule (ADOS [14]). ADI-R-based diagnostic classification of subjects as ASD followed criteria published by Risi and colleagues [15]. Specifically, individuals who almost met ADI criteria for autism were classified as ASD if; (1) they met criteria on social and either communication or repetitive behaviour domains; or (2) met criteria on the social domain and were within 2 points of criteria for communication, or met criteria on the communication domain and were within 2 points of social criteria, or within 1 point on both social and communication domains. The *Spectrum* class included all individuals who met *Strict* criteria and those individuals who were classified as ASD or autism on both the ADI-R and ADOS or who were not evaluated on one of the instruments but were diagnosed with autism on the other instrument. A summary of the sample sizes for the Discovery and Replication datasets for each Diagnostic/Ancestry subset is shown in TABLE 1.

As described elsewhere [7], ancestry for these individuals was determined for the proband by using 5,239 widely-spaced, independent SNPs that had a genotype completion rate of ≥99.9%. The software used was Spectral-GEM [16], which estimated 5 significant dimensions of ancestry. Subsequent clustering on dimensions of ancestry identified 9 clusters; 5

clusters were used to describe European ancestry and the remaining clusters best reflect Asian, African (East/West) and Latin American origins. The *All ancestry* class included all individuals including those who met the *European ancestry* criteria.

Genotyping and Association Analysis (Transmission Disequilibrium Test)

The discovery sample were genotyped using the Illumina Infinium 1M-single SNP microarray, the replication sample were genotyped on a either the Illumina Infinium 1M-single SNP microarray as well as the Illumina 1M-duo microarray. All quality control (QC) procedures were maintained across datasets; in addition QC marker sets from both the discovery and replication datasets were matched and only those markers meeting QC for both the discovery and replication datasets were carried forward to analysis. Additional QC details are described elsewhere [7]. A total of 856932 SNPs passed QC on both the discovery and replication sample. Transmissions disequilibrium test statistics were generated using PLINK v1.07 [17].

Pedigree SNP ratio Test (pedSRT)

The pedSRT is a modification to the SRT described by O'Dushlaine and colleagues [12] which is applicable to family-based data. Briefly, the SRT tests the ratio of the number of associated SNPs to the total number of SNPs in a pre-defined set of genes. A marker is considered "associated" if the association statistic is observed below a given threshold. The threshold used is arbitrary, but is set by default at an unadjusted $P \leq 0.05$. The significance of the ratio is determined through permutation using an empirical P-value derived from the proportion of the ratios for the permuted datasets that are greater than or equal to the observed ratio [12]. We performed 10000 permuted GWAS analyses for each of the diagnostic, ancestry strata for both the discovery and replication datasets. The pedSRT determines association using the transmission disequilibrium test (TDT) [18] as implemented

in PLINK [17]. In a case-control model, permutation is performed using case-randomisation. In the TDT design case-randomisation is performed by creating a pseudo-sibling. The pseudo-sibling is created from the non-transmitted alleles from the parents. Within each permutation cycle either the proband or pseudo-sibling is considered the "case". Alternate case-randomisation for the TDT are implemented in PLINK using the alternate phenotype routine.

It is important to note that to reduce type-I error in the SRT due to inflation of the original association signal, for each permutation "associated" SNPs are assigned according to their rank in the dataset [12]. In short, the numbers of SNPs (*T*) that meet the "associated" threshold are calculated from the primary dataset. For each permuted dataset the top *T* SNPs are termed "associated".

All SNP ratio statistics were calculated using custom scripts in STATA version 10 (Stata Corp, TX, USA).

Gene Tagging

Individual SNP codes from the Illumina 1M Infinium SNP array platform were updated to reflect build 130 of dbSNP. SNPs were assigned to genes using gene criteria from the dbSNP/NCBI criteria; namely if the SNP resides within the locus containing the gene transcript including 2kb 5' and 500bp 3' of the transcript. The gene assignment protocol was performed using NCBI criteria and facilitated using the file b130_SNPContigLocusId_36_3.bcp available at ftp://ftp.ncbi.nih.gov/snp/organisms/human_9606/database/organism_data/.

Gene-Set Selection

Gene-sets were described using the gene ontology database (GO; www.geneontology.org)[19]. Gene lists were obtained from the OBO format 1.2 database

release available from http://www.geneontology.org/GO.downloads.ontology.shtml (build release date 15-12-2009). Gene-ontology terms are structured in a semi-hierarchical relationship within the *cellular component, molecular function* and *biological process* nodes. Daughter ontology terms are more specialized and parent ontology terms are less specialized. But unlike a hierarchy, a term may have more than one parent term.

Parent terms were populated by their daughter terms to describe a composite list of genes for each term. SNP ratios were calculated on gene-ontology terms with greater than 20 SNPs but less than 2000 SNPs and greater than 1 gene but no more than 1000 genes. A total of 6853 GO terms met these criteria. To account for identity of terms we merged those GO terms containing identical gene lists; in total the list of unique terms is 5955.

Simulation of GO Terms

As mentioned above, the GO terms used in this study can show considerable overlap due to term redundancy, biological overlap and the hierarchical nature of the database. Simulations were performed to calculate the null distribution and subsequent expectancy for the total number of associated GO terms at a given threshold in a single study given the GO terms used.

We performed 1000 pedSRT permutations on a case-randomised sample derived from 1248 families from the Discovery dataset. A GWAS TDT was performed on each dataset followed by pedSRT using 10000 additional permutations on the 5995 GO terms. For each of the 1000 original permutations the proportion of the 5955 GO terms that met a significance threshold of $P \leq 0.05$ in the subsequent 10000 was calculated. The mean proportion across the 1000 permutations was used to predict the expected number of associated GO terms in a dataset.

Pathway Enrichment Map Generation

Visual representation of overlap in enriched GO terms was performed using the EnrichmentMap (http://baderlab.org/Software/EnrichmentMap [20]) plugin for Cytoscape 2.8.0 (http://www.cytoscape.org/ [21]). Consistent with the author's recommendations for use with the Gene Ontology database, nodes were joined if the overlap coefficient was ≥0.5.

Results

Across all analysis in the discovery dataset, 1035 unique GO terms show association enrichment at SRT-P-value≤0.05. Examination of those GO terms that show strong enrichment (SRT-P-value<0.001) highlights diverse processes such as regulation of cell division (mitosis and meiosis), ribosome processing and apoptosis. A visual representation of enriched pathways is shown in SUPPLEMENTARY FIGURE 1. A summary of the total number of GO terms that show enrichment at SRT-P-value≤0.05 is given in TABLE 2. Based on simulated data, 4.46% (SD=0.8%) of the 5995 unique but non-independent pathways are expected to be associated at SRT-P≤0.05 level. Given this level we would expect 267 GO terms to be associated per experiment. To provide a greater distinction of potentially important GO terms we examined the overlap of enriched GO terms in an independent replication dataset. Based on 4.46% of GO terms showing enrichment we would expect to observe replication for 12 of the 5995 pathways. All individual discovery samples show more GO terms associated than would be expected by chance (see Expected 1; TABLE 2). Moreover, the overlap between the discovery and replication sample also show enrichment over what would be expected by chance (see Expected Replication 2; TABLE2). When we use a more cautious interpretation based on the total number of observed associated GO terms in the discovery data and the predicted replication of 4.46% we would expect to replicate is between 15 and 17 pathways (see Expected 3; TABLE 2).

Under this model we still show enriched replication for each "*Diagnosis* | *Ancestry"* groupings. Overall compared to simulated data we observe between 1.5 and 3.2-fold enrichment in the overlap of pathways in the discovery and replication dataset above what can be expected by chance.

A summary of the replicated pathways, summary statistics, gene number and genes tagged in this analysis is shown in TABLE 4-7 (full lists of replicated pathways can be found in SUPPLEMENTARY TABLES 1A-1C). A total of eighty-eight unique GO terms were shown to be replicated within analytic groupings (see SUPPLEMENTARY TABLE 2), twenty-two GO terms were replicated within two of the analytic groupings and four GO terms were replicated within three of the analytic groupings (see TABLE 3). Replication was only considered within strata, such that for example, GO terms identified in the discovery Strict | European analyses were examined in the Strict | European replication dataset. The four GO terms that show enrichment across three groupings are GO:0006090, GO:0032872, GO:0032874 and GO:0042156, involved in pyruvate metabolism, regulation of the MAPK cascade and zinc-mediated transcriptional activation. A visual representation of replicated enriched pathways is shown in SUPPLEMENTARY FIGURE 2.

Discussion

The interpretation of GWAS data purely on the strength of association data is challenging where the distribution of association is close to or barely exceeding what is expected by the number of tests. In the absence of clear association enrichment across the entire dataset, interpretation has relied upon rank-order or via the application of sub-optimal significance thresholds which juggle type-I and type-II error. The principle of association enrichment approaches is to discover whether within this milieu of data there are underlying patterns to

the association. In these approaches we ask whether SNPs that are linked to genes of common function show greater proportion of nominal association than expected by chance. Although a modest association signal at an individual SNP within a gene may not warrant further investigation, the cumulative association of SNPs within a gene-family may offer insight into the biology of the disorder.

Gene enrichment approaches have been primarily developed to aid interpretation of data from microarray expression studies. In this context each gene is tagged by either one or a small number of probes regardless of gene size. However, when applying these technologies to SNP-based data we do not measure gene-wise variation or gene-wise association; instead we can potentially examine multiple points of association at any given gene using many tagging SNPs. This brings additional challenges and bias. When applying association enrichment we must account for and correct for these potential bias in these data. Firstly, when examining larger genes we utilise more SNP markers to tag the variation than for smaller genes. If we choose a maximum association signal approach per gene, we observe by chance, an inflated signal for the larger genes. By calculating the ratio of associated to not associated SNPs we can adjust each GO term to the total number of SNPs examined per GO term. Secondly, where multiple markers tag a gene, one might observe multiple strong association signals due to strong linkage disequilibrium between the associated markers. To reduce this effect we calculate significance of the data through permutation. Permutation is performed by case-randomisation within families where a pseudo-control sibling is created from the alleles that are not transmitted to the proband. By using the non-transmitted alleles we retain the linkage disequilibrium structure across the genome thereby retaining linkage-disequilibrium-related inflation in the original association

signal.

We have applied the SNP ratio-test to family-based data from the AGP to identify eighty-eight gene sets from the gene-ontology database that show a replicated enrichment for association signal. Of the overlapping GO terms, we observe enrichment in sets involved in diverse biological processes including pyruvate metabolism, transcription factor activation, cell-signalling and cell-cycle regulation.

One of the strongest findings from the Discovery and Replication findings was observed across the "*Strict* diagnosis | All ancestries" grouping for the GO term GO:0031146; SCF-dependent proteasomal ubiquitin-dependent protein catabolic process (Discovery SRT-P=0.0001; Replication SRT-P=0.0009). GO:0031146 is described by only two genes (*FBXO31* and *FBXO6*). Both genes are members of the F-Box protein family, which are involved in a variety of molecular and cellular functions including protein degradation, synapse formation and circadian rhythm [22]. *FBXO6* has also been suggested as a putative biomarker for autism [23] as one of thirteen genes highlighted in the work of Nishimura and Brown [24] who show differential expression at this gene in the lymphoblastoid cell lines from individuals with both the *FMR1* mutation and autism compared to typically developing controls.

Those GO terms that show replication across multiple diagnostic and ancestral groups are also noteworthy as they are robust to differences in sampling used in our analyses. Four replicated GO terms were observed in three analytic groupings (see TABLE 3). These include GO:0006090, GO:0032872, GO:0032874 and GO:0042156. GO:0006090 (pyruvate metabolic process) describes a group of 39-tagged genes (see SUPPLEMENTARY TABLE 3) covered by 589 SNPs. These genes are involved in the

biological processes connecting the chemical reactions and pathways involving pyruvate. Pyruvate metabolism is a component of the energy metabolism pathway which has received considerable attention with respect to autism. The biological plausibility of the pyruvate metabolic process association enrichment is supported by numerous studies showing evidence of aberration in pyruvate levels in individuals with autism [25]. The GO term GO:0042156 (zinc-mediated transcriptional activator activity) describes a group of three genes tagged by 37 SNPs (*MTF1*, *RNF4*, *and ZNF384*). One of the constituent genes, *MTF1*, human metal-regulatory transcription-factor-1, has previously warranted investigation as putative candidate gene for autistic disorder under an environmental exposure model of autism [26]. Finally, GO:0032872 (regulation of stress-activated MAPK cascade) and GO:0032874 (positive regulation of stress-activated MAPK cascade), which differ by a single gene (see SUPPLEMENTARY TABLE 3) describe 10 and 9 genes, and 122 and 116 SNPs respectively. These pathways are involved in increasing the signalling of the stress-related mitogen-activated protein kinase (MAPK) signalling pathway. Stress-activated MAPKs are thought to play a critical role in modulating inflammation, DNA damage response, apoptosis in cancer [27] and negative regulation of cell cycle progression [28,29]. Cell cycle progression and DNA damage response are also highlighted in enriched replicated GO terms in these analyses, for example GO:0032404 (mismatch repair complex binding) and GO:0031571 (G1/S DNA damage checkpoint).

In a recent study by this group we explored enrichment in GO terms for rare deleted CNVs [30]. Using individuals from the Discovery Group we identified twenty-four enriched GO terms that show enrichment in rare CNV at FDR q<.05 that highlighted five biological domains; namely cell proliferation, cell projection and motility, MHC-I, GTPase/RAS

signalling, and Kinase activation/regulation. We do not observe any overlap between the eighty-eight gene-sets showing replicated enrichment in the GWAS data with the twenty-four significant GO terms identified for rare structural variation. However, we do observe some overlap for GO terms enriched only in the Discovery dataset. These include overlap in "cell migration", "cell motility", "cell morphogenesis" and GO terms identified as having a role in protein kinase regulation.

We can take some encouragement that highlighted pathways are supported in the autism literature. We have emphasized biological plausibility of some of these pathways with autism and ASD. However, one major caveat when interpreting these data is whether this overlapping evidence reflects the considerable literature surrounding autism research and is therefore coincidence, or is biologically meaningful concordance.

Pathway approaches, such as the SRT and pedSRT, can be applied to research questions using candidate gene list. Candidate genes rely upon the selection of genes and markers based on previous knowledge of biology, function and position of the gene or marker. The pathway approach in the form used in this manuscript applies a "hypothesis-free" design, in which we examine all GO terms regardless of putative role. In a recent autism GWAS described by Wang and colleagues [4], the authors applied a hypothesis-testing candidate gene approach using their own methodology [11] to examine whether a group of cadherin and neurexin genes showed enrichment in their association data. The authors conclude that there was association enrichment for both a group of cadherin, and cadherin plus neurexin genes (P=0.02, P=0.004 respectively). We applied our approach to these gene-lists in our data (data not shown). Using the pedSRT, which differs in statistical method and gene-to-SNP assignment to that of Wang and colleagues, we do not observe significant association

signal enrichment in either the discovery or replication dataset for any of the analytic groupings.

To further explore potential overlap of our data and other GWAS we examined whether previously implicated genes from recent autism GWAS were present in the GO terms identified in this study. None of the genes that overlap with the top-associated SNPs from previous GWAS described by Wang and colleagues [4] (*CDH22*, *CTNNA3*, *DMD*, *FEZF2*, *LOC100132914*, *LRRC1* and *SYT17*) and Weiss and colleagues [6] (*ACTN2*, *ADA*, *CENPC1*, *CRIM1*, *CTNNA3*, *CUGBP2*, *GAS2*, *IQGAP2*, *JARID2*, *SGCD* and *XG*) appeared in the eighty-eight unique GO terms showing overlap in these analyses. Moreover, we do not observe overlap with those genes highlighted by the authors as residing close to their maximal association peaks, namely *SEMA5A*, *TAS2R1* and *CDH9*, *CDH10*.

The gene-ontology database is continuously updated as evidence is gathered on gene biology. The build of the database used in these analyses contains information on 17703 genes, compared to less than 5000 for databases such as KEGG. However, not all genes are tagged to GO terms. This is exemplified by the *MACROD2* gene, which contained SNPs showing the strongest association signal from our previous GWAS analyses [7]. Over time more information will be gathered on the biological role and interactions between these genes to further annotate these terms.

In addition to single gene effects such as *MACROD2*, data presented in this analysis may offer some additional insight into biological processes, within which genetic risk for autism may lie. This can include hypothesis-free gene-lists such as those in the gene-ontology dataset, or more hypothesis driven candidate gene lists highlighting previous linkage, association or biology. The application of pedSRT to our GWAS data has highlighted

biological processes previously implicated in autism and offers impetus to re-examine these processes based on evidence from genome-wide investigation. Association enrichment analysis provides additional evidence from GWAS data to identify genetic risk variants and genes and prioritise biological processes for further research into areas such as biomarker discovery, gene-gene interaction analyses and identification of putative drug targets.

Financial Disclosure
The authors confirm no financial conflicts related to the findings of this work.

Acknowledgments
The authors gratefully acknowledge the families participating in the study and the main funders of the AGP: Autism Speaks (USA), the Health Research Board (HRB; Ireland; AUT/2006/1, AUT/2006/2, PD/2006/48), The Medical Research Council (MRC; UK), Genome Canada/Ontario Genomics Institute, and the Hilibrand Foundation (USA). Additional support for individual groups was provided by the US National Institutes of Health (NIH grants: HD055751, HD055782, HD055784, MH52708, MH55284, MH061009, MH06359, MH066673, MH080647, MH081754, MH66766, NS026630, NS042165, NS049261), the Canadian Institutes for Health Research (CIHR), Assistance Publique - Hôpitaux de Paris (France), Autism Speaks UK, Canada Foundation for Innovation/Ontario Innovation Trust, Deutsche Forschungsgemeinschaft (grant: Po 255/17-4) (Germany), EC Sixth FP AUTISM MOLGEN, Fundação Calouste Gulbenkian (Portugal), Fondation de France, Fondation FondaMental (France), Fondation Orange (France), Fondation pour la Recherche Médicale (France), Fundação para a Ciência e Tecnologia (Portugal), the Hospital for Sick Children Foundation and University of Toronto (Canada), INSERM (France), Institut Pasteur (France), the Italian Ministry of Health (convention 181 of 19.10.2001), the John P Hussman Foundation (USA), McLaughlin

Centre (Canada), Ontario Ministry of Research and Innovation (Canada), the Seaver Foundation (USA), the Swedish Science Council, The Centre for Applied Genomics (Canada), the Utah Autism Foundation (USA) and the Wellcome Trust core award 075491/Z/04 (UK). D.P. is supported by fellowships from the Royal Netherlands Academy of Arts and Sciences (TMF/DA/5801) and the Netherlands Organization for Scientific Research (Rubicon 825.06.031). S.W.S holds the GlaxoSmithKline-CIHR Pathfinder Chair in Genetics and Genomics at the University of Toronto and the Hospital for Sick Children (Canada).

References

1. Fombonne E: Epidemiology of pervasive developmental disorders. *Pediatr Res* 2009; **65:** 591-598.

2. Fernell E, Gillberg C: Autism spectrum disorder diagnoses in Stockholm preschoolers. *Res Dev Disabil* 2010; **31:** 680-685.

3. Bailey A, Le Couteur A, Gottesman I *et al*: Autism as a strongly genetic disorder: evidence from a British twin study. *Psychol Med* 1995; **25:** 63-77.

4. Wang K, Zhang H, Ma D *et al*: Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature* 2009; **459:** 528-533.

5. Ma D, Salyakina D, Jaworski JM *et al*: A genome-wide association study of autism reveals a common novel risk locus at 5p14.1. *Ann Hum Genet* 2009; **73:** 263-273.

6. Weiss LA, Arking DE, Daly MJ, Chakravarti A: A genome-wide linkage and association scan reveals novel loci for autism. *Nature* 2009; **461:** 802-808.

7. Anney R, Klei L, Pinto D *et al*: A genome-wide scan for common alleles affecting risk for autism. *Hum Mol Genet* 2010; **19:** 4072-4082.

8. Purcell SM, Wray NR, Stone JL *et al*: Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 2009; **460:** 748-752.

9. Holmans P, Green EK, Pahwa JS *et al*: Gene ontology analysis of GWA study data sets provides insights into the biology of bipolar disorder. *Am J Hum Genet* 2009; **85:** 13-24.

10. Wang K, Li M, Hakonarson H: Analysing biological pathways in genome-wide association studies. *Nat Rev Genet* 2010; **11:** 843-854.

11. Wang K, Li M, Bucan M: Pathway-Based Approaches for Analysis of Genomewide Association Studies. *Am J Hum Genet* 2007; **81**.

12. O'Dushlaine C, Kenny E, Heron EA *et al*: The SNP ratio test: pathway analysis of genome-wide association datasets. *Bioinformatics* 2009; **25:** 2762-2763.

13. Lord C, Rutter M, Le Couteur A: Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J Autism Dev Disord* 1994; **24:** 659-685.

14. Lord C, Rutter M, Goode S *et al*: Autism diagnostic observation schedule: a standardized observation of communicative and social behavior. *J Autism Dev Disord* 1989; **19:** 185-212.

15. Risi S, Lord C, Gotham K *et al*: Combining information from multiple sources in the diagnosis of autism spectrum disorders. *J Am Acad Child Adolesc Psychiatry* 2006; **45:** 1094-1103.

16. Lee AB, Luca D, Klei L, Devlin B, Roeder K: Discovering genetic ancestry using spectral graph theory. *Genet Epidemiol* 2010; **34:** 51-59.

17. Purcell S, Neale B, Todd-Brown K *et al*: PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81:** 559-575.

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

19. Ashburner M, Ball CA, Blake JA *et al*: Gene ontology: tool for the unification of

biology. The Gene Ontology Consortium. *Nat Genet* 2000; **25:** 25-29.

20. Merico D, Isserlin R, Stueker O, Emili A, Bader GD: Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS One* 2010; **5:** e13984.

21. Shannon P, Markiel A, Ozier O *et al*: Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003; **13:** 2498-2504.

22. Ho MS, Ou C, Chan YR, Chien CT, Pi H: The utility F-box for protein destruction. *Cell Mol Life Sci* 2008; **65:** 1977-2000.

23. Nishimura Y, Martin CL, Vazquez-Lopez A *et al*: Genome-wide expression profiling of lymphoblastoid cell lines distinguishes different forms of autism and reveals shared pathways. *Hum Mol Genet* 2007; **16:** 1682-1698.

24. Brown V, Jin P, Ceman S *et al*: Microarray identification of FMRP-associated brain mRNAs and altered mRNA translational profiles in fragile X syndrome. *Cell* 2001; **107:** 477-487.

25. Haas RH: Autism and mitochondrial disease. *Dev Disabil Res Rev* 2010; **16:** 144-153.

26. Serajee FJ, Nabi R, Zhong H, Huq M: Polymorphisms in xenobiotic metabolism genes and autism. *J Child Neurol* 2004; **19:** 413-417.

27. Dhillon AS, Hagan S, Rath O, Kolch W: MAP kinase signalling pathways in cancer. *Oncogene* 2007; **26:** 3279-3290.

28. Bulavin DV, Fornace AJ, Jr.: p38 MAP kinase's emerging role as a tumor suppressor. *Adv Cancer Res* 2004; **92:** 95-118.

29. Bradham C, McClay DR: p38 MAPK in development and cancer. *Cell Cycle* 2006; **5:** 824-828.

30. Pinto D, Pagnamenta AT, Klei L *et al*: Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* 2010; **466:** 368-372.

Tables

TABLE 1: Sample size for the Discovery and Replication samples for each Diagnostic/Ancestral subset. Each sample is made up of parent-proband trios with atleast a spectrum diagnosis of ASD (Spectrum\|All). Individuals were selected from the total if they were determined to be of European ancestry (Spectrum\|European) or if they show a more restrictive diagnosis of Autism (Strict\|All) or if they met both restrictive ancestral and diagnostic thresholds (Strict\|European).

Diagnosis | Ancestry	Discovery (n)	Replication (n)	Combined (n)
Total	1248	874	2122
Spectrum | All	1248	874	2122
Spectrum | European	1169	704	1873
Strict | All	775	591	1366
Strict | European	726	477	1203

TABLE 2. Summary of enriched GO terms and overlap in the Discovery and Replication sample. (1) Expected associated GO terms given 4.46% of 5995 pathways showing enrichment, (2) Expected replication given 4.46% of 5995 unique pathways showing enrichment in simulated dataset. (3) Expected replication given proportion of associated pathways observed in the Discovery and an expected Replication-set enrichment of 4.46%.

Diagnosis | Ancestry	Observed Enriched GO Terms in Discovery Sample	Expected Discovery(1)	Observed Replicated GO Terms	Expected Replication(2)	Expected Replication(3)
Spectrum | All	389	267	38	12	17
Spectrum | European	392	267	34	12	17
Strict | All	329	267	22	12	15
Strict | European	354	267	24	12	16

TABLE 3: GO Terms showing replicated enrichment in two or more analytic groupings. Tagged gene lists for each of the GO Term are summarised in SUPPLEMENTARY TABLE 3.

Go Term	GO Name	Tagged Genes	Groupings
GO:0042156	zinc-mediated transcriptional activator activity	3	All & *Spectrum* |European
GO:0006090	pyruvate metabolic process	39	European & *Strict* | All
GO:0032872	regulation of stress-activated MAPK cascade	10	European & *Strict* | All
GO:0032874	positive regulation of stress-activated MAPK cascade	9	European & *Strict* | All
GO:0004758	serine C-palmitoyltransferase activity	3	European
GO:0017059	serine C-palmitoyltransferase complex	5	European
GO:0032494	response to peptidoglycan	6	European
GO:0043330	response to exogenous dsRNA	8	European
GO:0045070	positive regulation of viral genome replication	4	European
GO:0006730	one-carbon metabolic process	98	Spectrum
GO:0008276	protein methyltransferase activity	47	Spectrum
GO:0010712	regulation of collagen metabolic process	11	Spectrum
GO:0016278	lysine N-methyltransferase activity	31	Spectrum
GO:0031489	myosin V binding	2	Spectrum
GO:0032059	bleb	3	Spectrum
GO:0032400	melanosome localization	4	Spectrum
GO:0032402	melanosome transport	9	Spectrum
GO:0032755	positive regulation of interleukin-6 production	18	Spectrum
GO:0042054	histone methyltransferase activity	38	Spectrum
GO:0044246	regulation of multicellular organismal metabolic process	14	Spectrum
GO:0051648	vesicle localization	5	Spectrum
GO:0060346	bone trabecula formation	2	Spectrum
GO:0070382	exocytic vesicle	4	Spectrum
GO:0005876	spindle microtubule	27	Strict
GO:0015232	heme transporter activity	3	Strict
GO:0046888	negative regulation of hormone secretion	25	Strict

TABLE 4: Top 10 association enrichments of pedSRT for overlapping GO terms for analyses of families of all ancestries with a proband with a Spectrum diagnosis. Ratio refers to the ratio of associated to non-associated SNPs per GO term. P refers to the empirical significance generated from 10000 permutations.

GO Term	GO Category	GO Name	Identity	Total SNPs	Discovery		Replication		Total Genes	Tagged Genes
					Ratio	P	Ratio	P		
GO:0006730	Biological process	one-carbon metabolic process		1701	0.1053	0.0001	0.0665	0.0437	107	98
GO:0042156	Molecular function	zinc-mediated transcriptional activator activity		37	0.6818	0.0001	0.2333	0.0201	3	3
GO:0032059	Cellular component	bleb		64	0.3333	0.0004	0.1636	0.0284	3	3
GO:0000318	Molecular function	protein-methionine-R-oxide reductase activity		44	0.4194	0.0048	0.2571	0.0354	2	2
GO:0050705	Biological process	regulation of interleukin-1 alpha secretion	GO:0050717	56	0.2444	0.0085	0.1915	0.0225	3	3
GO:0032650	Biological process	regulation of interleukin-1 alpha production	GO:0032730	58	0.2340	0.0091	0.1837	0.0238	4	4
GO:0007567	Biological process	parturition		124	0.1376	0.0152	0.1273	0.0244	7	7
GO:0031489	Molecular function	myosin V binding		56	0.2174	0.0167	0.3023	0.0020	2	2
GO:0032402	Biological process	melanosome transport	GO:0051904	118	0.1456	0.0189	0.1683	0.0083	9	9
GO:0046628	Biological process	positive regulation of insulin receptor signaling pathway		43	0.1944	0.0194	0.1944	0.0180	3	3

TABLE 5: Top 10 association enrichments of pedSRT for overlapping GO terms for analyses of families of European ancestries with a proband with a Spectrum diagnosis. Ratio refers to the ratio of associated to non-associated SNPs per GO term. P refers to the empirical significance generated from 10000 permutations.

GO Term	GO Category	GO Name	Identity	Total SNPs	Discovery		Replication		Total Genes	Tagged Genes
					Ratio	P	Ratio	P		
GO:0042156	Molecular function	zinc-mediated transcriptional activator activity		37	0.6087	0.0001	0.2333	0.0195	3	3
GO:0006730	Biological process	one-carbon metabolic process		1701	0.0890	0.0017	0.0698	0.0297	107	98
GO:0009566	Biological process	fertilization		609	0.1362	0.0044	0.0973	0.0398	44	39
GO:0032872	Biological process	regulation of stress-activated MAPK cascade		122	0.1619	0.0059	0.1091	0.0476	10	10
GO:0006090	Biological process	pyruvate metabolic process		589	0.1113	0.0065	0.1219	0.0013	40	39
GO:0042788	Cellular component	polysomal ribosome		22	0.3750	0.0070	0.5714	0.0006	3	3
GO:0031235	Cellular component	intrinsic to internal side of plasma membrane		147	0.2049	0.0074	0.1667	0.0241	3	3
GO:0032874	Biological process	positive regulation of stress-activated MAPK cascade		116	0.1600	0.0074	0.1154	0.0412	9	9
GO:0032494	Biological process	response to peptidoglycan		78	0.2000	0.0080	0.2000	0.0079	6	6
GO:0045070	Biological process	positive regulation of viral genome replication		50	0.2500	0.0090	0.1905	0.0252	4	4

TABLE 6: Top 10 association enrichments of pedSRT for overlapping GO terms for analyses of families of all ancestries with a proband with a Strict diagnosis. Ratio refers to the ratio of associated to non-associated SNPs per GO term. P refers to the empirical significance generated from 10000 permutations.

GO Term	GO Category	GO Name	Identity	Total SNPs	Discovery		Replication		Total Genes	Tagged Genes
					Ratio	P	Ratio	P		
GO:0031146	Biological process	SCF-dependent proteasomal ubiquitin-dependent protein catabolic process		21	0.9091	0.0001	0.5000	0.0009	2	2
GO:0031571	Biological process	G1/S DNA damage checkpoint		43	0.3030	0.0007	0.1944	0.0219	4	4
GO:0015232	Molecular function	heme transporter activity		29	0.2083	0.0011	0.1600	0.0070	3	3
GO:0030276	Molecular function	clathrin binding		489	0.1425	0.0016	0.1063	0.0231	8	7
GO:0004738	Molecular function	pyruvate dehydrogenase activity	GO:0004739	22	0.1579	0.0088	0.1000	0.0393	3	3
GO:0016139	Biological process	glycoside catabolic process		35	0.2963	0.0091	0.2069	0.0282	3	3
GO:0000244	Biological process	assembly of spliceosomal tri-snRNP		34	0.2593	0.0193	0.2593	0.0184	4	4
GO:0032642	Biological process	regulation of chemokine production		139	0.1301	0.0301	0.1301	0.0306	15	14
GO:0033630	Biological process	positive regulation of cell adhesion mediated by integrin		67	0.1964	0.0327	0.2182	0.0253	3	2
GO:0032872	Biological process	regulation of stress-activated MAPK cascade		122	0.1193	0.0333	0.1509	0.0065	10	10

TABLE 7: Top 10 association enrichments of pedSRT for overlapping GO terms for analyses of families of European ancestries with a proband with a Strict diagnosis. Ratio refers to the ratio of associated to non-associated SNPs per GO term. P refers to the empirical significance generated from 10000 permutations.

GO Term	GO Category	GO Name	Identity	Total SNPs	Discovery		Replication		Total Genes	Tagged Genes
					Ratio	P	Ratio	P		
GO:0034142	Biological process	toll-like receptor 4 signaling pathway	GO:0070427	22	0.8333	0.0001	0.5714	0.0012	2	2
GO:0033083	Biological process	regulation of immature T cell proliferation		31	0.4762	0.0002	0.2917	0.0161	3	3
GO:0040036	Biological process	regulation of fibroblast growth factor receptor signaling pathway		133	0.3300	0.0004	0.1982	0.0235	7	7
GO:0070423	Biological process	nucleotide-binding oligomerization domain containing signaling pathway	GO:0070431	41	0.3667	0.0011	0.2424	0.0272	4	4
GO:0004758	Molecular function	serine C-palmitoyltransferase activity	GO:0016454	145	0.1885	0.0032	0.1328	0.0308	3	3
GO:0070555	Biological process	response to interleukin-1		608	0.1115	0.0035	0.0975	0.0112	17	17
GO:0045879	Biological process	negative regulation of smoothened signaling pathway		111	0.3059	0.0051	0.2472	0.0197	5	5
GO:0017059	Cellular component	serine C-palmitoyltransferase complex	GO:0031211	164	0.1631	0.0060	0.1310	0.0256	5	5
GO:0070391	Biological process	response to lipoteichoic acid		41	0.2813	0.0074	0.1714	0.0437	4	4
GO:0010830	Biological process	regulation of myotube differentiation	GO:0010832	152	0.1783	0.0093	0.1343	0.0404	3	3

Author List

Richard J.L. Anney [1]†, Elizabeth A. Heron [1], Ricardo Segurado [1], Elaine M. Kenny [1], Colm O'Dushlaine [1,2], Brian L. Yaspan[3], Elena Parkhomenko[4], **The Autism Genome Project**‡, Joseph D. Buxbaum[4], James S. Sutcliffe[3], Michael Gill [1], Louise Gallagher [1].

1. Department of Psychiatry, Trinity College Dublin, Dublin, IRELAND

2. The Eli and Edythe L. Broad Institute, Massachussettes Institute of Technology, Cambridge, Massachussettes, USA

3. Department of Molecular Physiology and Biophysics, Vanderbilt Kennedy Centre, Nashville, Tennessee, USA

4. The Seaver Autism Center for Research and Treatment, Department of Psychiatry, Mount Sinai School of Medicine, New York, USA

†Corresponding Author; Richard JL Anney (anneyr@tcd.ie)

Postal Address: Autism Genetics Group, Department of Psychiatry, Institute of Molecular Medicine

Trinity Centre for Health Sciences, St. James' Hospital, James' Street, Dublin 8, Ireland

Tel: ++353 1 896-8461

Fax: ++353 1 896-3405

‡Full list of Autism Genome Project Members is given below

Author Contributions

RJLA wrote the manuscript. EAH, RS and RJLA developed statistical methods applied in the manuscript. RJLA designed the experiment and scripted all code used within the

manuscript. EAH, RS, EMK and COD developed the principle of the SRT experiments and RJLA developed the method for family-based data (pedSRT). EAH, RS, EMK, COD, BLY, MG and LG aided in manuscript preparation. RJLA, EAH, RS, EMK COD, BY, EP, JDB and JSS discussed research strategies and data through the "pathway-based analysis working group". Additional intellectual support and guidance was provided through the AGP including BD, ADP, EHC, PS, JTG, CK, KW, HH and EM

The Autism Genome Project
§Member of Senior Investigator Committee for the Autism Genome Project

Canada
Department of Psychiatry, University of British Columbia, V6T 2A1, Canada.
Anthony J. Bailey§

Disciplines of Genetics and Medicine, Memorial University of Newfoundland, St John's Newfoundland A1B 3V6, Canada.
Bridget A. Fernandez

Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton, Ontario L8N 3Z5, Canada.
Peter Szatmari§, Ann P. Thompson

The Centre for Applied Genomics and Program in Genetics and Genomic Biology, The Hospital for Sick Children, Toronto, Ontario M5G 1L7, Canada.
Stephen W. Scherer§, Andrew Patterson§, Christian R. Marshall, Dalila Pinto

Centre for Addiction and Mental Health, Clarke Institute and Department of Psychiatry, University of Toronto, Toronto, Ontario M5G 1X8, Canada.
John B. Vincent

Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5S 1A1, Canada.
Stephen W. Scherer§

Division of Psychiatry, McGill University, Montreal, Quebec H3A 1A1, Canada.
Eric Fombonne

France
INSERM U952 and CNRS UMR 7224 and UPMC Univ Paris 06, Paris 75005, France.
Catalina Betancur§

INSERM U955, Fondation FondaMental, APHP, Hôpital Robert Debré, Child and Adolescent Psychiatry, 75019 Paris, France.
Richard Delorme

INSERM U995, Department of Psychiatry, Groupe Hospitalier Henri Mondor-Albert Chenevier, AP-HP; University Paris 12, Fondation FondaMental, Créteil 94000, France.
Marion Leboyer

Human Genetics and Cognitive Functions, Institut Pasteur; University Paris Diderot-Paris 7, CNRS URA 2182, Fondation FondaMental, 75015 Paris, France.
Thomas Bourgeron

Octogone/CERPP (Centre d'Eudes et de Recherches en Psychopathologie), University de Toulouse Le Mirail, Toulouse Cedex 31058, France.
Carine Mantoulan, Bernadette Roge, Maïté Tauber

Germany
Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, J.W. Goethe University Frankfurt, 60528 Frankfurt, Germany.
Christine M. Freitag§, Fritz Poustka, Eftichia Duketis

Division of Molecular Genome Analysis, German Cancer Research Center (DKFZ), Heidelberg 69120, Germany.
Sabine M. Klauck§, Annemarie Poustka‡

Greece
University Department of Child Psychiatry, Athens University, Medical School, Agia Sophia Children's Hospital, 115 27 Athens, Greece.
Katerina Papanikolaou, John Tsiantis

Ireland
Autism Genetics Group, Department of Psychiatry, School of Medicine, Trinity College, Dublin 8, Ireland.
Louise Gallagher§, Michael Gill§, Richard Anney, Nadia Bolshakova, Sean Brennan, Gillian Hughes,

Jane McGrath, Alison Merikangas, Elizabeth A. Heron

School of Medicine and Medical Science, University College, Dublin 4, Ireland.
Sean Ennis§, Andrew Green, Jillian P. Casey, Judith M. Conroy, Regina Regan, Naisha Shah

Italy
Department of Biology, University of Bologna, 40126 Bologna, Italy.
Elena Maestrini§, Elena Bacchelli, Fiorella Minopoli

Neuropsichiatria Infantile, Ospedale Santa Croce, 61032 Fano, Italy.
Vera Stoppioni

Stella Maris Institute for Child and Adolescent Neuropsychiatry, 56128 Calambrone (Pisa), Italy.
Agatino Battaglia§, Roberta Igliozzi, Barbara Parrini, Raffaella Tancredi

Portugal
Hospital Pediátrico de Coimbra, 3000 – 076 Coimbra, Portugal.
Guiomar Oliveira§, Joana Almeida, Frederico Duque

Instituto Nacional de Saude Dr Ricardo Jorge, Av Padre Cruz 1649-016, Lisbon, Portugal & BioFIG—Center for Biodiversity, Functional and Integrative Genomics, Campus da FCUL, C2.2.12, Campo Grande, 1749-016 Lisboa, Portugal & Instituto Gulbenkian de Cîencia, Rua Quinta Grande, 2780-156 Oeiras, Portugal.
Astrid Vicente§, Tiago Magalhaes, Catarina Correia, Ana F. Sequeira

Sweden
Gillberg Neuropsychiatry Centre, Sahlgrenska Academy, University of Gothenburg, S41345 Gothenburg, Sweden.
Christopher Gillberg, Gudrun Nygren

The Netherlands
Department of Child and Adolescent Psychiatry, University Medical Center, Utrecht 3508 GA, The Netherlands.
Maretha de Jonge, Herman Van Engeland, Jacob A.S.Vorstman

United Kingdom
Autism Centre for Education and Research, School of Education, University of Birmingham, B15 2TT, UK.
Kerstin Wittemeyer

Newcomen Centre, Guy's Hospital, London SE1 9RT, UK.
Gillian Baird

Department of Child and Adolescent Psychiatry, Institute of Psychiatry, King's College London, London SE5 8AF, UK.
Patrick F. Bolton

Social, Genetic and Developmental Psychiatry Centre, Institute Of Psychiatry, King's College London, London SE5 8AF, UK.
Michael L. Rutter

Academic Department of Child Psychiatry, Booth Hall of Children's Hospital, Blackley, Manchester M9 7AA, UK.
Jonathan Green

Centre for Integrated Genomic Medical Research, University of Manchester, Manchester M13 9PT, UK.
Janine A. Lamb

Department of Medicine, School of Epidemiology and Health Science, University of Manchester, Manchester M13 9PT, UK.
Andrew Pickles

Institute of Neuroscience, and Institute of Health and Society, Newcastle University, Newcastle Upon Tyne, NE1 7RU, UK.
Jeremy R. Parr, Ann Le Couteur, Tom Berney, Helen McConachie

Department of Psychiatry, University of Oxford, Warneford Hospital, Headington, Oxford OX3 7JX, UK.
Simon Wallace, Marc Coutanche, Suzanne Foley, Kathy White

Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, UK.
Anthony P. Monaco§, Richard Holt, Penny Farrar, Alistair T. Pagnamenta, Ghazala K. Mirza, Jiannis

Ragoussis, Inês Sousa, Nuala Sykes, Kirsty Wing

United States of America
Department of Psychiatry, Division of Child and Adolescent Psychiatry and Child Development, Stanford University School of Medicine, Stanford, California 94304, USA.
Joachim Hallmayer§

Department of Human Genetics, University of California—Los Angeles School of Medicine, Los Angeles, California 90095, USA.
Rita M. Cantor§, Stanley F. Nelson

Program in Neurogenetics, Department of Neurology and Center for Autism Research and Treatment, Semel Institute, David Geffen School of Medicine at UCLA.
Daniel H. Geschwind§, Brett S. Abrahams

Child Study Centre, Yale University, New Haven, Connecticut 06520, USA.
Fred Volkmar

Department of Psychiatry, University of Miami School of Medicine, Miami, FL 33136, USA
Susan E. Folstein

The John P. Hussman Institute for Human Genomics, University of Miami School of Medicine, Miami, Florida 33136, USA.
Margaret A. Pericak-Vance§, Michael L. Cuccaro, John Gilbert

Institute for Juvenile Research, Department of Psychiatry, University of Illinois at Chicago, Chicago, Illinois 60612, USA.
Edwin H. Cook§, Stephen J. Guter, Suma Jacob

Department of Psychiatry, Indiana University School of Medicine, Indianapolis, Indiana 46202, USA.
John I. Nurnberger Jr§, Christopher J. McDougle, David J. Posey

Autism and Communicative Disorders Centre, University of Michigan, Ann Arbor, Michigan 48109-2054, USA.
Catherine Lord, Christina Corsello, Vanessa Hus

Departments of Genetics and Genomic Sciences and Neuroscience, Mount Sinai School of Medicine, New York 10029, USA
Joseph D. Buxbaum§

The Seaver Autism Center for Research and Treatment and Department of Psychiatry, Mount Sinai School of Medicine, New York 10029, USA.
Joseph D. Buxbaum§, Alexander Kolevzon, Latha Soorya, Elena Parkhomenka

Nathan Kline Institute for Psychiatric Research (NKI), 140 Old Orangeburg Road, Orangeburg, New York 10962, USA & Department of Child and Adolescent Psychiatry, New York University and NYU Child Study Center, 550 First Avenue, New York, New York 10016, USA.
Bennett L. Leventhal

Autism Speaks, New York 10016, USA & Department of Psychiatry, University of North Carolina, Chapel Hill, North Carolina 27599-3366, USA.
Geraldine Dawson

Battelle Center for Mathematical Medicine, The Research Institute at Nationwide Children's Hospital and The Ohio State University, Columbus, Ohio 43205, USA.
Veronica J. Vieland§

Department of Pediatrics, Children's Hospital of Philadelphia, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA.
Hakon Hakonarson§

The Center for Applied Genomics, Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA.
Joseph T. Glessner, Cecilia Kim, Kai Wang, Hakon Hakonarson§

Pathology and Laboratory Medicine, University of Pennsylvania, Pennsylvania 19104, USA.
Gerard D. Schellenberg§

Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15213, USA
Bernie Devlin§, Lamburtus Klei

Departments of Psychiatry and Neurology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15213, USA
Nancy Minshew

Department of Molecular Physiology and Biophysics, and Center for Human Genetics Research, Vanderbilt University School of Medicine, Nashville, Tennessee, 37232 USA
James S. Sutcliffe§, Jonathan L. Haines§, Sabata C. Lund, Susanne Thomson, Brian L. Yaspan

Psychiatry Department, University of Utah Medical School, Salt Lake City, Utah 84108, USA.
Hilary Coon§, Judith Miller, William M. McMahon

Department of Psychiatry and Behavioural Sciences, University of Washington, Seattle, Washington 98195, USA.
Jeff Munson

Department of Speech and Hearing Sciences, University of Washington, Seattle, Washington 98195, USA.
Annette Estes

Departments of Biostatistics and Medicine, University of Washington, Seattle, Washington 98195, USA
.
Ellen M. Wijsman§