Human Molecular Genetics, 2012, Vol. 21, No. 21 ***4781–4792***
doi:10.1093/hmg/dds301
Advance Access published on July 26, 2012

Individual common variants exert weak effects on the risk for autism spectrum disorders

Richard Anney[1,†], Lambertus Klei[2], Dalila Pinto[3,†,‡], Joana Almeida[4], Elena Bacchelli[5], Gillian Baird[6], Nadia Bolshakova[1], Sven Bölte[7], Patrick F. Bolton[8], Thomas Bourgeron[10,11], Sean Brennan[1], Jessica Brian[12], Jillian Casey[13], Judith Conroy[13], Catarina Correia[14,15], Christina Corsello[16], Emily L. Crawford[17], Maretha de Jonge[19], Richard Delorme[20], Eftichia Duketis[7], Frederico Duque[4], Annette Estes[21], Penny Farrar[25], Bridget A. Fernandez[26], Susan E. Folstein[27], Eric Fombonne[28], John Gilbert[29], Christopher Gillberg[30], Joseph T. Glessner[31], Andrew Green[13], Jonathan Green[32], Stephen J. Guter[33], Elizabeth A. Heron[1], Richard Holt[25], Jennifer L. Howe[3], Gillian Hughes[1], Vanessa Hus[16], Roberta Igliozzi[15], Suma Jacob[33], Graham P. Kenny[1], Cecilia Kim[31], Alexander Kolevzon[34], Vlad Kustanovich[35], Clara M. Lajonchere[35], Janine A. Lamb[36], Miriam Law-Smith[1], Marion Leboyer[38], Ann Le Couteur[39], Bennett L. Leventhal[40,41], Xiao-Qing Liu[42], Frances Lombard[1], Catherine Lord[43], Linda Lotspeich[44], Sabata C. Lund[17], Tiago R. Magalhaes[14,15], Carine Mantoulan[45], Christopher J. McDougle[46,¶], Nadine M. Melhem[2], Alison Merikangas[1], Nancy J. Minshew[2,47], Ghazala K. Mirza[25], Jeff Munson[22], Carolyn Noakes[12], Gudrun Nygren[30], Katerina Papanikolaou[48], Alistair T. Pagnamenta[25], Barbara Parrini[49], Tara Paton[3], Andrew Pickles[37], David J. Posey[46], Fritz Poustka[7], Jiannis Ragoussis[25], Regina Regan[13], Wendy Roberts[12], Kathryn Roeder[50], Bernadette Roge[45], Michael L. Rutter[9], Sabine Schlitt[7], Naisha Shah[13], Val C. Sheffield[51], Latha Soorya[34], Inês Sousa[25], Vera Stoppioni[52], Nuala Sykes[25], Raffaella Tancredi[49], Ann P. Thompson[53], Susanne Thomson[17], Ana Tryfon[34], John Tsiantis[48], Herman Van Engeland[19], John B. Vincent[54], Fred Volkmar[55], JAS Vorstman[56], Simon Wallace[67], Kirsty Wing[25], Kerstin Wittemeyer[67], Shawn Wood[2], Danielle Zurawiecki[34], Lonnie Zwaigenbaum[57], Anthony J. Bailey[58,‖], Agatino Battaglia[49,‖], Rita M. Cantor[59,‖], Hilary Coon[61,‖], Michael L. Cuccaro[29], Geraldine Dawson[62,63,‖], Sean Ennis[13,‖], Christine M. Freitag[7,‖], Daniel H. Geschwind[60,‖], Jonathan L. Haines[18,‖], Sabine M. Klauck[65,‖], William M. McMahon[61,‖], Elena Maestrini[5,‖], Judith Miller[61,§], Anthony P. Monaco[25,66,‖], Stanley F. Nelson[59,‖], John I. Nurnberger Jr.[46,‖], Guiomar Oliveira[4,‖], Jeremy R. Parr[39,‖], Margaret A. Pericak-Vance[29,‖], Joseph Piven[64,‖], Gerard D. Schellenberg[68,‖], Stephen W. Scherer[3,‖], Astrid M. Vicente[14,15,‖], Thomas H. Wassink[70,‖], Ellen M. Wijsman[23,24,‖],

*To whom correspondence should be addressed at: Department of Psychiatry, University of Pittsburgh School of Medicine, 3811 O'Hara St., Pittsburgh, PA 15213. Tel: +1 4122466642; Fax: +1 4122466640; Email: devlinbj@upmc.edu
†Writing Committee.
‡Present address: The Seaver Autism Center for Research and Treatment, Department of Psychiatry and The Friedman Brain Institute, Mount Sinai School of Medicine, New York NY 10029, USA.
¶Present address: Lurie Center for Autism, Department of Psychiatry, Massachusetts General Hospital, Harvard Medical School, Lexington, MA 0242 USA.
‖Lead AGP investigators.
§Present address: Center for Autism Research, Children s Hospital of Philadelphia, Philadelphia, PA 19104, USA.
#Genotype production.

© *The Author 2012. Published by Oxford University Press.*
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.5), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Catalina Betancur[71,72,73†‖], Joseph D. Buxbaum[34†‖], Edwin H. Cook[33†‖], Louise Gallagher[1†‖], Michael Gill[1†‖], Joachim Hallmayer[44†‖], Andrew D. Paterson[3†‖], James S. Sutcliffe[17†‖], Peter Szatmari[53†‖], Veronica J. Vieland[74†‖], Hakon Hakonarson[31,69‖#] and Bernie Devlin[2*†‖]

[1]Autism Genetics Group, Department of Psychiatry, School of Medicine, Trinity College, Dublin 8, Ireland, [2]Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA, 15232, USA, [3]The Centre for Applied Genomics and Program in Genetics and Genomic Biology, The Hospital for Sick Children and Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada M5G 1L7, [4]Hospital Pediátrico de Coimbra, 3000–076 Coimbra, Portugal, [5]Department of Biology, University of Bologna, 40126 Bologna, Italy, [6]Guy's and St Thomas' NHS Trust & King's College, London SE1 9RT, UK, [7]Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, J.W. Goethe University Frankfurt, 60528 Frankfurt, Germany, [8]Department of Psychiatry, [9]MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King's College, London SE5 8AF, UK, [10]Human Genetics and Cognitive Functions, Institut Pasteur and [11]University Paris Diderot-Paris 7, CNRS URA 2182, Fondation FondaMental, 75015 Paris, France, [12]Autism Research Unit, The Hospital for Sick Children and Bloorview Kids Rehabilitation, University of Toronto, Toronto, ON, Canada M5G 1Z8, [13]School of Medicine, Medical Science University College, Dublin 4, Ireland, [14]Instituto Nacional de Saude Dr Ricardo Jorge and Instituto Gulbenkian de Cîencia, 1649-016 Lisbon, Portugal, [15]BioFIG—Center for Biodiversity, Functional and Integrative Genomics, Campus da FCUL, C2.2.12, Campo Grande, 1749-016 Lisboa, Portugal, [16]Department of Psychology, University of Michigan, Ann Arbor, MI 48109, USA, [17]Department of Molecular Physiology and Biophysics, Vanderbilt Kennedy Center, and Centers for Human Genetics Research and Molecular Neuroscience and [18]Center for Human Genetics Research, Vanderbilt University Medical Centre, Nashville, TN 37232, USA, [19]Department of Child Psychiatry, University Medical Center, Utrecht, 3508 GA, The Netherlands, [20]Child and Adolescent Psychiatry, APHP, Hôpital Robert Debré, 75019 Paris, France, [21]Department of Speech and Hearing Sciences, [22]Department of Psychiatry and Behavioral Sciences, [23]Department of Biostatistics and [24]Department of Medicine, University of Washington, Seattle, WA 98195, USA, [25]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK, [26]Disciplines of Genetics and Medicine, Memorial University of Newfoundland, St John's, NL, Canada A1B 3V6, [27]Department of Psychiatry, University of Miami School of Medicine, Miami, FL 33136, USA, [28]Division of Psychiatry, McGill University, Montreal, QC, Canada H3A 1A1, [29]The John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL 33101, USA, [30]Gillberg Neuropsychiatry Centre, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden, [31]The Center for Applied Genomics, Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA, [32]Academic Department of Child Psychiatry, University of Manchester, Manchester M9 7AA, UK, [33]Department of Psychiatry, Institute for Juvenile Research, University of Illinois at Chicago, Chicago, IL 60608, USA, [34]The Seaver Autism Center for Research and Treatment, Department of Psychiatry, The Friedman Brain Institute, Mount Sinai School of Medicine, New York NY 10029, USA, [35]Autism Genetic Resource Exchange, Autism Speaks, Los Angeles, CA 90036-4234, USA, [36]Centre for Integrated Genomic Medical Research and [37]Department of Medicine, School of Epidemiology and Health Science, University of Manchester, Manchester M13 9PT, UK, [38]Department of Psychiatry, Groupe hospitalier Henri Mondor-Albert Chenevier, INSERM U995, AP-HP; University Paris 12, Fondation FondaMental, Créteil 94000, France, [39]Institutes of Neuroscience and Health and Society, Newcastle University, Newcastle upon Tyne NE1 7RU, UK, [40]Nathan Kline Institute for Psychiatric Research (NKI), 140 Old Orangeburg Road, Orangeburg, NY 10962, USA, [41]Department of Child and Adolescent Psychiatry, New York University, NYU Child Study Center, New York, NY 10016, USA, [42]Department of Obstetrics, Gynecology, and Reproductive Sciences, University of Manitoba, Winnipeg, MB, Canada, [43]Center for Autism and the Developing Brain, Weill Cornell Medical College, White Plains, NY, USA, [44]Department of Psychiatry, Division of Child and Adolescent Psychiatry and Child Development, Stanford University School of Medicine, Stanford, CA 94304, USA, [45]Centre d'Eudes et de Recherches en Psychopathologie, University de Toulouse Le Mirail, Toulouse 31200, France, [46]Department of Psychiatry, Indiana University School of Medicine, Indianapolis, IN 46202, USA, [47]Department of Neurology, University of Pittsburgh School of Medicine, Pittsburgh, PA 15213, USA [48]University Department of Child Psychiatry, Athens University, Medical School, Agia Sophia Children's Hospital, 115 27 Athens, Greece, [49]Stella Maris Institute for Child and Adolescent Neuropsychiatry, 56128 Calambrone (Pisa), Italy, [50]Department of Statistics, Carnegie Mellon University, Pittsburgh, PA, USA, [51]Department of Pediatrics and Howard Hughes Medical Institute

Carver College of Medicine, University of Iowa, Iowa City, IA 52242, USA, [52]Neuropsichiatria Infantile, Ospedale Santa Croce, 61032 Fano, Italy, [53]Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton, ON, Canada L8N 3Z5, [54]Centre for Addiction and Mental Health, Clarke Institute and Department of Psychiatry, University of Toronto, Toronto, ON, Canada M5G 1X8, [55]Child Study Centre, Yale University, New Haven, CT 06520, USA, [56]Department of Psychiatry, Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, Utrecht 3584 CX, The Netherlands, [57]Department of Pediatrics, University of Alberta, Edmonton, AB, Canada T6G 2J3, [58]BC Mental Health and Addictions Research Unit, University of British Columbia, Vancouver, BC, Canada V5Z4H4, [59]Department of Human Genetics and [60]Department of Neurology, Los Angeles School of Medicine, University of California, Los Angeles, CA 90095, USA, [61]Psychiatry Department, University of Utah Medical School, Salt Lake City, UT 84108, USA, [62]Autism Speaks, New York, NY, USA, [63]Department of Psychiatry and [64]Carolina Institute for Developmental Disabilities, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3366, USA, [65]Division of Molecular Genome Analysis, German Cancer Research Center (DKFZ), Heidelberg 69120, Germany, [66]Office of the President, Tufts University, Boston, MA, USA, [67]Department of Psychiatry, University of Oxford, Warneford Hospital, Headington, Oxford, OX3 7JX, UK, [68]Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA, [69]Department of Pediatrics, Children's Hospital of Philadelphia, University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA, [70]Department of Psychiatry, Carver College of Medicine, Iowa City, IA 52242, USA, [71]INSERM U952, [72]CNRS UMR 7224 and [73]UPMC Univ Paris 06, Paris 75005, France and [74]Battelle Center for Mathematical Medicine, The Research Institute at Nationwide Children's Hospital and The Ohio State University, Columbus, OH 43205, USA

Received March 22, 2012; Revised July 13, 2012; Accepted July 19, 2012

While it is apparent that rare variation can play an important role in the genetic architecture of autism spectrum disorders (ASDs), the contribution of common variation to the risk of developing ASD is less clear. To produce a more comprehensive picture, we report Stage 2 of the Autism Genome Project genome-wide association study, adding 1301 ASD families and bringing the total to 2705 families analysed (Stages 1 and 2). In addition to evaluating the association of individual single nucleotide polymorphisms (SNPs), we also sought evidence that common variants, *en masse*, might affect the risk. Despite genotyping over a million SNPs covering the genome, no single SNP shows significant association with ASD or selected phenotypes at a genome-wide level. The SNP that achieves the smallest *P*-value from secondary analyses is rs1718101. It falls in *CNTNAP2*, a gene previously implicated in susceptibility for ASD. This SNP also shows modest association with age of word/phrase acquisition in ASD subjects, of interest because features of language development are also associated with other variation in *CNTNAP2*. In contrast, allele scores derived from the transmission of common alleles to Stage 1 cases significantly predict case status in the independent Stage 2 sample. Despite being significant, the variance explained by these allele scores was small (Vm< 1%). Based on results from individual SNPs and their *en masse* effect on risk, as inferred from the allele score results, it is reasonable to conclude that common variants affect the risk for ASD but their individual effects are modest.

INTRODUCTION

Genetic analyses over the past two decades have connected a plethora of rare mutations with the risk for autism spectrum disorder (ASD) (1–9). The rate of discovery has accelerated remarkably with the use of genome-wide screening tools capable of identifying sub-microscopic structural DNA variation (10). Many of the identified risk variants occur *de novo* and implicate many genes. From the distribution of *de novo* copy number variants (CNVs) in affected individuals and their unaffected siblings, Sanders *et al*. (11) conclude that there are hundreds of ASD risk loci in the human genome. This magnitude for risk genes is also supported by the distribution of *de novo* sequence events in ASD probands (12–15). Indeed, over the past year evidence supporting rare variation affecting risk, both *de novo* and inherited, continues to accumulate through both copy number and sequence-based studies (11–20).

In contrast, genome-wide evidence for common variation affecting the risk for ASD is limited. In a recent study by this group, the Autism Genome Project (AGP), we reported on a genome-wide association study (GWAS) of almost 1400 strictly defined ASD families genotyped for one million single nucleotide polymorphisms (SNPs) (21) (herein referred to as Stage 1). We performed four primary association

analyses based on ancestry and clinical thresholds and identified a single, marginally significant association for SNP rs4141463 located within *MACROD2*. Other large GWAS also highlighted significant associations at single loci, namely SNPs at 5p14.1 (22) and 5p15.2 (23). In our previous Stage 1 GWAS, we did not find evidence to support either of these earlier associations (21). Likewise, a subsequent study in an independent collection of 1170 individuals of European ancestry with ASD and 35307 non-ASD controls finds no support for the *MACROD2* marker rs4141463 (24).

The objective of the AGP has been to characterize the genetic architecture of ASD, and therefore we sought to determine whether both common variants and rare CNVs account for risk and, if so, by how much. Thus far, we can conclude that rare variation plays a substantial role (9,25), complementing many other studies (26,27). What we can say about common variation is less definitive. Based on the analysis of published GWAS, Devlin *et al*. (28) concluded that common genomic variants having a substantial impact on risk (relative risk >1.5) do not exist; there could be a small number of common risk variants, as yet unidentified, having an intermediate impact on risk (relative risk 1.2–1.5); yet there could be many common variants of modest impact (relative risk <1.2). To consistently and reliably detect common variants of modest effect will require samples in the tens of thousands (29).

To produce a more comprehensive picture of how common variants affect the risk for ASD, we report the second stage of the AGP GWAS, amounting to an additional 1301 ASD families as a combined mega-analysis of Stage 1 and Stage 2. In addition to the single SNP analyses, we use the two-stage design to evaluate the SNPs, *en masse*, for evidence that a large number of common variants exert at least weak effects on risk. The allele-score method, as previously described by Purcell *et al*. (30), calculates an accumulated score for each individual based on the presence and count of associated alleles (for family-based association, the count is the over-transmitted allele). With their data, the International Schizophrenia Consortium found that those common variants assessed (M) could explain up to 3.18% of the additive genetic variance (V_M) for schizophrenia (30). To estimate how much of the variance for ASD can be explained by common variants, the allele score derived from the Stage 1 data (21) was used to predict case status in the Stage 2 data and calculate the V_M.

RESULTS

Properties of the Stage 1 sample are reported elsewhere (21). When combined with the Stage 2 data, we report the largest family-based sample of ASD to date (Table 1). Ancestry analysis, estimated from numerous genetic markers by using standard techniques (31–33) and mapped back onto European recruitment sites to anchor the common ancestry of subjects, showed that 94.9% of the probands are of European ancestry (Supplementary Material, Fig. S1). The primary analyses target four overlapping data sets, based on diagnosis (*spectrum* versus *strict*) and ancestry (only European versus any ancestry). Of these four samples, the most inclusive allows for a *spectrum* diagnosis for subjects of any ancestry (Table 1), whereas the least inclusive allows for only a *strict* diagnosis of autism for subjects of European ancestry (see Supplementary Material, Fig. S2 for experimental design).

Association analyses of these four classes do not identify any single SNP to be significantly associated with autism as judged by an accepted threshold for genome-wide significance, $P < 5 \times 10^{-8}$. Instead the distributions of test statistics are consistent with that expected by chance (Supplementary Material, Fig. S3). Only four genomic regions show association signals at $P < 10^{-6}$, two falling in genes (Table 2; Supplementary Material, Figs S4 and 5). When we evaluated the corresponding Stage 2 results with SNPs highlighted by our Stage 1 study (21), only one shows evidence of a modest association, namely rs4150167, which is a non-synonymous variant in *TAF1C* (Spectrum, All ancestries; rs4150167-A; Stage 1 OR = 0.37, $P = 7.87 \times 10^{-7}$; Stage 2 OR = 0.65, $P = 0.014$).

Exploratory analyses also produce no significant findings after adjusting for multiple testing (Table 3). A detailed summary of these findings is given in the Supplementary Material, Table S1; Fig. S5). Results from Stage 2 analysis alone are presented in Supplementary Material, Table S2, but are not discussed in the main manuscript. Of the promising results from other GWAS studies (22,23,34), only rs1328244 (13q33.3) and rs6646569 (Xq24) described by Wang *et al*. (22) garner support from the combined AGP sample, although it is important to bear in mind that the studies overlap somewhat in samples/families and thus correlated statistics are expected (Table 4).

Under the hypothesis that individual common variants exert only a limited effect on risk and that many such variants are present and exert these effects independently, we would expect these effects to be detectable in an analysis of common variants *en masse*. To address this question, allele scores were generated for all AGP probands based on the four Stage 1 primary analyses with the goal of determining whether the score derived from the Stage 1 results predicted the Stage 2 case versus control status. Allele scores based on markers associated at 10 significance thresholds ranging from $P < 0.5$ to $P < 0.00001$ were evaluated. Used as a positive control for the method, the Stage 1 scores showed high to perfect predictive value for case status in the Stage 1 subjects (data not shown). When examined against Stage 2 individuals, the Stage 1 scores were significant predictors of case status (Fig. 1) and thus explain a significant portion of variance in case and control status of Stage 2 samples. In general, the variance explained increases with an increased number of markers in the model. Still the markers explain only a small proportion of the variance—always $<1\%$, with the greatest signal observed in the smallest yet most homogeneous group, namely European ancestry individuals with a *Strict* diagnosis of autism (Vm = 0.78%; Empirical $P < 0.001$). Analyses stratifying by quintile of minor allele frequency showed that most of the variance explained accrues to the quintile 0.2–0.3 (Supplementary Material, Fig. S6). Although these results were noisy (Supplementary Material, Fig. S6), they suggest that many common variants affect risk.

Importantly the allele-score analysis assumes that the distribution of liability for pseudo controls is similar to that of the general population. Various factors could skew this

Table 1. The sample size for specific family-based analyses

Analysis	Stage 2				Combined Stages 1 and 2			
	No. of families	Cases per family[a]			No. of families	Cases per family		
		1	2	3		1	2	3
Spectrum	1301	1289	11	1	2705	2678	26	1
Strict	785	782	3	0	1673	1667	6	0
Spectrum/European	1108	1096	11	1	2384	2359	24	1
Strict/European	660	657	3	0	1458	1452	6	0
Verbal	951	947	3	1	1884	1868	15	1
Non-verbal	333	331	2	0	799	796	3	0
Verbal/European	830	826	3	1	1686	1672	13	1
Non-verbal/European	266	264	2	0	682	679	3	0
Upper IQ	594	594	0	0	1191	1187	4	0
Lower IQ	307	307	0	0	672	670	2	0
Upper IQ/European	528	528	0	0	1067	1064	3	0
Lower IQ/European	221	221	0	0	551	549	2	0
Word	1068	1061	6	1	2241	2222	18	1
Phrase	946	941	4	1	1893	1877	15	1

[a]All families were required to have maternal and paternal genotypes. In addition, the sample size listed here are the families and samples that passed quality control criteria.

Table 2. Top hits from the primary analyses in the combined AGP GWA sample

Chr	Position	SNP	Gene	Diagnosis	Ancestry	*P*	OR	OR1	OR2
2	205794415	rs4675502	*PARD3B*	Strict	All	4.340e-07	1.28 (1.16–1.41)	1.29	1.27
5	162016095	rs7711337	NONE	Spectrum	All	8.254e-07	0.82 (0.76–0.89)	0.83	0.81
8	72724295	rs7834018	NONE	Strict	European	7.538e-07	0.64 (0.53–0.77)	0.63	0.65
16	82771184	rs4150167	*TAF1C*	Spectrum	All	2.910e-07	0.51 (0.39–0.66)	0.37	0.65

The SNPs listed are restricted to top associated SNP if more than one SNP within a linkage disequilibrium (LD) block exists; for additional associated SNPs in LD with the index SNP, see Supplementary Material. OR1, odds ratio from Stage 1 data; OR2, odds ratio for Stage 2 data; OR is the overall odds ratio (95% confidence interval); *P*, the *P*-value for OR. The distribution of the association signals around these SNPs for these ancestral and diagnostic categories is shown in Supplementary Material, Figure S4. Graphical representation of the association signal for these highlighted SNPs across stages and diagnostic groups is plotted in Supplementary Material, Figure S5.

distribution and thus bias downwards the estimated Vm, a notable one being an excess of multiplex families relative to population frequencies (35,36). Multiplex families comprise at least 33% of the families studied here (47.5% simplex, 19.5% unknown), and thus could substantially diminish the estimated Vm. In addition, as pointed out by Ripke *et al*. (37), the use of a larger sample size to develop the predictor (i.e. model from the Stage 1 sample) should produce a more accurate allele score and thus increase the amount of variance in the Stage 2 case and control status explained.

DISCUSSION

For this report the AGP genotyped 1301 additional families for almost one million SNPs, bringing the total analysed herein to 2678 families. From this large data set, we observed no genome-wide significant association with specific common variants. Four independent signals in the primary analyses yielded uncorrected *P*-values $<1 \times 10^{-6}$ and 23 independent signals in the exploratory analyses crossed this threshold (Tables 2 and 3). Some of these SNPs fall in intriguing genes, including *TAF1C* from the primary analyses and *CNTNAP2* from the exploratory analyses. *TAF1C* encodes the TATA box-binding protein-associated factor RNA polymerase I subunit C. We previously highlighted and discussed *TAF1C* and the corresponding SNP (rs4150167) (21). Of interest is that the common allele for rs4150167 was over-transmitted and is represented on roughly 98% of the chromosomes. The same pattern, although not as extreme, is present in the Stage 2 sample and again cannot be traced to a genotyping error (Supplementary Material, Fig. S7). If this were a true causal effect, it suggests that the minor allele at rs4150167 is somehow protective. The protective allele results in a non-synonymous glycine to arginine change at amino acid position 523 and has also been implicated in alternate splicing (38), both of which offer a starting point for investigation of the impact of this variation on gene function. On the other hand the association should be viewed cautiously because the power of this study to detect a protective effect is small. Statistically, therefore, the posterior probability that this observation is a false positive is large.

The strongest uncorrected association observed across all analyses was for rs1718101 in European individuals with

Table 3. The most associated SNPs from the secondary analyses in the combined sample

Chr	Position	SNP	Gene	Diagnosis/subgroup	Inheritance	*P*	OR	OR1	OR2
1	116327194	rs12726299	*SLC22A15*	Verb[a]/Eur[b]	Paternal	1.496e−07	1.51 (1.29–1.76)	1.61	1.43
1	204764871	rs11118968	*RASSF5*	Spect[c]/All	Maternal	2.452e−07	0.44 (0.32–0.61)	0.43	0.45
2	230226476	rs6752370	*DNER*	Strict/All	Maternal	8.526e−07	1.62 (1.33–1.96)	1.56	1.67
2	31000000	rs10205350	*GALNT14*	Nverb[d]/All	Maternal	3.942e−07	0.50 (0.38–0.66)	0.57	0.40
3	185006350	rs263035	*YEATS2*	Spect/All	Paternal	2.258e−07	1.39 (1.22–1.57)	1.29	1.49
4	147068744	rs12331851	*ZNF827*	Verb/Eur	Paternal	6.081e−07	0.38 (0.26–0.57)	0.49	0.28
5	2162900	rs6879627	NONE	IQ+[e]/Eur	Maternal	3.992e−07	0.52 (0.40–0.67)	0.63	0.43
5	65575602	rs29456	NONE	Strict/All	Maternal	1.226e−07	1.65 (1.37–1.99)	1.54	1.75
7	145753720	rs1718101	*CNTNAP2*	IQ+/Eur	Additive	7.783e−09	2.13 (1.63–2.80)	2.16	2.11
10	119652081	rs1936295	NONE	Spect/Eur	Maternal	6.636e−07	1.69 (1.37–2.09)	1.70	1.66
10	55971935	rs1930165	*PCDH15*	IQ+/All	Maternal	9.861e−08	0.60 (0.50–0.73)	0.57	0.63
11	106363884	rs11211996	*GUCY1A2*	Nverb/Eur	Paternal	4.287e−07	2.11 (1.56–2.85)	1.67	3.00
11	113578084	rs3782000	*ZBTB16*	IQ+/All	Additive	1.842e−07	1.41 (1.24–1.61)	1.37	1.45
16	4941380	rs9635542	NONE	Nverb/All	Paternal	3.268e−07	0.44 (0.32–0.61)	0.39	0.52
16	61230112	rs288604	NONE	Spect/All	Maternal	2.975e−07	1.58 (1.32–1.88)	1.59	1.57
17	68575820	rs9302952	*SLC39A11*	Verb/All	Maternal	4.039e−07	1.64 (1.35–1.99)	1.77	1.54
18	34259508	rs932026	NONE	IQ+/Eur	Maternal	1.473e−07	0.44 (0.32–0.60)	0.35	0.52
18	66531220	rs17083037	NONE	IQ+/All	Additive	1.368e−07	0.47 (0.35–0.63)	0.51	0.43
18	7024945	rs600695	*LAMA1*	IQ+/All	Paternal	3.578e−07	1.61 (1.34–1.94)	1.58	1.64
20	14810971	rs6110458	*MACROD2*	Spect/All	Maternal	1.806e−07	1.46 (1.27–1.69)	1.56	1.39
20	14813155	rs14135	*CR596518/MACROD2*	Spect/All	Maternal	1.778e−07	1.49 (1.28–1.74)	1.61	1.40
20	14815421	rs1475531	*CR596518/MACROD2*	Strict/All	Maternal	2.011e−07	1.53 (1.30–1.79)	1.85	1.28
21	38787019	rs2836439	*ERG*	IQ-[f]/Euro	Paternal	6.638e−07	0.42 (0.30–0.60)	0.39	0.47

SNPs listed are restricted to top associated SNP if more than one SNP within a linkage disequilibrium (LD) block exists; for additional associated SNPs in LD with the index SNP, see Supplementary Material. Note that *P*-values are not corrected for multiple testing and none of the results reported here should be considered significant. OR1, odds ratio from Stage 1 data; OR2, odds ratio for Stage 2 data; OR is the overall odds ratio (95% confidence interval); *P*, the *P*-value for OR.
[a]Verbal.
[b]European.
[c]Spectrum.
[d]Non-verbal.
[e]Higher IQ.
[f]Lower IQ.

Table 4. Top hits from regions previously implicated as autism susceptibility regions from non-AGP studies

Chr	Position	SNP	Original *P*-value	Reference	*P*-value	Best *P* ($\pm$ 10 kb)
5	9677904	rs10513025[a]	2.1×10^{-7}	(22)	0.546	rs41457; $P = 0.0569$
5	26003459	rs4307059	2.1×10^{-10}	(21)[b]	0.2033	rs7732252; $P = 0.0915$
13	108881898	rs1328244	8.2×10^{-7}	(21)[b]	0.0094	rs1328244; $P = 0.0094$
14	39901753	rs7147817	2.7×10^{-7}	(21)[b]	0.5591	rs4903707; $P = 0.2756$
16	19116069	rs9932538	1.9×10^{-7}	(21)[b]	0.2459	rs7188617; $P = 0.2028$
X	4940800	rs11798405	9.0×10^{-7}	(21)[b]	0.5664	rs5961595; $P = 0.0406$
X	32391976	rs5972577	2.7×10^{-7}	(21)[b]	0.695	rs6653872; $P = 0.1086$
X	119133917	rs6646569	9.7×10^{-7}	(21)[b]	0.0022	rs6646569; $P = 0.0022$

[a]Used proxy rs2234235 which is in perfect or almost perfect LD with rs10513025.
[b]In the Stage 1 sample reported in Anney *et al*. (21) 111 AGRE families overlapped with Wang *et al*. (22) or 8.1%; in the entire AGP sample the overlap is more substantial, 858 spectrum subjects (analysed in a case–control setting) or 31%.

higher IQ (*P*-value $= 7.8 \times 10^{-9}$; OR = 2.13). This SNP is not in substantial linkage disequilibrium (LD) with other genotyped SNPs and thus no other SNP genotyped by this study supports its association. Examination of the genotype intensity plots, however, showed good quality clustering (Supplementary Material, Fig. S7). The SNP resides within *CNTNAP2* (7q35), which encodes the Contactin-associated protein-like 2, a member of the neurexin-family thought to play a role in axonal differentiation and guidance. *CNTNAP2* is one of the largest genes in the human genome encompassing ~1.5% of chromosome 7 and has previously been implicated in ASD, in part because it resides within regions of suggestive linkage (39). CNVs falling in other neurexin genes including *NRXN1* have been implicated in ASD (4,8,9,25,40,41). More directly, rare variation in *CNTNAP2*, including CNVs, other structural disruptions and deleterious sequence variation, has been identified in subjects with autism, epilepsy, intellectual disability, Pitt-Hopkins-like syndrome, schizophrenia and Tourette syndrome (42–48), albeit with differing degrees of evidence for effect on these traits. Although the most compelling

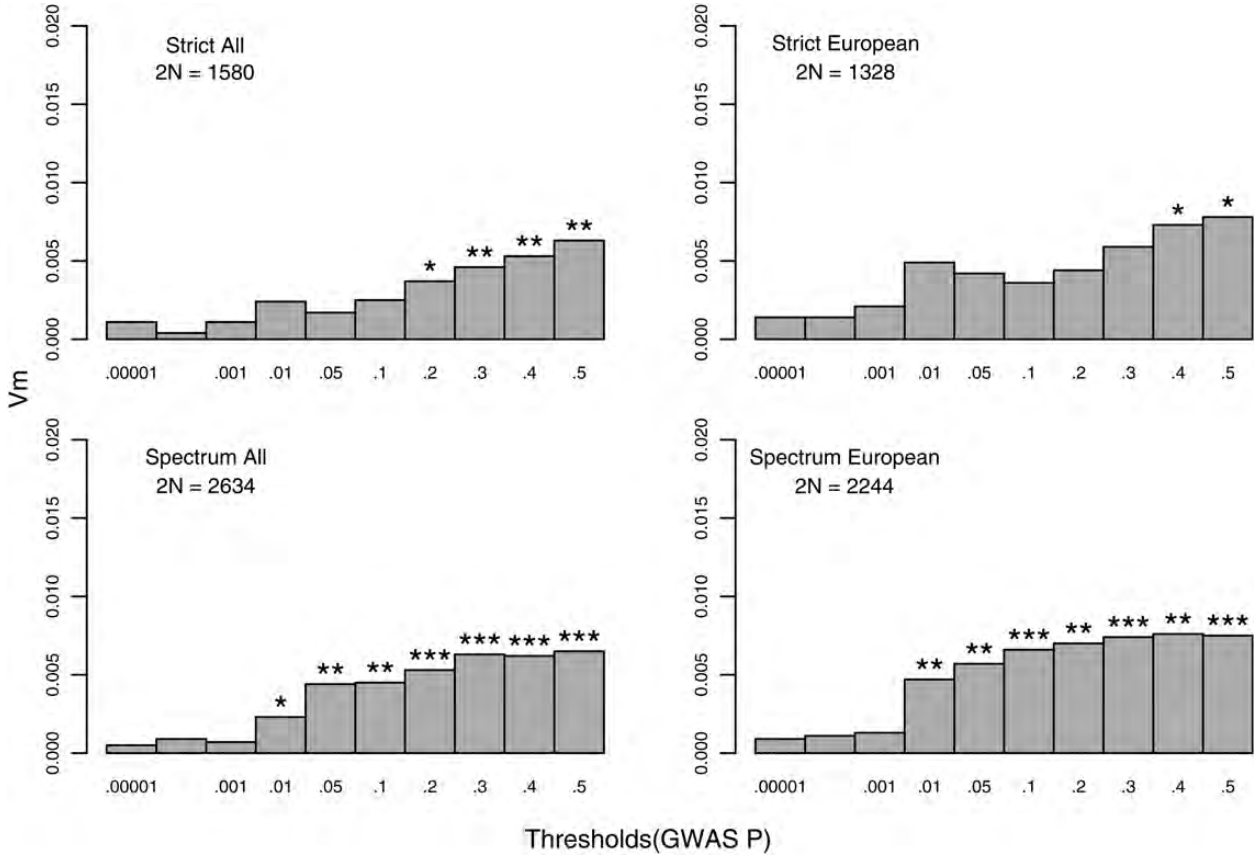


Figure 1. Additive genetic variance at marker (VM) explained by the Stage 1-derived allele score in Stage 2 probands and pseudo controls for a given ancestry and diagnostic classification. Significant association of the allele score denoted by ($^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$). Note that the number of predictors, at a given threshold, is roughly the threshold times the number of SNPs meeting quality control criteria (947 233).

evidence is for recessively inherited Pitt-Hopkins-like syndrome, the other studies are more consistent with dominant or additive effects. Common variants in *CNTNAP2* have also been associated with ASD (49–51) and specific-language impairment (52). In light of the literature, *post hoc* we identified the common SNPs reported in the early studies (49,50,52) to be associated with the risk for ASD or language impairment (Table 5) and evaluated their association to ASD and age-at-first-word and -phrase for all three inheritance patterns. Of the three SNPs reported in the literature, one (rs2710102) shows very modest association with the risk for ASD. For the language outcomes, rs1718101 shows modest, significant association with age-at-first-phrase, acting either additively or dominantly but not recessively; rs17236239 shows modest association with age-at-first-phrase under a recessive model (Table 5).

In our previous study, we highlighted association with SNPs within *MACROD2*. The most highly associated *MACROD2* SNP from primary Stage 1 analyses, rs4141463 (P-value = 2.40×10^{-8}; OR = 0.65), obtained by analyses of subjects with a strict diagnosis of autism and of European ancestry, showed little if any signal in the Stage 2 sample (P-value = 0.206, OR = 0.91). Thus, for the mega-analysis reported here, the association is less compelling (P-value = 1.2×10^{-6}; OR = 0.77) than previously reported (21). Within this large gene three SNPs show substantial but still non-significant over-transmission of specific alleles from mothers (Table 3). The region encompassed by these three SNPs captures the 3′ region of a putative antisense RNA, specifically *CR596518*, and one SNP falls in its translated region (rs14135). Because antisense RNA can play a role in the mechanics of imprinting (53), this observation is of some note. Nonetheless, without additional biological or statistical evidence, the result is unconvincing. Thus, whether SNPs in *MACROD2* or in its intra-genic antisense RNA genes play a role in the risk for ASD remains an open question. If they do, however, there can now be little doubt that their effect size is modest on the basis of our results and others (24).

That no individual common variants are significantly associated with the risk for ASD in our data was anticipated by earlier analyses (28), based on results from three GWAS studies and statistical theory regarding the relationships among sample size, effect size and power. These analyses predicted that few if any common variants have an impact on risk exceeding 1.2 (or below its inverse). Our results from >2700 trios, together with the results from other published GWAS (22,23,34), bear out this prediction. The analyses laid out in Devlin *et al*. (28) cannot be precise when predicting the number of loci with common variants of modest impact on risk (0.8< risk <1.2); the modelling is consistent with a range of loci from zero to many thousands.

Table 5. Family-based quantitative trait analysis at previously implicated CNTNAP2 SNPs for age at first word and phrase for ASD subjects and three models of inheritance, as well as qualitative analysis of ASD risk

SNP	Genotype (risk[a])	Chr	Position	Minor	Major	MAF	Z(ADD)	p(ADD)	Z(DOM)	p(DOM)	Z(REC)	p(REC)
Age-at-first-word												
rs1718101[f]	R(A)	7	145753721	A	G	0.054	−1.78	*0.074*	−1.51	0.130	−1.72	*0.086*
rs7794745[g]	W(T)[e]	7	146120539	T	A	0.345	1.09	0.276	0.45	0.654	1.36	0.173
rs2710102[h]	R(G)	7	147205323	A	G	0.483	1.49	0.136	0.53	0.595	1.69	*0.091*
rs17236239[i]	R(G)	7	147213238	G	A	0.344	−0.81	0.419	−1.14	0.253	0.25	0.805
Age-at-first-phrase												
rs1718101[f]	R(A)	7	145753721	A	G	0.054	−2.40	**0.017**[c]	−2.42	**0.015**[c]	−0.32	0.752
rs7794745[g]	W(T)[e]	7	146120539	T	A	0.345	0.12	0.907	0.07	0.945	0.12	0.909
rs2710102[h]	R(G)	7	147205323	A	G	0.483	1.08	0.279	0.02	0.984	1.70	*0.089*
rs17236239[i]	R(G)	7	147213238	G	A	0.344	0.37	0.711	−1.17	0.241	2.37	**0.018**[b]

SNP	Genotype (risk[a])	Chr	Position	Minor	Major	MAF	OR (95%CI)	p(ADD)
All ASD								
rs1718101[f]	R(A)	7	145753721	A	G	0.054	1.25 (1.06–1.49)	**0.010**[b]
rs7794745[g]	W(T)[e]	7	146120539	T	A	0.345	0.90 (0.79–1.03)	0.120
rs2710102[h]	R(G)	7	147205323	A	G	0.483	1.08 (1.00–1.16)	**0.045**[b]
rs17236239[i]	R(G)	7	147213238	G	A	0.344	0.95 (0.88–1.03)	0.200
European, higher IQ								
rs1718101[f]	R(A)	7	145753721	A	G	0.054	2.13 (1.63–2.80)	**7.78E-9**[d]
rs7794745[g]	W(T)[e]	7	146120539	T	A	0.345	0.86 (0.71–1.04)	0.116
rs2710102[h]	R(G)	7	147205323	A	G	0.483	1.02 (0.91–1.15)	0.740
rs17236239[i]	R(G)	7	147213238	G	A	0.344	1.05 (0.93–1.19)	0.410

[a]Risk allele reported in initial association study.
[b]Association effect observed in same direction as initial association.
[c]Association effect observed in opposite direction as initial association.
[d]Association is initial study.
[e]Risk allele for ambiguous SNP (A/T) defined according to minor allele frequency.
Studies are [f]this study; [g]Arking *et al*. (50); [h]Alarcon *et al*. (49); [i]Vernes *et al*. (52).
P-values < 0.1 are italicized, those < 0.05 also in bold.

Secondary approaches are being developed and applied to GWAS data to explore further the role of common variation within ASD. The AGP and others have applied diverse analytic approaches to determine whether there is evidence of enriched association in specific genes and groups of genes (54–57), whether there are regions of extended homozygosity within affected individuals that may implicate regions harbouring putatively recessive alleles (58) or to explore more discrete trait-based phenotypes within genome-wide data (59–61). To seek evidence for or against common variants having this modest impact on risk, we approached the problem by constructing an allele score, based on the transmission properties of SNP alleles in Stage 1 data, and asking whether these composite scores from putative risk alleles could predict case and pseudo-control status in Stage 2. In other words, do common variants over-transmitted in Stage 1 also tend to be observed in Stage 2? If so, this may provide evidence that common variants affect the risk for ASD. Moreover, when combined with the published GWAS results and the theory in Devlin *et al*. (28), they illuminate how common variants affect risk: individually they have very small effect, but *en masse* they exert a detectable impact. This logical circle is now closed. We find that allele scores derived in Stage 1 do indeed predict case and pseudo-control status in Stage 2, making the case that common variants affect the risk for ASD. The score cannot account for much of the variance, $<1\%$, and only about a third of that recently explained for schizophrenia (30). Thus, while the existence of common variants affecting the risk of ASD is almost assured, their individual effects are modest and their collective effects could be smaller than that for rare variation.

On the other hand, complementing the allele-score analysis, gene-set enrichment and related analyses for AGP and other ASD data sets have found significant evidence for common variants affecting risk (54–57), suggesting that common variants account for a non-trivial proportion of risk and that many true positive associations of small effect could be buried in the noise of stochastic variation. Qualitatively, recent studies of rare variants also find it challenging to distinguish risk variation from the background stochastic noise (12–15). Given these challenges, it is reasonable to conjecture that even if all of the samples analysed here were also sequenced at the whole genome level, it would still be impossible to discern how much risk accrues from common versus rare variation, at least from current knowledge. Thus a precise estimate of the relative contribution of rare and common variation to risk will require further studies.

MATERIALS AND METHODS

Clinical evaluation

Participants in this AGP study were recruited at centres in North America and Europe. Subjects with known karyotypic

abnormalities and fragile X mutations were typically excluded. Likely affected individuals were assessed using the Autism Diagnostic Interview-Revised (62) and Autism Diagnostic Observation Schedule (63). Cognitive functioning was established for the majority of subjects using a range of cognitive measures from which a categorical classification of intellectual capacity was derived for the analyses, as described below and elsewhere (21).

Genotyping

The Stage 1 sample was genotyped using the Illumina Infinium 1M-single SNP microarray; the Stage 2 sample was genotyped on either the Illumina Infinium 1M-single or the Illumina 1M-duo microarray. All quality control (QC) procedures were consistent across data sets and are described elsewhere (21). For subjects inferred to be of European ancestry, markers showing Fst >0.02 across recruitment sites were excluded from analyses. A total of 947 233 SNPs passed QC for the combined data sets.

Population structure

Ancestry for both Stage 1 and Stage 2 samples was estimated by using 5176 SNPs, which had a genotype completion rate of $\geq$99.9%, and a minor allele frequency MAF ≥ 0.05 and were separated by at least 500 kb. The initial dacGem (33) analysis identified three dimensions of ancestry and separated the data into five ancestry groups (Supplementary Material, Fig. S1). Groups A through C, illustrated in Supplementary Material, Figure S1, delineate 1285, 771 and 486 families of European ancestry as determined by the representation of European recruitment sites. All other families were assigned to be non-European for purposes of association analysis (described below). By using genotypes from the same panel of 5176 SNPs, we searched the data set for duplicate samples between Stages 1 and 2, as well as within the Stages, and found none.

Phenotype treatment

As described in Anney *et al*. (21), our primary analyses grouped families into two nested diagnostic classes: s*trict*, in which affected individuals met both ADI-R and ADOS criteria for autism; and *spectrum*, combining individuals meeting *strict* criteria with individuals with a broader ASD diagnosis as determined by responses to the ADI-R and ADOS. As described above, we also grouped individuals on the basis of ancestry. To evaluate the effect of transmitted alleles on the risk of ASD, the primary analyses assume an additive model and evaluated four overlapping data sets: two levels of diagnosis (*spectrum* versus *strict* diagnosis) by two levels of ancestry (*all ancestries* versus *European ancestry*).

Secondary analyses (Supplementary Material, Fig. S2) were also performed based on parent-of-origin, verbal/ non-verbal status, IQ, age-at-first-words and age-at-first-phrases. The parent-of-origin analysis treated paternal- and maternal-specific transmissions separately for both *strict* and *spectrum* diagnostic classes. Stratification by verbal/non-verbal status as described in Liu *et al*. (64) was conducted in the total sample. IQ-based analyses included data from subjects whose IQ >80 (higher-IQ) and those with $25<$ IQ <70 (lower-IQ). IQ categorization was determined using available measures of verbal, non-verbal (performance) and full-scale IQ assessments. A score >80 for any of these three measures resulted in classification in the higher IQ group; otherwise, individuals classified into the lower IQ group had IQ estimates on at least two measures in the range 25–70. The age-at-first word and the age-at-first-phrase were analysed as quantitative variables (distributions shown in Supplementary Material, Figs S8 and 9).

Genetic association analysis

Association analyses were performed as described elsewhere (20). Briefly, family-based analyses were performed using FBAT (65) and an in-house programme written for family-based association (66) that implements methods described by Cordell *et al*. (67). For parent-of-origin analyses, only the latter was used. A summary of the sample size for analyses involving Stage 2 and both Stages is given in Table 1.

Allele score analysis

Four allele scores were defined based on the GWA signals from the four primary association analyses from Stage 1 described above. For each SNP, the risk allele was defined as the over-transmitted allele. The SNP-specific component of the score was calculated as the dosage of the allele multiplied by the corresponding log of the odds ratio, and was computed for all autosomal SNPs. The individual-specific score was calculated by summing over the SNP-specific component values without mean imputation for missing data, using PLINK version 1.07 (68). Ten scores were created for each individual using association thresholds of the *P*-value: <0.00001, <0.0001, <0.001, <0.01, <0.05, <0.1, <0.2, <0.3, <0.4, <0.5). The allele-score approach requires controls as well as cases; to account for the family-based design, we generated a pseudo control from the untransmitted parental alleles using PLINK v1.07. For any particular SNP, 'pseudo control' is the term used for the three genotypes that could have been formed by the parental alleles, but were not; i.e. if parents have genotypes a_1a_2 and a_3a_4, and their offspring is a_1a_3, then the pseudo controls are a_1a_4, a_2a_3 and a_2a_4. These controls form the foundation for likelihood calculations for family-based association analysis.

Logistic regression of Stage 2 case status on mean score was performed, with covariates being the number of non-missing genotypes for all SNPs used to calculate the score and four dimensions of ancestry defined using PLINK v1.07. An approximation of the additive genetic variance explained by the observed markers was calculated using the difference in the Nagelkerke's pseudo R-squared from a model including the score and covariates versus a model including only the covariates. Empirical significance of these findings was estimated based on 1000 permutations through case–control randomization within each data set.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the families participating in the study.

Conflict of Interest statement. None declared.

FUNDING

The authors thank 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: HD035469, HD055748, HD055751, HD055782, HD055784, MH52708, MH55284, MH057881, 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, The National Childrens Research Centre (Ireland), 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). Computational infrastructure for RJLA was supported by the Trinity Centre for High Performance Computing (http://www.tchpc.tcd.ie/) funded through grants from Science Foundation Ireland. Funding to pay the Open Access publication charges for this article was provided by Autism Speaks.

REFERENCES

1. Jamain, S., Quach, H., Betancur, C., Rastam, M., Colineaux, C., Gillberg, I.C., Soderstrom, H., Giros, B., Leboyer, M., Gillberg, C. *et al.* (2003) Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nat. Genet.*, **34**, 27–29.
2. Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., Yamrom, B., Yoon, S., Krasnitz, A., Kendall, J. *et al.* (2007) Strong association of de novo copy number mutations with autism. *Science*, **316**, 445–449.
3. Durand, C.M., Betancur, C., Boeckers, T.M., Bockmann, J., Chaste, P., Fauchereau, F., Nygren, G., Rastam, M., Gillberg, I.C., Anckarsater, H. *et al.* (2007) Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. *Nat. Genet.*, **39**, 25–27.
4. Marshall, C.R., Noor, A., Vincent, J.B., Lionel, A.C., Feuk, L., Skaug, J., Shago, M., Moessner, R., Pinto, D., Ren, Y. *et al.* (2008) Structural variation of chromosomes in autism spectrum disorder. *Am. J. Hum. Genet.*, **82**, 477–488.
5. Weiss, L.A., Shen, Y., Korn, J.M., Arking, D.E., Miller, D.T., Fossdal, R., Saemundsen, E., Stefansson, H., Ferreira, M.A., Green, T. *et al.* (2008) Association between microdeletion and microduplication at 16p11.2 and autism. *N. Engl. J. Med.*, **358**, 667–675.
6. Kumar, R.A., KaraMohamed, S., Sudi, J., Conrad, D.F., Brune, C., Badner, J.A., Gilliam, T.C., Nowak, N.J., Cook, E.H. Jr., Dobyns, W.B. *et al.* (2008) Recurrent 16p11.2 microdeletions in autism. *Hum. Mol. Genet.*, **17**, 628–638.
7. Fernandez, B.A., Roberts, W., Chung, B., Weksberg, R., Meyn, S., Szatmari, P., Joseph-George, A.M., Mackay, S., Whitten, K., Noble, B. *et al.* (2010) Phenotypic spectrum associated with de novo and inherited deletions and duplications at 16p11.2 in individuals ascertained for diagnosis of autism spectrum disorder. *J. Med. Genet.*, **47**, 195–203.
8. Glessner, J.T., Wang, K., Cai, G., Korvatska, O., Kim, C.E., Wood, S., Zhang, H., Estes, A., Brune, C.W., Bradfield, J.P. *et al.* (2009) Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature*, **459**, 569–573.
9. Pinto, D., Pagnamenta, A.T., Klei, L., Anney, R., Merico, D., Regan, R., Conroy, J., Magalhaes, T.R., Correia, C., Abrahams, B.S. *et al.* (2010) Functional impact of global rare copy number variation in autism spectrum disorders. *Nature*, **466**, 368–372.
10. Cook, E.H. Jr. and Scherer, S.W. (2008) Copy-number variations associated with neuropsychiatric conditions. *Nature*, **455**, 919–923.
11. Sanders, S.J., Ercan-Sencicek, A.G., Hus, V., Luo, R., Murtha, M.T., Moreno-De-Luca, D., Chu, S.H., Moreau, M.P., Gupta, A.R., Thomson, S.A. *et al.* (2011) Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron*, **70**, 863–885.
12. Sanders, S.J., Murtha, M.T., Gupta, A.R., Murdoch, J.D., Raubeson, M.J., Willsey, A.J., Ercan-Sencicek, A.G., DiLullo, N.M., Parikshak, N.N., Stein, J.L. *et al.* (2012) Disruptive de novo point mutations, revealed by whole-exome sequencing, are strongly associated with Autism Spectrum Disorders. *Nature*, **485**, 237–241.
13. O'Roak, B.J., Vives, L., Girirajan, S., Karakoc, E., Krumm, N., Coe, B.P., Levy, R., Ko, A., Lee, C., Smith, J.D. *et al.* (2012) Sporadic autism exomes reveal a highly interconnected network of de novo mutations. *Nature*, **485**, 246–250.
14. Neale, B., Kou, Y., Liu, L., Ma'ayan, A., Samocha, K.E., Sabo, A., Lin, C.F., Stevens, C., Wang, L.S., Makarov, V. *et al.* (2012) Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature*, **485**, 242–245.
15. Iossifov, I., Ronemus, M., Levy, D., Wang, Z., Hakker, I., Rosenbaum, J., Yamrom, B., Lee, Y.H., Narzisi, G., Leotta, A. *et al.* (2012) De novo gene disruptions in children on the autistic spectrum. *Neuron*, **74**, 285–299.
16. Pagnamenta, A.T., Khan, H., Walker, S., Gerrelli, D., Wing, K., Bonaglia, M.C., Giorda, R., Berney, T., Mani, E., Molteni, M. *et al.* (2011) Rare familial 16q21 microdeletions under a linkage peak implicate cadherin 8 (CDH8) in susceptibility to autism and learning disability. *J. Med. Genet.*, **48**, 48–54.
17. Talkowski, M.E., Mullegama, S.V., Rosenfeld, J.A., van Bon, B.W., Shen, Y., Repnikova, E.A., Gastier-Foster, J., Thrush, D.L., Kathiresan, S., Ruderfer, D.M. *et al.* (2011) Assessment of 2q23.1 microdeletion syndrome implicates MBD5 as a single causal locus of intellectual disability, epilepsy, and autism spectrum disorder. *Am. J. Hum. Genet.*, **89**, 551–563.
18. Vorstman, J.A., van Daalen, E., Jalali, G.R., Schmidt, E.R., Pasterkamp, R.J., de Jonge, M., Hennekam, E.A., Janson, E., Staal, W.G., van der Zwaag, B. *et al.* (2011) A double hit implicates DIAPH3 as an autism risk gene. *Mol. Psychiatry*, **16**, 442–451.

19. Vaags, A.K., Lionel, A.C., Sato, D., Goodenberger, M., Stein, Q.P., Curran, S., Ogilvie, C., Ahn, J.W., Drmic, I., Senman, L. *et al.* (2012) Rare deletions at the neurexin 3 locus in autism spectrum disorder. *Am. J. Hum. Genet.*, **90**, 133–141.
20. Sato, D., Lionel, A.C., Leblond, C.S., Prasad, A., Pinto, D., Walker, S., O'Connor, I., Russell, C., Drmic, I.E., Hamdan, F.F. *et al.* (2012) SHANK1 deletions in males with autism spectrum disorder. *Am. J. Hum. Genet.*, **90**, 879–887.
21. Anney, R., Klei, L., Pinto, D., Regan, R., Conroy, J., Magalhaes, T.R., Correia, C., Abrahams, B.S., Sykes, N., Pagnamenta, A.T. *et al.* (2010) A genome-wide scan for common alleles affecting risk for autism. *Hum. Mol. Genet.*, **19**, 4072–4082.
22. Wang, K., Zhang, H., Ma, D., Bucan, M., Glessner, J.T., Abrahams, B.S., Salyakina, D., Imielinski, M., Bradfield, J.P., Sleiman, P.M. *et al.* (2009) Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature*, **459**, 528–533.
23. Weiss, L.A., Arking, D.E., Daly, M.J. and Chakravarti, A. (2009) A genome-wide linkage and association scan reveals novel loci for autism. *Nature*, **461**, 802–808.
24. Curran, S., Bolton, P., Rozsnyai, K., Chiocchetti, A., Klauck, S.M., Duketis, E., Poustka, F., Schlitt, S., Freitag, C.M., Lee, I. *et al.* (2011) No association between a common single nucleotide polymorphism, rs4141463, in the *MACROD2* gene and autism spectrum disorder. *Am. J. Med. Genet. B Neuropsychiatr. Genet.*, **156B**, 633–639.
25. Szatmari, P., Paterson, A.D., Zwaigenbaum, L., Roberts, W., Brian, J., Liu, X.-Q., Vincent, J.B., Skaug, J.L., Thompson, A.P., Senman, L. *et al.* (2007) Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nat. Genet.*, **39**, 319–328.
26. Betancur, C. (2011) Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. *Brain Res.*, **1380**, 42–77.
27. Devlin, B. and Scherer, S.W. (2012) Genetic architecture in autism spectrum disorder. *Curr. Opin. Genet. Devel.*, **22**, 229–237.
28. Devlin, B., Melhem, N. and Roeder, K. (2011) Do common variants play a role in risk for autism? Evidence and theoretical musings. *Brain Res.*, **1380**, 78–84.
29. Sullivan, P. (2012) Don't give up on GWAS. *Mol. Psychiatry*, **17**, 2–3.
30. Purcell, S.M., Wray, N.R., Stone, J.L., Visscher, P.M., O'Donovan, M.C., Sullivan, P.F. and Sklar, P. (2009) Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, **460**, 748–752.
31. Luca, D., Ringquist, S., Klei, L., Lee, A.B., Gieger, C., Wichmann, H.E., Schreiber, S., Krawczak, M., Lu, Y., Styche, A. *et al.* (2008) On the use of general control samples for genome-wide association studies: genetic matching highlights causal variants. *Am. J. Hum. Genet.*, **82**, 453–463.
32. Lee, A.B., Luca, D., Klei, L., Devlin, B. and Roeder, K. (2010) Discovering genetic ancestry using spectral graph theory. *Genet. Epidemiol.*, **34**, 51–59.
33. Klei, L., Kent, B.P., Melhem, N., Devlin, B. and Roeder, K. (2011) GemTools: a fast and efficient approach to estimating genetic ancestry. http://arxiv.org/pdf/1104.1162.pdf.
34. Ma, D., Salyakina, D., Jaworski, J.M., Konidari, I., Whitehead, P.L., Andersen, A.N., Hoffman, J.D., Slifer, S.H., Hedges, D.J., Cukier, H.N. *et al.* (2009) A genome-wide association study of autism reveals a common novel risk locus at 5p14.1. *Ann. Hum. Genet.*, **73**, 263–273.
35. Risch, N. (2001) Implications of multilocus inheritance for gene-disease association studies. *Theor. Popul. Biol.*, **60**, 215–220.
36. Ferreira, M.A., Sham, P., Daly, M.J. and Purcell, S. (2007) Ascertainment through family history of disease often decreases the power of family-based association studies. Behav. Genet. *37*, 631–636.
37. Ripke, S., Sanders, A.R., Kendler, K.S., Levinson, D.F., Sklar, P., Holmans, P.A., Lin, D.Y., Duan, J., Ophoff, R.A., Andreassen, O.A. *et al.* (2011) Genome-wide association study identifies five new schizophrenia loci. Nat. Genet. *43*, 969–976.
38. Hiller, M., Huse, K., Szafranski, K., Jahn, N., Hampe, J., Schreiber, S., Backofen, R. and Platzer, M. (2006) Single-nucleotide polymorphisms in NAGNAG acceptors are highly predictive for variations of alternative splicing. *Am. J. Hum. Genet.*, **78**, 291–302.
39. Liu, J., Nyholt, D.R., Magnussen, P., Parano, E., Pavone, P., Geschwind, D., Lord, C., Iversen, P., Hoh, J., Ott, J. *et al.* (2001) A genomewide screen for autism susceptibility loci. *Am. J. Hum. Genet.*, **69**, 327–340.
40. Wisniowiecka-Kowalnik, B., Nesteruk, M., Peters, S.U., Xia, Z., Cooper, M.L., Savage, S., Amato, R.S., Bader, P., Browning, M.F., Haun, C.L. *et al.* (2010) Intragenic rearrangements in NRXN1 in three families with autism spectrum disorder, developmental delay, and speech delay. *Am. J. Med. Genet. B Neuropsychiatr. Genet.*, **153B**, 983–993.
41. Ching, M.S., Shen, Y., Tan, W.H., Jeste, S.S., Morrow, E.M., Chen, X., Mukaddes, N.M., Yoo, S.Y., Hanson, E., Hundley, R. *et al.* (2010) Deletions of NRXN1 (neurexin-1) predispose to a wide spectrum of developmental disorders. *Am. J. Med. Genet. B Neuropsychiatr. Genet.*, **153B**, 937–947.
42. Mikhail, F.M., Lose, E.J., Robin, N.H., Descartes, M.D., Rutledge, K.D., Rutledge, S.L., Korf, B.R. and Carroll, A.J. (2011) Clinically relevant single gene or intragenic deletions encompassing critical neurodevelopmental genes in patients with developmental delay, mental retardation, and/or autism spectrum disorders. *Am. J. Med. Genet. A*, **155A**, 2386–2396.
43. Gregor, A., Albrecht, B., Bader, I., Bijlsma, E.K., Ekici, A.B., Engels, H., Hackmann, K., Horn, D., Hoyer, J., Klapecki, J. *et al.* (2011) Expanding the clinical spectrum associated with defects in CNTNAP2 and NRXN1. *BMC Med. Genet.*, **12**, 106.
44. Sehested, L.T., Moller, R.S., Bache, I., Andersen, N.B., Ullmann, R., Tommerup, N. and Tumer, Z. (2010) Deletion of 7q34-q36.2 in two siblings with mental retardation, language delay, primary amenorrhea, and dysmorphic features. *Am. J. Med. Genet. A*, **152A**, 3115–3119.
45. Mefford, H.C., Muhle, H., Ostertag, P., von Spiczak, S., Buysse, K., Baker, C., Franke, A., Malafosse, A., Genton, P., Thomas, P. *et al.* (2010) Genome-wide copy number variation in epilepsy: novel susceptibility loci in idiopathic generalized and focal epilepsies. *PLoS. Genet.*, **6**, e1000962.
46. Zweier, C., de Jong, E.K., Zweier, M., Orrico, A., Ousager, L.B., Collins, A.L., Bijlsma, E.K., Oortveld, M.A., Ekici, A.B., Reis, A. *et al.* (2009) CNTNAP2 and NRXN1 are mutated in autosomal-recessive Pitt-Hopkins-like mental retardation and determine the level of a common synaptic protein in Drosophila. *Am. J. Hum. Genet.*, **85**, 655–666.
47. Poot, M., Beyer, V., Schwaab, I., Damatova, N., Van't Slot, R., Prothero, J., Holder, S.E. and Haaf, T. (2010) Disruption of CNTNAP2 and additional structural genome changes in a boy with speech delay and autism spectrum disorder. *Neurogenetics*, **11**, 81–89.
48. Jackman, C., Horn, N.D., Molleston, J.P. and Sokol, D.K. (2009) Gene associated with seizures, autism, and hepatomegaly in an Amish girl. *Pediatr. Neurol.*, **40**, 310–313.
49. Alarcon, M., Abrahams, B.S., Stone, J.L., Duvall, J.A., Perederiy, J.V., Bomar, J.M., Sebat, J., Wigler, M., Martin, C.L., Ledbetter, D.H. *et al.* (2008) Linkage, association, and gene-expression analyses identify CNTNAP2 as an autism-susceptibility gene. *Am. J. Hum. Genet.*, **82**, 150–159.
50. Arking, D.E., Cutler, D.J., Brune, C.W., Teslovich, T.M., West, K., Ikeda, M., Rea, A., Guy, M., Lin, S., Cook, E.H. *et al.* (2008) A common genetic variant in the neurexin superfamily member CNTNAP2 increases familial risk of autism. *Am. J. Hum. Genet.*, **82**, 160–164.
51. Bakkaloglu, B., O'Roak, B.J., Louvi, A., Gupta, A.R., Abelson, J.F., Morgan, T.M., Chawarska, K., Klin, A., Ercan-Sencicek, A.G., Stillman, A.A. *et al.* (2008) Molecular cytogenetic analysis and resequencing of contactin associated protein-like 2 in autism spectrum disorders. *Am. J. Hum. Genet.*, **82**, 165–173.
52. Vernes, S.C., Newbury, D.F., Abrahams, B.S., Winchester, L., Nicod, J., Groszer, M., Alarcon, M., Oliver, P.L., Davies, K.E., Geschwind, D.H. *et al.* (2008) A functional genetic link between distinct developmental language disorders. *N. Engl. J. Med.*, **359**, 2337–2345.
53. Williamson, C.M., Ball, S.T., Dawson, C., Mehta, S., Beechey, C.V., Fray, M., Teboul, L., Dear, T.N., Kelsey, G. and Peters, J. (2011) Uncoupling antisense-mediated silencing and DNA methylation in the imprinted Gnas cluster. *PLoS Genet.*, **7**, e1001347.
54. Anney, R.J., Kenny, E.M., O'Dushlaine, C., Yaspan, B.L., Parkhomenka, E., Buxbaum, J.D., Sutcliffe, J., Gill, M. and Gallagher, L. (2011) Gene-ontology enrichment analysis in two independent family-based samples highlights biologically plausible processes for autism spectrum disorders. *Eur. J. Hum. Genet.*, **19**, 1082–1089.
55. Yaspan, B.L., Bush, W.S., Torstenson, E.S., Ma, D., Pericak-Vance, M.A., Ritchie, M.D., Sutcliffe, J.S. and Haines, J.L. (2011) Genetic analysis of biological pathway data through genomic randomization. *Hum. Genet.*, **129**, 563–571.

56. Hussman, J.P., Chung, R.H., Griswold, A.J., Jaworski, J.M., Salyakina, D., Ma, D., Konidari, I., Whitehead, P.L., Vance, J.M., Martin, E.R. *et al.* (2011) A noise-reduction GWAS analysis implicates altered regulation of neurite outgrowth and guidance in autism. *Mol. Autism*, **2**, 1.
57. Davis, L.K., Gamazon, E.R., Kistner-Griffin, E., Badner, J.A., Liu, C., Cook, E.H., Sutcliffe, J.S. and Cox, N.J. (2012) Loci nominally associated with autism from genome-wide analysis show enrichment of brain expression quantitative trait loci but not lymphoblastoid cell line expression quantitative trait loci. *Mol. Autism*, **3**, 3.
58. Casey, J.P., Magalhaes, T., Conroy, J.M., Regan, R., Shah, N., Anney, R., Shields, D.C., Abrahams, B.S., Almeida, J., Bacchelli, E. *et al.* (2011) A novel approach of homozygous haplotype sharing identifies candidate genes in autism spectrum disorder. *Hum. Genet*.
59. Vieland, V.J., Hallmayer, J., Huang, Y., Pagnamenta, A.T., Pinto, D., Khan, H., Monaco, A.P., Paterson, A.D., Scherer, S.W., Sutcliffe, J.S. *et al.* (2011) Novel method for combined linkage and genome-wide association analysis finds evidence of distinct genetic architecture for two subtypes of autism. *J. Neurodev. Disord.*, **3**, 113–123.
60. Chapman, N.H., Estes, A., Munson, J., Bernier, R., Webb, S.J., Rothstein, J.H., Minshew, N.J., Dawson, G., Schellenberg, G.D. and Wijsman, E.M. (2011) Genome-scan for IQ discrepancy in autism: evidence for loci on chromosomes 10 and 16. *Hum. Genet.*, **129**, 59–70.
61. Liu, X.Q., Georgiades, S., Duku, E., Thompson, A., Devlin, B., Cook, E.H., Wijsman, E.M., Paterson, A.D. and Szatmari, P. (2011) Identification of genetic loci underlying the phenotypic constructs of autism spectrum disorders. *J Am. Acad. Child Adolesc. Psychiatry*, **50**, 687–696 e613.
62. Lord, C., Rutter, M. and Le Couteur, A. (1994) Autism diagnostic interview-revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J. Autism Dev. Disord.*, **24**, 659–685.
63. Lord, C., Rutter, M., Goode, S., Heemsbergen, J., Jordan, H., Mawhood, L. and Schopler, E. (1989) Autism diagnostic observation schedule: a standardized observation of communicative and social behavior. *J. Autism Dev. Disord.*, **19**, 185–212.
64. Liu, X.Q., Paterson, A.D. and Szatmari, P. (2008) Genome-wide linkage analyses of quantitative and categorical autism subphenotypes. *Biol. Psychiatry*, **64**, 561–570.
65. Laird, N.M., Horvath, S. and Xu, X. (2000) Implementing a unified approach to family-based tests of association. *Genet. Epidemiol.*, **19**(Suppl. 1), S36–42.
66. Crossett, A., Kent, B.P., Klei, L., Ringquist, S., Trucco, M., Roeder, K. and Devlin, B. (2010) Using ancestry matching to combine family-based and unrelated samples for genome-wide association studies. *Stat. Med.*, **29**, 2932–2945.
67. Cordell, H.J., Barratt, B.J. and Clayton, D.G. (2004) Case/pseudocontrol analysis in genetic association studies: a unified framework for detection of genotype and haplotype associations, gene-gene and gene-environment interactions, and parent-of-origin effects. *Genet. Epidemiol.*, **26**, 167–185.
68. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.