Molecular Psychiatry (2013), 1–8
© 2013 Macmillan Publishers Limited All rights reserved 1359-4184/13
www.nature.com/mp

ORIGINAL ARTICLE

Common genetic variants on 1p13.2 associate with risk of autism

K Xia[1,2,11], H Guo[1,2,3,11], Z Hu[1,2,11], G Xun[4], L Zuo[5], Y Peng[1,2], K Wang[6], Y He[4], Z Xiong[1,2], L Sun[6], Q Pan[1], Z Long[1], X Zou[7], X Li[1], W Li[1], X Xu[1], L Lu[1], Y Liu[1], Y Hu[1], D Tian[1], L Long[1], J Ou[4], Y Liu[4], X Li[4], L Zhang[1], Y Pan[1], J Chen[1], H Peng[1], Q Liu[1], X Luo[4], W Su[1], L Wu[1], D Liang[1], H Dai[1], X Yan[1,8], Y Feng[1,8], B Tang[1,8], J Li[1,2], Z Miedzybrodzka[9], J Xia[1], Z Zhang[1], X Luo[5], X Zhang[6], D St Clair[3,9], J Zhao[1,4] and F Zhang[3,10]

Autism is a highly heritable neurodevelopmental disorder, and known genetic variants, mostly rare, account only for a small proportion of cases. Here we report a genome-wide association study on autism using two Chinese cohorts as gene discovery ($n = 2150$) and three data sets of European ancestry populations for replication analysis of top association signals. Meta-analysis identified three single-nucleotide polymorphisms, rs936938 ($P = 4.49 \times 10^{-8}$), non-synonymous rs6537835 ($P = 3.26 \times 10^{-8}$) and rs1877455 ($P = 8.70 \times 10^{-8}$), and related haplotypes, *AMPD1-NRAS-CSDE1*, *TRIM33* and *TRIM33-BCAS2*, associated with autism; all were mapped to a previously reported linkage region (1p13.2) with autism. These genetic associations were further supported by a *cis*-acting regulatory effect on the gene expressions of *CSDE1*, *NRAS* and *TRIM33* and by differential expression of *CSDE1* and *TRIM33* in the human prefrontal cortex of post-mortem brains between subjects with and those without autism. Our study suggests *TRIM33* and *NRAS-CSDE1* as candidate genes for autism, and may provide a novel insight into the etiology of autism.

Molecular Psychiatry advance online publication, 5 November 2013; doi:10.1038/mp.2013.146

Keywords: association fine mapping; autism; common genetic variants; genome-wide association study; human genetics

INTRODUCTION

Autism (OMIM 209850) is a childhood neurodevelopmental disorder characterized by impairment in language communication, social interaction and responsiveness, and restricted and repetitive patterns of interest or behavior.[1] The disorder presents clinically during the first 3 years of life and is about three times more common in boys than in girls. The prevalence has increased from no more than 5 per 10 000 individuals throughout the 1980s to 1 in 50 school-age children in the United States according to a recent report in 2013,[2] although the estimate may be partly due to a change in the practices of diagnosis and ascertainment. The fact that autism concordance in MZ twins approaches 92% in contrast to 10% in DZ twins suggests a strong genetic basis.[3]

Evidence from different cases supports the fact that chromosomal abnormality contributes to the risk for autism,[4] and linkage studies and cytogenetic analysis have led to identification of several novel candidate genes including neurexins (*NRXNs*) and neuroligins (*NLGNs*). Through the genomic linkage scan, significant linkages have been reported on 2q31, 3q and 7q (22, 34).[5,6] Several regions of interest, including 13q, 16p, 17q, 1p13.2, 1q31.1, 5p13, 8q24, 15q, 19p and Xq, although not consistently, have been suggested in more than one study sample. The majority of these regions, except 1p13.2 and 2q32, have been reported with chromosomal abnormality.[6]

Several genome-wide studies have been carried out to identify the genetic variants associated with risk for autism or autistic spectrum disorders. Copy number variations (CNVs) including several large recurrent deletions or duplications have been found disrupting either single gene or a chromosomal region containing multiple genes, and the best established autism-associated CNVs include 7q11.23, 15q11–13, 16p11.2 and 22q11.2 loci, and *NRXN1*, *CNTN4*, *NLGNs* and *SHANK3* genes.[7–13] Recently conducted exome-sequencing studies suggest that hundreds of *de novo* mutations have some role in the development of autism, and solid evidence implicates a few specific genes (*CHD8*, *KATANAL2*, *SCN2A*, *NTNG1*).[14–17] Those structural variants or *de novo* mutations, many of which are high penetrant or altering protein but individually rare, together account for a limited proportion of the genetic risk for autism. In contrast, only a modest number of common variants have been reported at *CDH9-CDH10*, *SEMA5A* and *MACROD2* loci through genome-wide association (GWA) studies.[12,18,19]

Autism 'with marked phenotypic heterogeneity' is etiologically multifactorial. Hypothesizing that multiple common variants collectively or interacting with environmental factors account for a certain proportion of risk for autism,[20,21] we performed a GWA analysis of two independent cohorts of a Chinese population sample, and top signals were replicated in three data sets of European ancestry populations. Meta-analysis of single-nucleotide

[1]State Key Laboratory of Medical Genetics, Central South University, Changsha, Hunan, China; [2]School of Biological Science and Technology, Central South University, Hunan, China; [3]Division of Intramural Research Programs, National Institute of Mental Health, The National Institutes of Health, Bethesda, MD, USA; [4]Mental Health Institute, The Second Xiangya Hospital, Central South University, Hunan, China; [5]Department of Psychiatry, Yale University, New Haven, CT, USA; [6]State Key Laboratory Incubation Base of Dermatology, Anhui, China; [7]Department of Pediatrics, No. 3 Hospital of the Sun Yat-sen University, Guangdong, China; [8]The Xiangya Hospital, Central South University, Hunan, China; [9]University of Aberdeen, Royal Cornhill Hospital, Aberdeen, UK and [10]The Lieber Institute for Brain Development, Johns Hopkins University Medical Campus, Baltimore, MD, USA. Correspondence: Dr K Xia, State Key Laboratory of Medical Genetics, Central South University, 110 Xiangya Road, Changsha, Hunan 410078, China or Dr J Zhao, Mental Health Institute, The Second Xiangya Hospital, Central South University, 139 Middle Renming Road, Changsha, Hunan 410011, China or Dr F Zhang, Lieber Institute for Brain Development, 855 North Wolfe Street, Baltimore, MD 21205, USA.
E-mail: xiakun48@163.com or zhaojingpingcsu@163.com or fzhang20@jhmi.edu
[11]These authors contributed equally to this work.
Received 10 March 2013; revised 25 August 2013; accepted 17 September 2013

Table 1. Study sample characteristics and genotyping by cohort

Data set	*Analysis*	*Design*	*Illumina chips*	*Subjects*[a]	*Trios*	*SNPs*[b]	*Study population*
SKLMGtrios	GWAS	Family-based	370K[c]	825	275	702 234	Chinese
SKLMGcc	GWAS	Case–control	370K/610K[c]	1120		702 234	Chinese
AGRE	Meta-analysis	Family-based	550K	855	285	513 312	European ancestry
SFARI	Meta-analysis	Family-based	1M	2067	689	781 108	European ancestry
AGP	Meta-analysis	Family-based	1M	975	325	786 230	European ancestry

Abbreviations: AGP, Autism Genome Project; AGRE, Autism Genetic Research Exchange; GWAS, genome-wide association studies; SKLMGcc, case–control cohort (State Key Laboratory of Medical Genetics); SKLMGtrios, case–parent triad; SFARI, Simons Foundation for Autism Research Initiative; SNP, single-nucleotide polymorphism. [a]Number of subjects used in the analysis. [b]Number of SNPs on autosomes after genome-wide quality control. [c]Genotyping differences resolved by genome-wide imputation; $\approx$1.06 million SNPs successfully imputed and passed genome-wide quality control; analyzed 702K SNPs on the Illumina HumanHap 1M chip.

polymorphisms (SNPs) and haplotypes showing a consistent trend of association in the five cohorts identified three SNPs and haplotypes associated with autism at genome-wide significance. Significant *cis*-acting regulatory effects and differential expression between autism cases and controls were found for *NRAS-CSDE1* and *TRIM33*, providing molecular evidence for these genetic associations. Our findings provide a novel insight into the genetic etiology of autism.

MATERIALS AND METHODS

Study design and subjects

Subjects used for genome-wide gene discovery included one cohort of case–parent triad families ($n = 870$ subjects) and one cohort of cases with autism and unrelated controls ($n = 1280$ subjects) in the Chinese population. Autism probands were diagnosed independently by two experienced psychiatrists using the DSM-IV-TR (American Psychiatric Association, 2000) criteria for autism. The diagnostic procedure also included an assessment using a series of tools—neurological examination, mental status examination, the Childhood Autism Rating Scale (CARS) and Autism Behavior Checklist (ABC). Subjects with Asperger's syndrome, Rett syndrome and pervasive developmental disorder not otherwise specified (PPD-NOS), such as fragile-X syndrome, were excluded from this study. All participating individuals provided written informed consent.

Three other cohorts of an autism family-based European ancestry population were used to validate the findings of discovery, and a meta-analysis of five cohorts was carried out to obtain combined evidence for association. These cohorts included samples from the Autism Genetic Research Exchange (AGRE), the Simons Foundation for Autism Research Initiative (SAFRI) Base and the Autism Genome Project (AGP). The sample recruitments of the three European cohorts have been described elsewhere,[9,13,18,19] and additional sample descriptions are also provided in the Supplementary Method.

DNA extraction, genotyping and quality control

Whole genomic DNA was extracted from the whole blood of all autism patients and their parents as well as from unrelated normal controls in the two Chinese cohorts using standard proteinase K digestion and the phenol–chloroform method. All autism patients and their parents as well as 110 unrelated controls were genotyped for SNP across the whole genome using Illumina HumanHap CNV370K BeadChip at the State Key Laboratory of Medical Genetics (SKLMG) in Changsha, China. An additional 1000 unrelated controls were genotyped using Illumina HumanHap 610 Quad BeadChip at the Genotyping Core Facility at the State Key Laboratory Incubation Base for Dermatology in Hefei, China. All SNPs in the HumanHap CNV370 are covered in the HumanHap 610 Quad BeadChip.

Genome-wide quality control was applied to each individual cohort. Individuals with missing SNP call rate $>2\%$ were excluded; SNPs were zeroed out if Mendelian errors were $>5\%$ and individuals were removed if Mendelian errors were $>3\%$ from the analysis. Sample duplication and cryptic relatedness were examined by the identity-by-state analysis of genotypes in autosomes. One of each related pair (IBD-sharing coefficients >0.10) was excluded. Individuals with sex error were identified, based on the heterozygote of the X chromosome, and excluded. For better utilization of the two Chinese cohorts, genome-wide imputation using IMPUTE2[22] and HapMap CHB data as reference was performed to increase the genotyping coverage. Identity-by-state and multidimensional scaling analyses were used to examine genetic heterogeneity in each cohort and to remove genetic outliers.

The genotyping and quality control of the three European cohorts have been described elsewhere.[9,18,19] In brief, AGRE samples were genotyped using Illumina 550K, and Simons Foundation for Autism Research Initiative (SFARI) and AGP samples were typed using Illumina 1M chips. The same quality control procedure was applied to the existing data sets before the analysis.

Genetic association analysis

We first performed the transmission disequilibrium test analysis of the case–parent triad cohort and logistic regression analyses for the case–control cohort from China using PLINK[23]; the genetic model was also assessed for the case–control data. Because either of the two Chinese cohorts alone may not be able to provide enough power according to our power analysis (Supplementary Method), we set $\alpha = 0.10$ for both the case–parent triad and the case–control cohort. Individual SNPs passed the threshold and combined analysis showed nominal significance of association ($P < 0.01$) were included for meta-analysis with SNPs that showed the same direction of transmitted allele in three additional family cohorts of European ancestry. The combined *P*-values were calculated using Stouffer's *Z*-score method for meta-analysis (Supplementary Method).

Haplotype analysis

As the GWA study was designed to find common genetic variants, the functional variants that link the genetic association with human diseases may at least partly lie in the nearby linked chromosomal region.[24] To search for such possible functional SNPs, a haplotype analysis was carried out in each of five cohorts for SNPs that were genotyped in the two Chinese cohorts in a chromosomal region around the top association signals. This haplotype analysis was also used to confirm the single SNP association, rather than as a discovery analysis. The analysis was performed using the sliding window approach, followed by meta-analysis of haplotype association. Haplotype-based transmission disequilibrium test association ('--hap-tdt') and haplotype-based case–control association ('--hap-assoc') analyses were performed using PLINK for family-based and case–control samples, respectively, and haplotype frequency was estimated based on all samples in each individual cohort using PLINK ('--hap-freq'). Although replication and meta-analysis of multiple cohorts were used to reduce possible false positives, associations of the final haplotypes identified at each locus with autism were also assessed by permutation tests to further correct for multiple testing. The permutation test was performed using FBAT for each of the four family-based cohorts and using PLINK for the case–control cohort separately, followed by a calculation of combined *P*-value. SNP functional prediction and genomic characteristics were assessed using bioinformatics tools developed on the basis of HapMap data.[25]

Post-mortem brain expression analysis

Both *cis*-association of genetic variants with *mRNA* expression and differential expression were examined for the genome-wide-associated SNPs and loci in post-mortem human brains. The *cis*-association analysis was performed on the basis of genotype and gene expression data from the post-mortem human prefrontal cortex of 224 postnatal subjects (109 Caucasians and 115 African Americans). The sample was collected at the

© 2013 Macmillan Publishers Limited

Table 2. Association signal ($P<10\times10^{-5}$) *P*-values and meta-analysis of five independent cohorts (genotyped SNPs only)

Chr.	*BP*	*SNP*	*Genes*	*MA*	Discovery (two Chinese cohorts)					Replications (three European ancestry cohorts)						
					OR_1	P_1	OR_2	P_2	P_{1+2}	OR_3	P_3	OR_4	P_4	OR_5	P_5	P_comb
1	11 474 9804	rs6537825	*TRIM33*	A	1.60	2.33×10^{-4}	1.38	1.00×10^{-3}	8.27×10^{-7}	1.80	1.68×10^{-2}	1.40	1.03×10^{-1}	1.23	1.72×10^{-1}	3.26×10^{-8}
1	11 485 4231	rs3827735	*TRIM33*	A	0.73	1.21×10^{-2}	0.68	6.37×10^{-3}	2.12×10^{-4}	0.92	6.44×10^{-1}	0.75	1.22×10^{-1}	0.77	4.03×10^{-2}	3.22×10^{-5}
1	114 875 399	rs10858047	*TRIM33*	C	1.45	3.37×10^{-3}	1.34	4.55×10^{-3}	4.51×10^{-5}	1.36	7.94×10^{-2}	1.05	7.49×10^{-1}	1.20	1.01×10^{-1}	2.23×10^{-5}
1	114 900 615	rs1877455	*TRIM33-BCAS2*	A	0.65	6.25×10^{-4}	0.57	7.40×10^{-5}	1.78×10^{-7}	0.83	3.88×10^{-1}	0.83	3.61×10^{-1}	0.66	5.00×10^{-3}	8.70×10^{-8}
1	114 973 172	rs7539721	*DENND2C*	C	1.41	5.89×10^{-3}	1.47	5.54×10^{-3}	9.27×10^{-5}	1.51	2.16×10^{-2}	1.09	6.12×10^{-1}	1.14	2.65×10^{-1}	2.39×10^{-5}
1	11 500 2331	rs6537841	*DENND2C*	A	1.41	5.70×10^{-3}	1.48	5.01×10^{-3}	8.17×10^{-5}	1.43	4.45×10^{-2}	1.15	4.05×10^{-1}	1.14	2.65×10^{-1}	2.04×10^{-5}
1	115 023 957	rs761755	*AMPD1-NRAS*	G	1.41	4.96×10^{-3}	1.45	1.08×10^{-2}	1.51×10^{-4}	1.42	4.61×10^{-2}	1.15	4.05×10^{-1}	1.15	2.20×10^{-1}	2.55×10^{-5}
1	115 041 339	rs926938	*AMPD1-NRAS*	A	1.48	1.39×10^{-3}	1.73	8.99×10^{-5}	4.92×10^{-7}	1.17	2.04×10^{-1}	1.27	3.42×10^{-2}	1.15	8.34×10^{-2}	4.49×10^{-8}
1	115 065 208	rs10489525	*CSDE1*	A	0.72	1.05×10^{-2}	0.60	4.54×10^{-4}	1.80×10^{-5}	0.82	1.43×10^{-1}	0.90	3.96×10^{-1}	0.85	7.42×10^{-2}	5.51×10^{-6}
1	185 324 470	rs10911998	*PLA2G4A*	T	1.50	4.68×10^{-3}	1.29	1.21×10^{-2}	1.61×10^{-5}	1.10	4.59×10^{-1}	1.14	2.50×10^{-1}	1.17	5.81×10^{-2}	4.49×10^{-5}
8	28 587 858	rs240919	*EXTL3*	C	0.76	2.90×10^{-2}	0.77	3.03×10^{-2}	2.10×10^{-3}	0.73	2.18×10^{-2}	0.80	9.71×10^{-2}	0.90	2.40×10^{-1}	2.25×10^{-5}
11	85 349 350	rs618679	*PICALM*	A	1.29	5.28×10^{-2}	1.54	1.64×10^{-4}	5.47×10^{-5}	1.11	4.96×10^{-1}	1.13	4.01×10^{-1}	1.25	2.66×10^{-2}	2.41×10^{-5}
11	85 350 031	rs2077815	*PICALM*	G	1.32	3.35×10^{-2}	1.60	1.52×10^{-4}	2.89×10^{-5}	1.08	5.99×10^{-1}	1.11	4.59×10^{-1}	1.19	8.13×10^{-2}	6.59×10^{-5}
11	85 393 384	rs527162	*PICALM*	C	1.29	5.08×10^{-2}	1.60	5.33×10^{-5}	2.26×10^{-5}	1.09	5.53×10^{-1}	1.11	4.53×10^{-1}	1.22	5.15×10^{-2}	3.30×10^{-5}
11	85 458 970	rs669556	*PICALM*	C	1.26	7.19×10^{-2}	1.56	5.33×10^{-5}	3.63×10^{-5}	1.11	4.96×10^{-1}	1.12	4.03×10^{-1}	1.23	3.67×10^{-2}	2.39×10^{-5}
12	95 958 351	rs6538761	*NEDD1*	A	1.37	1.12×10^{-2}	1.32	2.21×10^{-2}	6.44×10^{-4}	1.52	1.46×10^{-3}	1.24	7.11×10^{-2}	1.08	3.57×10^{-1}	1.58×10^{-6}
18	3 173 354	rs10853291	*MYOM1*	T	1.26	7.42×10^{-2}	1.32	1.43×10^{-2}	2.75×10^{-3}	1.21	1.15×10^{-1}	1.43	2.09×10^{-3}	1.09	3.19×10^{-1}	9.83×10^{-6}

Abbreviations: AGP, Autism Genome Project; BP, base pair; AGRE, Autism Genetic Research Exchange; Chr., chromosome; MA, minor allele; OR_{1-5}, odds ratio estimates for the five cohorts; *P*_comb, combined *P*-value from meta-analysis of SNPs in all five cohorts; P_{1+2}, combined *P*-value from meta-analysis of two Chinese cohorts; P_{1-5}, *P*-values derived, respectively, from SKLMGtrios, SKLMGcc, AGRE, AGP and SFARI cohorts; SKLMGcc, case–control cohort (State Key Laboratory of Medical Genetics); SKLMGtrios, case–parent triad; SFARI, Simons Foundation for Autism Research Initiative; SNP, single-nucleotide polymorphism.

Clinical Brain Disorders Branch of National Institute of Mental Health of NIH, and details of genotyping and the experiment have been described elsewhere (http://braincloud.jhmi.edu).[26] General linear regression analysis was performed to test for the association of SNP genotype and the mean level of gene expression while adjusting for age, sex, post-mortem interval (in h), RNA integrity number and race. Differential gene expression was analyzed for the genome-wide-associated loci in the frontal cortex of human brains from 16 cases with autism and 16 controls. Instead of assessing each probe separately and to avoid multiple testing issues, we tested for the mean difference in gene expression between cases and controls for nine probes in the genome-wide-associated loci simultaneously using multivariate analysis of variance while adjusting for covariates (e.g., sex, age, RNA integrity number and post-mortem interval). The sample collection, original data quality control and data processing methods have been described elsewhere.[27]

RESULTS

Genome-wide associations

We first performed transmission disequilibrium test of genome-wide SNPs in one cohort of cases with autism and their parents ($n=825$ subjects) and logistic regression analysis in one cohort of autism cases and unrelated normal controls ($n=1120$ subjects) in the Chinese population sample (Table 1). No genome-wide significant association ($P<5\times10^{-08}$) was found in either cohort (Supplementary Figures S1 and S2), and evidence for population stratification in the case–control cohort was minimal (*Lambda* GC = 1.017). Combined *P*-values of SNPs showing consistent trend of association ($P<0.1$) in both cohorts suggested strong association signals ($P<10\times10^{-7}$) for rs6537825 and rs10858046 at *TRIM33*, for rs1877455 and rs4839385 at *BCAS2* and for rs2268697 and rs926938 at *AMPD1-NRAS* (Supplementary Table S1). Although none of these SNPs reached genome-wide significance, all were genotyped and located at a linkage region (1p13.2) previously indicated for autism.[6] We did not find any CNVs in either of the Chinese cohorts, although some normal structural variants have been reported in this region.

Replication analysis of three European ancestry cohorts

To control false-positive findings and gain statistical power to detect genetic association, we performed a replication analysis of those top association signals in two Chinese cohorts combined ($P<0.01$), followed by a meta-analysis with SNPs showing consistent trend of association in three additional data sets of European ancestry (Table 1). Several top signals discovered in the Chinese samples were replicated ($P<0.05$) in at least one of the three European ancestry data sets, and the associations appeared weaker than in the Chinese population. However, the meta-analysis suggests an association with autism at, or very close to, genome-wide significance with three SNPs: rs926938 ($P=4.49\times10^{-8}$) at *AMPD1-NRAS*, rs6537835 ($P=3.26\times10^{-8}$) in *TRIM33* and rs1877455 ($P=8.70\times10^{-8}$) (Table 2). Whereas rs6537835 is non-synonymous and rs926938 is located upstream of *AMPD1* and downstream of *NRAS*, rs1877455 is an intronic variant of *DENND2C* between *TRIM33* and *BCAS2* loci.

Although these three genome-wide-associated SNPs are located in the linkage disequilibrium (LD) block of 1p13.2, a different LD structure is noted between Chinese and European ancestry populations (Figure 1). According to the HapMap data in the combined Han Chinese and Japanese populations (CHBJPT), the genome-wide-associated SNP, rs926938 at *AMPD1-NRAS*, is in moderate LD ($r^2>0.563$ and $D'=1$) with two others—rs6537835 at *TRIM33* and rs1877455 at *TRIM33-BCAS2*, in which the latter two SNPs are in strong LD with each other ($r^2>0.908$ and $D'=1$). All three SNPs are common in Chinese populations (minor allele frequency (MAF), 0.38–0.45). In contrast, in the European ancestry population (CEU), rs926938 is in less LD ($r^2<0.11$ and $D'=1$) with both rs6537835 and rs1877455, but the latter two SNPs showed

© 2013 Macmillan Publishers Limited

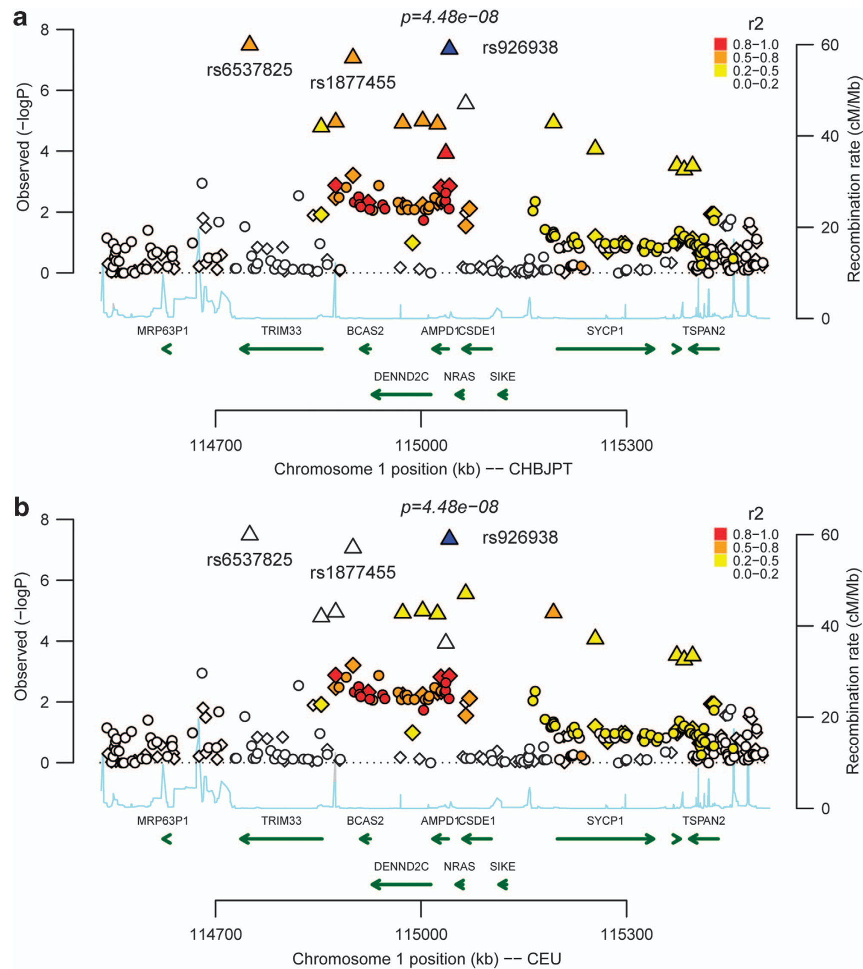

Figure 1. Regional association plot of negative logarithm of *P*-values with recombination in the 1p13.2 region. Linkage disequilibrium calculation is based on the genome-wide-associated SNP rs926938. (**a**, top panel) Based on HapMap data in Han Chinese and Japanese (CHBJPT) populations. (**b**, lower panel) Based on HapMap data in the European population. Triangle = meta-analysis *P*-values of SNPs, which shows consistent direction of association in five cohorts; circle = genotyped SNPs; and diamond = imputed SNPs in the Chinese triad cohort, which were not all replicated in the same direction for meta-analysis of five cohorts. CEU, European ancestry population.

the same strong LD ($r^2 > 0.908$, $D' = 1$) as in Asian populations (CHBJPT). It is worth noting that rs926938 is similarly common (MAF = 0.50) in both populations, but rs6537835 and rs1877455 have much lower allele frequency (MAF = 0.08) in the European ancestry population, which explains the weaker genetic association observed in the European ancestry samples.

Meta-analysis of single SNP associations in five cohorts also showed association signals ($P < 10 \times 10^{-5}$) that do not reach genome-wide significance at the 1p13.2 region (e.g., *DENND2C*, *CSDE1*, *SYCP1*) and at *PLA3G4A*, *EXTL3*, *PICALM*, *NEDD1* and *MYOM1* loci (Table 2). For those SNPs imputed in the two Chinese cohorts, meta-analyses did not find any GWAs with autism.

Functional haplotype for autism

We conducted a fast search for possible functional variants that may be in LD with the top association signals through haplotype analysis of 112 SNPs in the chromosomal region of 500 kb around the associated SNP rs926938. The haplotype analysis was also used to define the LD block spanned by the three associated SNPs, and was performed using up to 10-SNP sliding window approach in each of five sample cohorts.

Meta-analysis identified multiple haplotypes associated with autism at, or close to, genome-wide significance ($P < 10 \times 10^{-8}$). Interestingly, three most significant haplotypes were observed exactly around the three autism-associated SNPs (Table 3). According to further bioinformatics analyses in both populations,[25] the majority of SNPs involved with these three haplotypes are functional, but they were not detected as strong signals owing to lower MAF (e.g., 10489525 in *CSDE1*, $P = 5.51 \times 10^{-6}$). The *first* haplotype, AGG, was observed at rs926938 with rs8453 and rs10489525, and it spans the *AMPD1-NRAS-CSDE1* locus ($P = 9.33 \times 10^{-8}$). While SNPs rs926938 and rs10489525 are transcription-factor binding sites, rs8453 is conserved and appears with a high regulation potential (score = 0.35); both rs8453 and rs10489525 are located in *CSDE1*. The *second* haplotype, AGTTGTCCA, at rs6537825–rs11582563–rs11585926–rs11589568–rs7511633–rs6661053–rs11102800–rs3827735–rs11102807 was located at *TRIM33* ($P = 4.06 \times 10^{-8}$). In the E3 ubiquitin ligase *TRIM33*, rs6537825—the non-synonymous SNP associated with autism—causes the substitution of threonine for isoleucine. It should further be noted that these SNPs involved with the haplotype are in LD ($r^2 > 0.50$) with several other functional SNPs in *TRIM33* in HapMap data (Supplementary Tables S2 and S3), particularly in the CHB population. Finally, the *third* haplotype, TCT, involved rs10858047–rs11587400–rs1877455 at the *TRIM33-BCAS2* locus ($P = 8.36 \times 10^{-8}$). Although the SNP rs1877455 was the genome-wide-associated SNP with autism, the other two were

© 2013 Macmillan Publishers Limited

Table 3. Top associated haplotypes and frequency at three genome-wide-associated loci from the meta-analysis of five cohorts

Haplo	P_comb	SKLMGtrios			P-value	SKLMGcc			P-value	AGRE			P-value	SFARI			P-value	AGP			P-value
		Freq	T	U		Freq	F_A	F_U		Freq	T	U		Freq	T	U		Freq	T	U	
AMPD1-NRAS-CASDE1: rs926938|rs8453|rs10489525																					
AG	2.77×10^{-8}	0.481	164	111	0.0014	0.488	0.602	0.473	0.0001	0.489	147	126	0.2037	0.496	307	270	0.124	0.467	175	135	0.0231
AGG	9.33×10^{-8}	0.465	158	110	0.0034	0.480	0.597	0.469	0.0001	0.482	149	130	0.2472	0.483	328	283	0.069	0.454	184	149	0.0538
TRIM33: rs6537825|rs11582563|rs11585926|rs11589568|rs7511633|rs6661053|rs11102800|rs3827735|rs11102807																					
AGTTGT	1.98×10^{-7}	0.382	158	99	0.0002	0.396	0.461	0.387	0.0232	0.066	45	25	0.0168	0.083	96	78	0.172	0.075	56	40	0.1025
AGTTGTC	5.82×10^{-8}	0.382	158	97	0.0001	0.396	0.461	0.387	0.0232	0.066	45	25	0.0168	0.083	96	75	0.108	0.074	55	38	0.0779
AGTTGTCC	6.39×10^{-8}	0.382	158	97	0.0001	0.396	0.461	0.388	0.0256	0.066	45	25	0.0168	0.083	96	75	0.108	0.074	55	38	0.0779
AGTTGTCCA	4.06×10^{-8}	0.382	158	97	0.0001	0.395	0.464	0.391	0.0246	0.065	43	24	0.0203	0.082	96	74	0.092	0.071	53	35	0.0547
TRIM33-BCAS2: rs10858047|rs11587400|rs1877455																					
TCT	8.36×10^{-8}	0.444	105	159	0.0009	0.405	0.298	0.422	0.0002	0.080	40	47	0.4397	0.085	77	116	0.005	0.075	44	54	0.3124
CT	6.72×10^{-8}	0.449	106	162	0.0006	0.410	0.301	0.425	0.0002	0.081	39	47	0.3883	0.085	77	116	0.005	0.078	44	53	0.3608

Abbreviations: AGP, Autism Genome Project; AGRE, Autism Genetic Research Exchange; Freq, haplotype frequency in the sample; Haplo, haplotype; SKLMGcc, case–control cohort (State Key Laboratory of Medical Genetics); SKLMGtrios, case–parent triad; SFARI, Simons Foundation for Autism Research Initiative; SNP, single-nucleotide polymorphism; T, U, number of transmitted and non-transmitted haplotypes. SNPs boxed are genome-wide-associated loci in meta-analysis of five cohorts.

functional. The rs11587400 is conserved and may have high regulation potential (score = 0.45), and rs10858047 is also conserved and is in LD with several functional SNPs, including rs222493, a transcription-factor-binding site, and rs222493, an *miRNA*-binding site, at *TRIM33-BCAS2* ($r^2 = 0.88$ and $D' = 0.94$).

The permutation test confirmed that all three final haplotypes, AGG of rs926938–rs8453–rs10489525 at *AMPD1-NRAS-CSDE1* (permutation $P = 4.4 \times 10^{-5}$), AGTTGTCCA of rs6537825–rs11582563–rs11585926–rs11589568–rs7511633–rs6661053–rs11102800–rs3827735–rs11102807 at *TRIM33* (permutation $P = 3.9 \times 10^{-5}$) and TCT of rs10858047–rs11587400–rs1877455 at the *TRIM33-BCAS2* locus (permutation $P = 7.7 \times 10^{-6}$), were unlikely associated with autism by chance (Supplementary Table S4).

Cis-acting regulatory effects and differential expression in human brain

To find evidence of molecular mechanism for these genetic associations, we examined the *cis*-acting regulatory effect on *mRNA* gene expression in post-mortem human brains. As the true causal variants were not clear and statistical power to detect significant association with gene expression may be a concern when MAF is low and sample size is not large, genetic association with gene expression analysis was performed for all genes and SNPs that were involved in the three haplotypes associated with autism. Gene expression analyses of three SNPs, rs926938–rs8453–rs10489525, with each of three genes, *AMPD1-NRAS-CSDE1*, found that rs8453 was significantly associated with the gene regulations of *CSDE1* ($P = 0.0009$) and *NRAS* ($P = 0.0561$), and that rs10489525, was significantly associated with the gene regulations of *CSDE1* ($P = 0.0187$) and *NRAS* ($P = 0.0751$) (Figure 2). Although we did not observe significant association of rs6537825 with the expression of *TRIM33*, two common variants, rs11102800 and rs11102807 (MAF, 0.43–0.48), that formed an autism-associated haplotype with rs6537825 appear associated with *TRIM33* expression ($P < 0.054$). The analysis showed that disease-risk-associated alleles (rs8453, rs11102800 and rs11102807) appear associated with lower expressions of *CSDE1*, *NRAS* and *TRIM33*. Unfortunately, we did not find a significant association ($P > 0.20$) of the autism-implicated SNP, rs1877455, and two others (rs10858047 and rs11587400) on the haplotype, with the expressions of *TRIM33* and *BCAS2* (Supplementary Table S5).

The genome-wide-associated loci were differentially expressed in the post-mortem human frontal cortex between subjects with and those without autism. On the basis of the exact F-test for multivariate analysis of variance, we found an overall significant differential expression ($P = 0.0332$) of the nine probes in the six loci spanned by the genome-wide-associated SNPs (Supplementary Figure S3). Specifically, significant differential gene expressions were observed at *TRIM33* ($P = 0.0423$), *NRAS* ($P = 0.0463$) and *BCAS2* ($P = 0.0155$), and close to a significant level of differential expression ($P = 0.057$) at *CSDE1* (Supplementary Tables S6 and S7). This is worth noting given that the sample size (16 cases and 16 controls) is not large, and we also understand that MAF of the disease-associated SNPs is common in Chinese but much lower in the European ancestry population. More importantly, the differentially expressed genes in human brains were consistent with that where the *cis*-acting regulatory effects we found were on *TRIM33*, *NRAS* and *CSDE1* in an independent brain sample of normal subjects. Our findings also confirmed a finding in a recent study on post-mortem brain expression that *TRIM33*, in the prefrontal cortex, is differentially regulated ($P = 0.0049$, fold change $= -1.6$) in human brains between autism and normal subjects.[28]

Significant differential expression of *TRIM33* ($P = 0.025$) and *NRAS* ($P = 0.0099$) was also observed in the temporal cortex in a smaller number of samples comprising 10 autism cases and 11 normal controls (Supplementary Table S6).

© 2013 Macmillan Publishers Limited

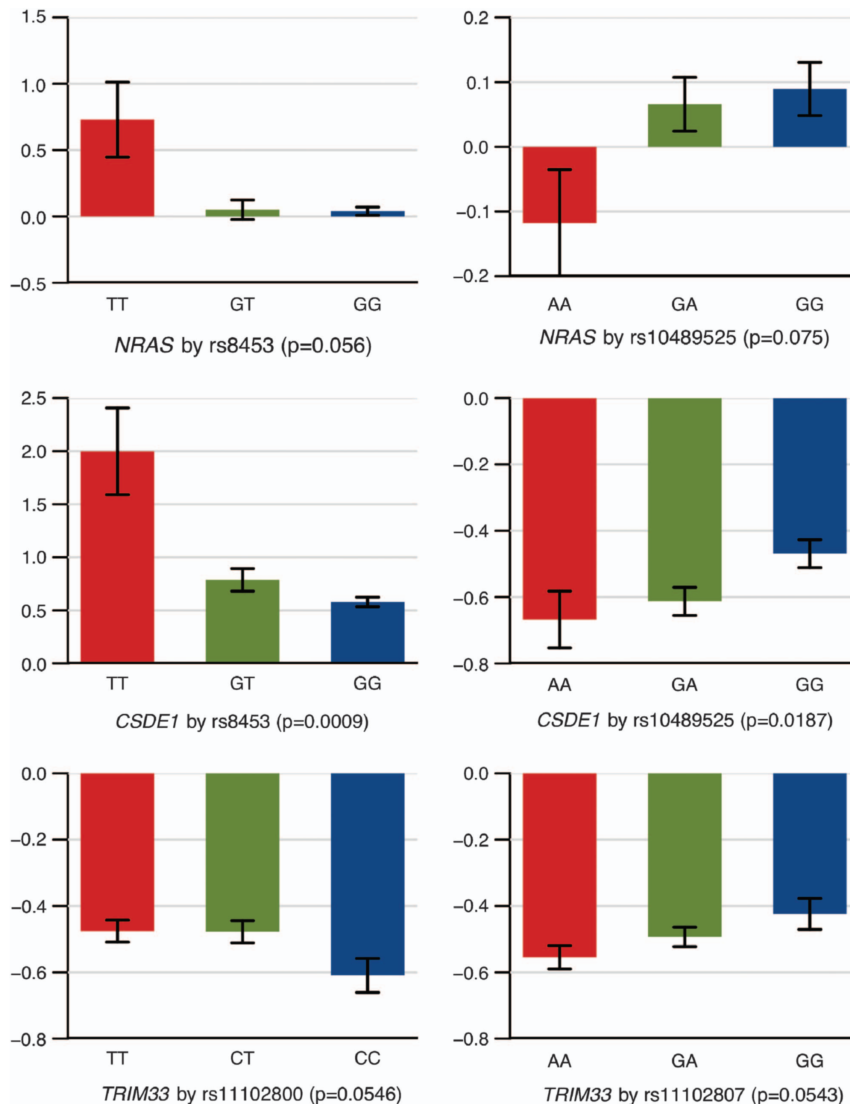

Figure 2. The *cis*-acting regulatory associations of genome-wide autism-associated loci at *NRAS-CSDE1* and *TRIM33* in postnatal human brains ($n = 224$). Plots were based on *post hoc* least-square estimates of mean by genotype from multiple linear regression models adjusted for covariates of sex, age (at brain sample collection), RNA integrity number and post-mortem intervals (in h).

DISCUSSION

Through GWA analysis of two cohort Chinese samples and replication analysis of top signals in three cohort European ancestry samples from previous GWA studies, our study revealed that multiple common genetic variants at *AMPD1-NRAS-CSDE1*, *TRIM33* and *TRIM33-BCAS2* are associated with autism, and all these loci are located in an LD block of 1p13.2 linkage region, previously reported as being linked to autism in multiple independent studies. In a whole genomic linkage scan of 147 autism-affected American sib pairs, using 362 microsatellite markers Risch *et al.*[29] reported the top signal (maximum multipoint LOD score 2.15) with microsatellite marker D1S1675.[29] This marker maps to within 1 Mb of *TRIM33*, which is the nearest gene. In a linkage study of 17 multiplex Finish families designed to replicate 10 previously reported linkage regions, a top linkage peak was observed with marker D1S1675, although none of the 10 loci were replicated.[30] Later, a high-density SNP genome-wide linkage scan in a large extended pedigree of seven individuals affected with autism also suggested 1p13.2 for autism (nonparametric linkage score $= 2.34$, $P = 0.0094$).[31] On the basis of our study, several genes in this linkage region could be potential candidates for autism susceptibility.

TRIM33, an E3 ubiquitin ligase, is the most likely candidate. Not only have mutations in ubiquitin E3 ligases been reported as being associated with other neuropsychiatric disorders such Angelman syndrome—a disorder with many features that overlap with autism (UBE3A),[32] Charcot–Marie–Tooth disease (LRSAM1),[33] Juvenile onset Parkinson's disease (PARK2)[34] and X linked-mental retardation (CUL4B)[35,36]—but also CNVs not found in controls affecting ubiquitin pathways, including UBE3A and PARK2, [8,10] have been found enriched in autism patients. The SNP in *TRIM33*, rs6537825, showing the strongest association with autism, is non-synonymous and causes an isoleucine-to-threonine substitution, and has significant *cis*-acting regulatory effect on *TRIM33* expression in the brain. This and other functional polymorphisms may slightly alter the function of *TRIM33* and make the ubiquitin–proteasome system less efficient. Finally, *TRIM33* was differentially expressed in human brains between autism patients and normal subjects, and was reported as one of the top loci associated with differential expressions across the whole genome in an independent sample of post-mortem human brains.[28]

© 2013 Macmillan Publishers Limited

In overlapping genes related to environmentally responsive, toxicogenomics and human immune and inflammatory response with 5300 genes in autism-linked regions, Herbert *et al.*[37] proposed *NRAS* as a candidate gene for potential gene–environment interactions in autism. However, we believe that *CSDE1*, or together with *NRAS*, is a likely candidate gene for autism. While a GWA was observed at functional SNP rs926938 between *AMPD1* and *NRAS*, this and two other functional SNPs, rs8453 and rs1048952 in *CSDE1*, formed a haplotype associated with autism at genome-wide significance. One of the functional SNPs in *CSDE1*, rs1048952, also showed as an association signal in multiple cohorts, and combined together in five cohorts it was just below the genome-wide significance. Moreover, two functional SNPs, rs8453 and rs1048952, have a *cis*-acting regulatory effect on *CSDE1* and *NARS*, both of which were differentially expressed in the frontal or temporal cortex of human brains between autism cases and normal controls. This provides further molecular evidence for *CSDE1* and *NRAS* as candidate genes for autism. A recent whole exome sequencing also identified that *de novo* loss-of-function mutation in *CSDE1* causes autism.[16]

Through this study, we also demonstrated the importance of further fine-mapping association analyses of genome-wide-associated loci through haplotype analysis in finding causal or functional variants that are often rare but are more likely to contribute to the GWA. The GWA study is designed to detect association with common genetic variants; however, in some cases the associations are difficult to understand owing to the lack of known functions for the associated SNPs. Some have argued by means of a simulation study that multiple rare mutations may account for the common genetic association through 'synthetic association'.[24] Although others may have an alternative argument against it, our study may provide an empirical evidence of the importance of fine-mapping association around the top signal. For example, one of our associated SNPs, rs926938, was found at *AMPD1-NRAS*; through haplotype analysis, we found a haplotype that minor allele of rs926938 formed with major alleles of two others in *CSDE1* associated with autism at genome-wide significance. Further downstream analyses of brain expressions tend to support *CSDE1* as a candidate gene in autism; unfortunately, the latter two SNPs were not detected as genome-wide-associated signals through a single SNP association analysis owing to lower MAF unless a haplotype analysis was performed.

Given that very few common genetic variants are found to be associated with autism, our study may provide novel insight into the parthenogenesis of common genetic variants and autism or other neuropsychiatric disorders.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

We are grateful to all the children with autism, their families and to the normal controls who participated in this study. We thank Autism Speaks for sharing resources from the Autism Genetic Resources Exchange (AGRE), the Simons Foundation for Autism Research Initiative (SFARI) for providing data from the Simons Simplex Collection (SCC), the NIH GWAS Data Repository (AGP data set: phs000267.v1.p1) and the Contributing Investigator(s) who contributed the phenotype and genotype data from his/her original studies. We also thank Mr Tianzhang Ye, Dr Carlo Colantuoni and Dr Joel E Kleinman for assisting in accessing the brain expression data and Dr Elizabeth Sherman for comments. The research was supported by the National Basic Research Program of China (2012CB517900, 2011CB510002), the National Natural Science Foundation of China (81330027, 81161120544), the National Alliance for Research on Schizophrenia and Depression (NARSAD) Award (17616 to LZ) and Intramural Research Program funding from the National Institute of Mental Health, The National Institutes of Health in the United States.

REFERENCES

1 Bailey A, Phillips W, Rutter M. Autism: towards an integration of clinical, genetic, neuropsychological, and neurobiological perspectives. *J Child Psychol Psychiatry* 1996; **37**: 89–126.
2 Blumberg S, Bramlett MD, Kogan MD, Schieve LA, Jones JR. Changes in prevalence of parent-reported autism spectrum disorder in school-aged U.S. children: 2007 to 2011–2012. *Natl Health Statist Rep* 2013; **65**: 1–12.
3 Hallmayer J, Glasson EJ, Bower C, Petterson B, Croen L, Grether J *et al.* On the twin risk in autism. *Am J Hum Genet* 2002; **71**: 941–946.
4 Veenstra-VanderWeele J, Cook Jr EH. Molecular genetics of autism spectrum disorder. *Mol Psychiatry* 2004; **9**: 819–832.
5 Gupta AR, State MW. Recent advances in the genetics of autism. *Biol Psychiatry* 2007; **61**: 429–437.
6 Veenstra-Vanderweele J, Christian SL, Cook Jr EH. Autism as a paradigmatic complex genetic disorder. *Annu Rev Genom Hum Genet* 2004; **5**: 379–405.
7 Durand CM, Betancur C, Boeckers TM, Bockmann J, Chaste P, Fauchereau F *et al.* Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. *Nat Genet* 2007; **39**: 25–27.
8 Glessner JT, Wang K, Cai G, Korvatska O, Kim CE, Wood S *et al.* Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature* 2009; **459**: 569–573.
9 Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, Moreno-De-Luca D *et al.* Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron* 2011; **70**: 863–885.
10 Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T *et al.* Strong association of *de novo* copy number mutations with autism. *Science* 2007; **316**: 445–449.
11 Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, Fossdal R *et al.* Association between microdeletion and microduplication at 16p11.2 and autism. *N Engl J Med* 2008; **358**: 667–675.
12 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.
13 Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R *et al.* Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* 2010; **466**: 368–372.
14 Muers M. Human genetics: fruits of exome sequencing for autism. *Nat Rev Genet* 2012; **13**: 377.
15 Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A *et al.* Patterns and rates of exonic *de novo* mutations in autism spectrum disorders. *Nature* 2012; **485**: 242–245.
16 Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ *et al. De novo* mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* 2012; **485**: 237–241.
17 O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, Girirajan S *et al.* Exome sequencing in sporadic autism spectrum disorders identifies severe *de novo* mutations. *Nat Genet* 2011; **43**: 585–589.
18 Wang K, Zhang H, Ma D, Bucan M, Glessner JT, Abrahams BS *et al.* Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature* 2009; **459**: 528–533.
19 Anney R, Klei L, Pinto D, Regan R, Conroy J, Magalhaes TR *et al.* A genome-wide scan for common alleles affecting risk for autism. *Hum Mol Genet* 2010; **19**: 4072–4082.
20 McClellan J, King MC. Genetic heterogeneity in human disease. *Cell* 2010; **141**: 210–217.
21 Geschwind DH. Autism: many genes, common pathways? *Cell* 2008; **135**: 391–395.
22 Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009; **5**: e1000529.
23 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559–575.
24 Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB. Rare variants create synthetic genome-wide associations. *PLoS Biol* 2010; **8**: e1000294.
25 Xu Z, Taylor JA. SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. *Nucleic Acids Res* 2009; **37**: W600–W605.
26 Colantuoni C, Lipska BK, Ye T, Hyde TM, Tao R, Leek JT *et al.* Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature* 2011; **478**: 519–523.
27 Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, Horvath S *et al.* Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* 2011; **474**: 380–384.
28 Chow ML, Pramparo T, Winn ME, Barnes CC, Li HR, Weiss L *et al.* Age-dependent brain gene expression and copy number anomalies in autism suggest distinct pathological processes at young versus mature ages. *PLoS Genet* 2012; **8**: e1002592.

© 2013 Macmillan Publishers Limited

29 Risch N, Spiker D, Lotspeich L, Nouri N, Hinds D, Hallmayer J *et al.* A genomic screen of autism: evidence for a multilocus etiology. *Am J Hum Genet* 1999; **65**: 493–507.
30 Auranen M, Nieminen T, Majuri S, Vanhala R, Peltonen L, Järvelä I. Analysis of autism susceptibility gene loci on chromosomes 1p, 4p, 6q, 7q, 13q, 15q, 16p, 17q, 19q and 22q in Finnish multiplex families. *Mol Psychiatry* 2000; **5**: 320–322.
31 Allen-Brady K, Miller J, Matsunami N, Stevens J, Block H, Farley M *et al.* A high-density SNP genome-wide linkage scan in a large autism extended pedigree. *Mol Psychiatry* 2009; **14**: 590–600.
32 Kishino T, Lalande M, Wagstaff J. UBE3A/E6-AP mutations cause Angelman syndrome. *Nat Genet* 1997; **15**: 70–73.
33 Guernsey DL, Jiang H, Bedard K, Evans SC, Ferguson M, Matsuoka M *et al.* Mutation in the gene encoding ubiquitin ligase LRSAM1 in patients with Charcot–Marie–Tooth disease. *PLoS Genet* 2010; **6**: e1001081.
34 Abbas N, Lücking CB, Ricard S, Dürr A, Bonifati V, De Michele G *et al.* A wide variety of mutations in the parkin gene are responsible for autosomal recessive parkinsonism in Europe. French Parkinson's Disease Genetics Study Group and the European Consortium on Genetic Susceptibility in Parkinson's Disease. *Hum Mol Genet* 1999; **8**: 567–574.
35 Tarpey PS, Raymond FL, O'Meara S, Edkins S, Teague J, Butler A *et al.* Mutations in CUL4B, which encodes a ubiquitin E3 ligase subunit, cause an X-linked mental retardation syndrome associated with aggressive outbursts, seizures, relative macrocephaly, central obesity, hypogonadism, pes cavus, and tremor. *Am J Hum Genet* 2007; **80**: 345–352.
36 Zou Y, Liu Q, Chen B, Zhang X, Guo C, Zhou H *et al.* Mutation in CUL4B, which encodes a member of cullin-RING ubiquitin ligase complex, causes X-linked mental retardation. *Am J Hum Genet* 2007; **80**: 561–566.
37 Herbert MR, Russo JP, Yang S, Roohi J, Blaxill M, Kahler SG *et al.* Autism and environmental genomics. *Neurotoxicology* 2006; **27**: 671–684.

Supplementary Information accompanies the paper on the Molecular Psychiatry website (http://www.nature.com/mp)

© 2013 Macmillan Publishers Limited