

ARTICLE

Gene-set analysis shows association between FMRP targets and autism spectrum disorder

Arija Jansen^{*,1,2}, Gwen C Dieleman², August B Smit³, Matthijs Verhage⁴, Frank C Verhulst², Tinca J C Polderman¹ and Danielle Posthuma^{1,2,4}

Autism spectrum disorder (ASD) is a heterogeneous group of disorders characterized by problems with social interaction, communication, and repetitive and restricted behavior. Despite its high heritability and the substantial progress made in elucidating genetic associations, the corresponding biological mechanisms are largely unknown. Our objective is to investigate the contribution of common genetic variation to biological pathways functionally involved in ASD. We conducted gene-set analyses to identify ASD-associated functional biological pathways using the statistical tools MAGMA and INRICH. Gene-set selection was based on previously reported associations with psychiatric disorders and resulted in testing of specific synaptic and glial sets, a glutamate pathway gene-set, mitochondrial gene-sets and gene-sets consisting of fragile X mental retardation protein (FMRP) targets. In total 32 gene-sets were tested. We used Psychiatric Genomics Consortium genome-wide association studies summary statistics of ASD. The study is based on the largest ASD sample to date ($N=5305$). We found one significantly associated gene-set consisting of FMRP-targeting transcripts (MAGMA: $p\text{ corr.} = 0.014$, INRICH: $p\text{ corr.} = 0.031$; all competitive P -values). The results indicate the involvement of FMRP-targeted transcripts in ASD in common genetic variation. This novel finding is in line with the literature as FMRP has been linked to fragile X syndrome, ASD and cognitive development in whole-exome sequencing and copy number variant studies. This gene-set has also been linked to Schizophrenia suggesting that FMRP-targeted transcripts might be involved in a general mechanism with shared genetic etiology between psychiatric disorders.

European Journal of Human Genetics (2017) 25, 863–868; doi:10.1038/ejhg.2017.55; published online 19 April 2017

INTRODUCTION

The heterogeneous manifestation of autism spectrum disorders (ASD) consists of several characteristic features including markedly abnormal social interaction, impaired communication abilities, and repetitive and restricted patterns of behavior and interests. Symptoms vary in severity per case and can occur with or without intellectual or language impairment.¹ Heritability is high, with an estimate of 60%,² which has led to extensive research into the genetic variants underlying ASD.

There have been several genome-wide studies including genome-wide association (GWA) and linkage studies, which reported promising associations.^{3–5} However, these associations explained only a small fraction of the genetic risk to ASD, showed little replicable results with P -values between 10^{-4} and 10^{-8} , and illustrated that most of the GWA studies up to date lack power to reliably determine a role for common variants in ASD.^{4–10}

The examination of copy number variants (CNVs) in ASD patients has been more successful by revealing enrichment of genes important for several, mostly synaptic, functions.^{11–14} However, causal CNVs occur in 5–10% of ASD cases and although they have a large effect on the liability to ASD, they are very rare, generally not specific to ASD, and cover large genomic areas including multiple genes.¹⁵

The most exciting genetic discoveries for ASD have been reported based on whole-exome sequencing (WES) data, in which several rare *de novo* variants in a diversity of genes have been linked to ASD.^{15–19} These WES findings implicated a role for genes involved in chromatin remodeling,^{16,17,19,20} synaptic formation,¹⁶ transcriptional regulation,¹⁶ and FMRP-associated genes^{16,17,20} (see for reviews refs 10,21).

The emerging picture is that ASD, like other psychiatric traits, is highly polygenic, and likely influenced by a mix of rare and common variants, of which functional implications still have to be determined.^{10,22} The identification of pathways with genes dysfunctional in ASD may increase by investigating the combined effect of multiple variants, using gene-set analysis. This, because it evaluates the joint effect of multiple genetic variants grouped according to biological or cellular function, thereby decreasing the multiple testing problem and increasing effect size.^{23–26}

In the current study, our objective is to investigate the contribution of common genetic variation to biological pathways functionally involved in ASD. To this end, we investigated whether the joint effect of common genetic variants grouped into *a priori* selected gene-sets is associated with ASD. Because of our hypothesis driven top down approach, we performed direct testing on all single nucleotide polymorphisms (SNPs) in a particular gene-set. This, instead of a

¹Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, VU University, Amsterdam, The Netherlands;

²Department of Child and Adolescent Psychiatry and Psychology, Erasmus University Medical Center-Sophia Children's Hospital, Rotterdam, The Netherlands; ³Department of Molecular and Cellular Neurobiology, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, VU University, Amsterdam, The Netherlands; ⁴Department of Clinical Genetics, VU University Medical Center, Amsterdam Neuroscience, Amsterdam, The Netherlands

*Correspondence: A Jansen, Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Neuroscience Campus Amsterdam, VU University, Amsterdam, The Netherlands, W&N Building, Room B-652, De Boelelaan 1085, 1081 HV Amsterdam, The Netherlands; Tel: +31 20 598 9898, Fax: +31 02 598 7112; E-mail: a.g.jansen@vu.nl

Received 12 October 2016; revised 14 February 2017; accepted 14 March 2017; published online 19 April 2017

functional enrichment analysis of top SNPs that does not require *a priori* hypotheses and uses top SNPs to define possible associated pathways based on functional enrichment.

We selected five categories of gene-sets, and limited ourselves to expert-curated gene-sets, resulting in testing 32 gene-sets. We included 19 curated synaptic²⁴ (category one) and three glial gene-sets (oligodendrocytes, astrocytes, and oligodendrocytes and astrocytes combined)²⁷ (category two), which have cell type-specific functions. Synaptic genes have been implicated to be among the top genes harboring variants associated with ASD^{14,16} and other psychiatric disorders.²⁸ The glial sets are an exploratory approach to provide general insights and starting points for more specific hypothesis formation although previous research has pointed in the direction of a role for astrocytes²⁹ and oligodendrocytes³⁰ in ASD, and for oligodendrocytes in schizophrenia (SCZ).³¹ The third category we selected consists of genes which gene-transcripts are targeted by fragile X mental retardation protein (FMRP). FMRP is an RNA binding protein expressed in the brain coded by the *FMRI* gene located at Xq27.3 that has been linked to fragile X syndrome (FXS)³² and evidence for an association with ASD is accumulating due to WES^{16,17,20} and CNV studies,³³ yet it has not been confirmed using common variants from genome-wide association studies (GWAS). Darnell *et al.*³⁴ and Ascano *et al.*³⁵ have provided insights into the biological underpinnings of FXS and ASD using human tissue and mouse models on FMRP and the RNA this binds to. Their research resulted in three gene-sets with FMRP target transcripts (Darnell gene-set, Ascano gene-set, Darnell and Ascano overlap gene-set). We aimed to test whether these sets of FMRP-targeted genes are associated with ASD in a large human sample not enriched with the *FMRI* variant. Our fourth gene-set category is a glutamate pathway. With glutamate being the most important excitatory agent in the brain, glutamate and its receptors have been suggested to have a role in psychiatric diseases including ASD.^{27,36} The fifth category (six gene-sets) is based on mitochondrial genes.²⁷ Mitochondria provide energy for the cell and with the brain being the organ using most of the energy, even a small reduction in energy production can result in impaired brain processes in the synapse.²⁷ Mitochondrial dysfunction has cautiously been associated with psychiatric diseases, including ASD, based on abnormal mitochondrial biomarker values and high prevalence of mitochondrial diseases in ASD patients compared to a healthy subpopulation.^{37,38}

In sum, our main goal is to directly test predefined sets of genes for their association with ASD. These gene-sets were selected because of previous associations with psychiatric diseases and test involvement of (1) synaptic processes, (2) glia cells, (3) FMRP, (4) glutamate and (5) mitochondrial involvement. The underlying hypothesis is that the

polygenic nature of ASD shows convergence of genetic effects in biologically meaningful sets of genes.

MATERIALS AND METHODS

Sample

We used the publicly available GWAS summary statistics (PGC.ASD.euro.all.25Mar2015.txt.gz) downloaded from <http://www.med.unc.edu/pgc/results-and-downloads> on 21 May 2015. More information on the sample can be found on the mentioned Psychiatric Genomics Consortium (PGC) website. Briefly, in their original study PGC used five cohorts: the Geschwind Autism Center of Excellence (ACE), the Autism Genome Project (AGP), the Autism Genetic Resource Exchange (AGRE), the NIMH Repository, the Montreal/Boston Collection (MONBOS), and the Simons Simplex Collection (SSC); see Table 1). The total number of ASD probands in this sample is 5305, and of pseudocontrols this is 5303. In a pseudocontrol setting, instead of a regular control group the non-transmitted parental allele is used as the control. All participants were of European descent. For the current gene-set analyses summary statistics (ie, *P*-values per SNP) of this PGC study were used.

Generation of gene-sets

We used 32 publicly available expert-curated gene-sets that we assigned to five distinct categories. The 19 synaptic gene-sets were published in previous studies^{14,24,28} in which they were defined based on assignment of subcellular function as determined by synaptic protein purification experiments and data mining for synaptic genes and gene ontology. Synaptic genes were subdivided into 19 functional groups (*N* genes 1047). The glial (146 genes), oligodendrocyte (52 genes), astrocyte (42 genes) and mitochondrial (six gene-sets, *N* genes 132) gene-sets were created and described by Duncan *et al.*²⁷ who conducted a database search in the gene ontology database and REACTOME. They supplemented the identified genes with genes found via an in-depth literature study. In addition, we included three gene-sets consisting of FMRP-targeting genes (*N* genes 1809) as defined by Darnell *et al.*³⁴ and Ascano *et al.*³⁵ with sequencing methods. All gene-sets as used in the present study are shown in Supplementary Tables S1 and S2.

MAGMA and INRICH gene-set analyses

Gene-set analyses can consist of self-contained testing and competitive testing. In a self-contained test the alternative hypothesis states that a gene-set is associated with the trait against the null hypothesis of no association, whereas in competitive testing the alternative hypothesis is that the gene-set is significantly stronger associated with the trait than genes not included in the gene-set.

We performed our analysis using two methods, MAGMA and INRICH, both providing competitive test results.³⁹

MAGMA (v1.01, <http://ctglab.nl/software/magma>) is a tool to perform gene and gene-set analysis which is distinguishable from other methods like INRICH, ALIGATOR, MAGENTA by having more statistical power, being less affected by linkage disequilibrium (LD; a SNP is in LD with another SNP when their specific alleles occur more often together than expected by

Table 1 Overview of the cohorts that were included in the initial analysis by the PGC

Cohort	Type of cohort	N cases/controls	Country	Ancestry	Diagnostic instrument
UCLA Autism Center of Excellence (ACE)	Parent-Parent-Proband Trios	391/391	USA	Caucasian	ADI-R and/or ADOS
Autism Genome Project (AGP)	Parent-Parent-Proband Trios	2272/2272	USA, Europe	Caucasian	ADI-R and ADOS
Autism Genetic Resource Exchange (AGRE)	Multiplex families	974/974	USA	Caucasian	ADI-R and ADOS
NIMH Repository, the Montreal/Boston Collection (MONBOS)	Simplex and multiplex families	1396/1396	USA/Canada	Caucasian	ADI-R (NIMH). Autism Screening Questionnaire, ADI-R ADOS (MONBOS)
Simons Simplex Collection (SSC)	Parent-Parent-Proband Trios	2231/2231	USA	Caucasian	ADI-R and ADOS

Abbreviations: ADI-R, Autism Diagnostic Interview-Revised; ADOS, Autism Diagnostic Observation Schedule; PGC, Psychiatric Genomics Consortium. The summary statistics, containing the *P*-values per SNP of this PGC analysis, were used for the gene-set analysis.

chance, implicating that the independent association assumption is violated and you can predict one of the specific alleles with high certainty dependent on the other known allele.) and multi-marker associations due to its multiple regression approach and being computationally less demanding as it does not use a permutation based approach.^{39,40} A significant hit in MAGMA indicates that multiple genes in the gene-set are associated. Although the SNPs included in these genes can have relatively high *P*-values, only together they are responsible for a positive signal in a gene-set. The 1000 genomes European panel (reference file) and NCBI 37.3 (gene location) (downloaded from <http://ctglab.nl/software/magma>) were used for SNP annotation to genes.

INRICH⁴¹ is a permutation based GWA analysis tool that tests whether functionally related genes compiled in gene-sets show a stronger association with a phenotype than expected by chance. A significant hit in

INRICH can occur with only a few highly associated SNPs in a gene-set. INRICH can be downloaded from <http://atgu.mgh.harvard.edu/inrich/downloads.html>. Utilizing Plink⁴² we computed several intervals for our analyses using different parameters for the LD clumping procedure. We applied the default INRICH values of 5000 first pass permutations, and 1000 s pass permutations.

Clumping is a method to reduce the amount of double signal in a data set due to LD. We assigned SNPs that are significant between a certain threshold to the same clump if they have an r^2 of 0.5 and are not yet assigned to another clump. For this clumping parameter we used several SNP *P*-value significance thresholds: 0.0001 and 0.01 (both fixed *P*-value thresholds) and 0.00896288 (1% cut-off *P*-value, computed in R studio v3.0.2, Boston, MA, USA) and 0.0008106764 (0.1% cut-off *P*-value, computed in R studio v3.0.2). These different thresholds influence which and how many SNPs are assigned to a

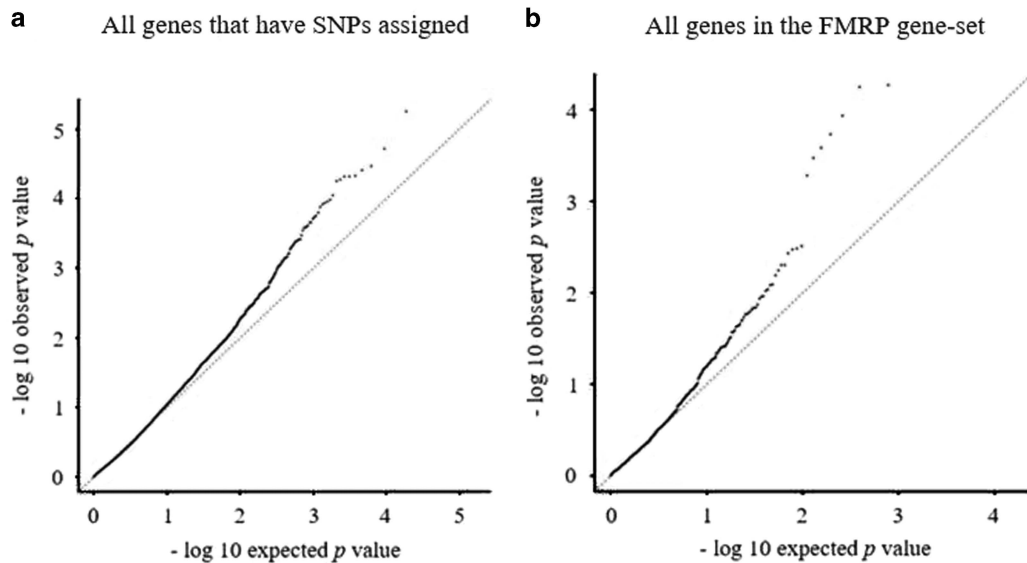


Figure 1 Visualisation of the polygenic pattern at gene level. (a) QQ plot of *P*-values of all genes that have SNPs from the PGC data set assigned to them regardless of inclusion in a gene-set. This plot compares the observed *P*-values to the expected *P*-values at gene level. The non-linear pattern (the deviation from the diagonal) visualizes the polygenic signal in the genes. (b) QQ plot of the genes in the FMRP (Darnell) gene-set. The earlier lift off and larger deviation from the diagonal compared to (a) illustrates the signal in a is driven in part by genes in the FMRP gene-set by Darnell.

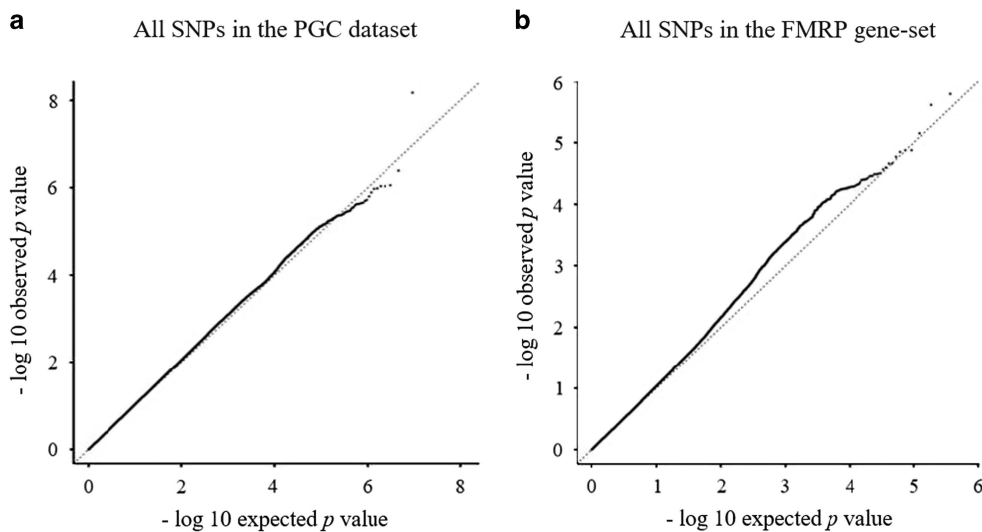


Figure 2 Visualisation of the polygenic pattern at SNP level. (a) QQ plot from all SNPs in the PGC data set. This plot compares the observed *P*-values to the expected *P*-values at SNP level. The non-linear pattern (the deviation from the diagonal) visualizes the polygenic signal in the SNPs. (b) QQ plot from the SNPs in the FMRP (Darnell) gene-set. The earlier lift off and larger deviation from the diagonal compared to (a) illustrates the signal in a is driven partly by SNPs in the FMRP gene-set by Darnell.

clump. A stricter *P*-value cut-off results in a clump with less SNPs. Again, we used NCBI 37.1 for gene location.

Statistical testing

We applied multiple testing following our hypothesis driven approach: per hypothesis we multiplied the *P*-value by the number of gene-sets that was tested for that particular hypothesis. Although this correction is not as strict as the Bonferroni correction, it provides sufficient correction as we constructed independent hypotheses generating independent results per hypothesis. This allows for multiple testing correction per hypothesis instead of a more stringent multiple testing correction over all hypotheses.

A consensus has not yet been reached so different studies parameterize and evaluate results differently.²⁵

RESULTS

MAGMA

Competitive gene-set analyses resulted in a statistically significant, multiple testing corrected association with the FMRP target gene-set by Darnell *et al.*³⁴ ($P=0.014$). None of the other gene-sets showed a statistically significant association, after multiple testing correction (for all results see Supplementary Table S3. Also, see Figure 1 (gene level) and Figure 2 (SNP level). These figures show the polygenic signal (Figures 1a and 2a) and the role of the FMRP gene-set in this signal (Figures 1b and 2b).

Gene-based tests in MAGMA for the genes included in the significant gene-set resulted in multiple significantly associated genes, indicating the results were not driven by a few highly associated genes and illustrating the value of testing groups of genes together (Supplementary Table S4).

INRICH

We performed this method with several parameter settings, as described in the method section. Overall, different parameter settings regarding clumping thresholds did not meaningfully change the results. However, the FMRP targets showed again a significant association with ASD after multiple testing correction, $P=0.031$, for the clumping SNP *P*-value significance threshold: 0.0001. For all results see Supplementary Table S5.

DISCUSSION

Previous research efforts have clearly shown that ASD is a highly polygenic disorder with reported heritability around 60%.² Some genetic variants have consistently been identified but functional implications of current genetic findings are as yet modest.^{14,22,25} In the present study, we tested whether the genetic variants of small effect on ASD tend to cluster in selected functional sets of genes. We tested expert-curated functional gene-sets that have been constructed previously in the context of their putative role in psychiatric disorders. Our multiple testing corrections were not as strict as a Bonferroni correction. If we had applied this correction no gene-set would be significantly associated. Still, we believe a sufficient correction was applied as we constructed, based on previous findings, independent hypotheses that as such generated independent results per hypothesis. Also, we applied competitive tests instead of the – far less stringent – self-contained tests, as usually used in this type of analyses. Moreover, the gene-sets are not independent, and thus a Bonferroni based on all gene-sets for all tested hypotheses is likely overly conservative.

The FMRP gene-set by Darnell was found to be significantly associated with the risk for ASD confirming previously reported associations in WES and CNV studies.^{17,33–35} This gene-set of

FMRP-targeting proteins was constructed by Darnell *et al.*³⁴ who identified FMRP interactions with mRNA in the mouse brain by means of high-throughput sequencing of RNAs isolated by cross-linking immunoprecipitation. Their study showed a connection between loss of function FMRP and ASD-associated symptoms in FXS and ASD patients. FMRP is important for translation of hundreds of neuronal mRNA's and its loss results in morphological and physiological neuronal defects resulting in FXS-like symptoms like cognitive impairment, seizures, anxiety and hyperactivity.^{43–45} The link between ASD and FXS seems intuitive as FXS is the leading form of monogenetic inherited intellectual disability with many cognitive and behavioral symptoms which are also manifest in ASD.⁴³ In addition, FXS shows comorbidity with ASD, about 30% of FXS patients are diagnosed with ASD, whereas 1 to 2% of ASD patients show FXS comorbidity.⁴³ As an FXS diagnosis was an exclusion criterion our results are not likely due to inclusion of patients with this monogenetic disorder. However, in the FXS there are pre-mutations, less CGG repeats than FXS patients but more than healthy individuals, causing diseases like fragile X-associated tremor ataxia syndrome, and increasing the incidence of ASD and ADHD.⁴⁶ Additional genetic phenotyping, including FMRP count, would be needed to ensure that samples are not enriched for FMRP pre-mutations.

A study on ASD rare variants⁴⁷ also reported associations with the Darnell FMRP gene-set, yet they found no evidence for overlap at the individual gene level. An association between this gene-set and SCZ has also been reported.^{47–50} Two SCZ CNV studies^{48,50} showed genetic overlap between ASD at gene-set and individual gene level. As a whole, these results might point in the direction of a common biological basis between ASD and SCZ making it of interest to look further into this possible overlap.

A general concern in psychiatric disorders is phenotypic heterogeneity and in ASD heterogeneity in intellectual disability (ID),⁵¹ is one of these concerns. As ID was not an exclusion criterion in our study, we cannot ensure with 100% certainty that our results are not partly driven by ID. Attempts have been made⁵² to stratify samples into low (IQ < 60) and high IQ (IQ > 60) but a downside is that subsequent decreasing sample sizes reduce statistical power. Unfortunately, we only had access to GWAS summary statistics and not to IQ scores and raw genotypes of participants we could not perform such analyses in our current study.

A final point to address regarding our FMRP hypothesis is that, out of three FMRP gene-sets, only the Darnell gene-set remained significant after multiple testing correction. The Ascano and Ascano autism overlap gene-sets did not generate significant results. Possible explanations for these findings are the different ways the gene-sets were constructed. Darnell³⁴ identified FMRP interactions within the mouse brain by means of high-throughput RNA sequencing and follow-up analysis, whereas Ascano *et al.*³⁵ examined FMR1 family protein binding sites to identify and rank FMRP targets in human embryonic kidney cells. These methods may have resulted in two different subsets of FMRP targets showing little overlap and expressing different biological properties. As a final remark, both gene-sets are large compared to all other tested sets. As the Ascano gene-set is larger than the Darnell gene-set, and given that effects of larger gene-sets are generally more easily to detect,⁴⁰ it is unlikely that gene-set size explains the association of the Darnell set. In addition, both MAGMA and INRICH have a correct type I error rate which is independent of gene-set size.⁴⁰

Our results do not support a role of the glutamate pathway, mitochondrial, synaptic or glial pathway in ASD, suggesting it is unlikely that there are large effects of these pathways on the risk of

ASD. However, gene-set definitions are dynamic, and with increased precision in pathway annotation, these results may change.²⁵

Taken together, the current results provide evidence for a role of FMRP-targeted transcripts in ASD. As FMRP is associated with several psychiatric conditions a more thorough exploration of genes in this gene-set and their association with different psychiatric disorders might provide useful information on an underlying shared genetic etiology between several disorders.

To conclude, we performed a gene-set analysis aiming to find common variation clustered in functional pathways associated with ASD. Our significant hit in common genetic variants is an FMRP-targeting gene-set that has been associated with ASD in rare variation and other psychiatric illnesses. These findings can point in the direction of a more general mechanism underlying psychiatric disorders making cross disorder research an important future component of the scientific repertoire.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We would like to thank the PGC for allowing us to use their summary statistics. Statistical analyses were carried out on the Genetic Cluster Computer (<http://www.geneticcluster.org>), which is financially supported by the Netherland Scientific Organization (NWO 480-05-003). We gratefully acknowledge financial support by the Sophia Stichting voor Wetenschappelijk Onderzoek under Grant Number 593 and the Netherland Scientific Organization (NWO VICI 453-14-005). ABS and MV received funding from the EU 7th Framework Program (HEALTH-F2-2009-242167 'SynSys').

DISCLAIMER

The manuscript is original, and not published nor under concurrent consideration elsewhere.

AUTHOR CONTRIBUTIONS

AJ and DP had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Study concept and design: DP; curation of synaptic gene-sets: MV, ABS; analysis of data: AJ; interpretation of results: all authors; drafting of the manuscript: AJ, TJCP, DP; critical revision of the manuscript for important intellectual content: all authors; obtained funding: GCD, FCV, MV, DP; study supervision: GCD, TJCP, DP.

- Sullivan PF, Daly MJ, O'Donovan M: Genetic architectures of psychiatric disorders: the emerging picture and its implications. *Nat Rev Genet* 2012; **13**: 537–551.
- Chow ML, Pramparo T, Winn ME *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.
- Gai X, Xie HM, Perin JC *et al*: Rare structural variation of synapse and neurotransmission genes in autism. *Mol Psychiatry* 2012; **17**: 402–411.
- Pinto D, Pagnamenta AT, Klei L *et al*: Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* 2010; **466**: 368–372.
- Sanders SJ, He X, Willsey AJ *et al*: Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. *Neuron* 2015; **87**: 1215–1233.
- Sanders SJ, Murtha MT, Gupta AR *et al*: *De novo* mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* 2012; **485**: 237–241.
- De Rubeis S, He X, Goldberg AP *et al*: Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* 2014; **515**: 209–215.
- Iossifov I, Ronemus M, Levy D *et al*: *De novo* gene disruptions in children on the autistic spectrum. *Neuron* 2012; **74**: 285–299.
- Neale BM, Kou Y, Liu L *et al*: Patterns and rates of exonic *de novo* mutations in autism spectrum disorders. *Nature* 2012; **485**: 242–245.
- O'Roak BJ, Vives L, Girirajan S *et al*: Sporadic autism exomes reveal a highly interconnected protein network of *de novo* mutations. *Nature* 2012; **485**: 246–250.
- Iossifov I, O'Roak BJ, Sanders SJ *et al*: The contribution of *de novo* coding mutations to autism spectrum disorder. *Nature* 2014; **515**: 216–221.
- Ronemus M, Iossifov I, Levy D, Wigler M: The role of *de novo* mutations in the genetics of autism spectrum disorders. *Nat Rev Genet* 2014; **15**: 133–141.
- De Rubeis S, Buxbaum JD: Recent advances in the genetics of autism spectrum disorder. *Curr Neurol Neurosci Rep* 2015; **15**: 36.
- Mooney MA, Wilmot B: Gene set analysis: a step-by-step guide. *Am J Med Genet Part B Neuro psychiatr Genet* 2015; **168**: 517–527.
- Ruano D, Abecasis GR, Glaser B *et al*: Functional gene group analysis reveals a role of synaptic heterotrimeric G proteins in cognitive ability. *Am J Hum Genet* 2010; **86**: 113–125.
- Sullivan PF, Posthuma D: Biological pathways and networks implicated in psychiatric disorders. *Curr Opin Behav Sci* 2015; 58–68.
- Torkamani A, Topol EJ, Schork NJ: Pathway analysis of seven common diseases assessed by genome-wide association. *Genomics* 2008; **92**: 265–272.
- Duncan L, Holmans PA, Lee PH *et al*: Pathway analyses implicate glial cells in schizophrenia. *PLoS ONE* 2014; **9**: e89441.
- Lips ES, Cornelisse LN, Toonen RF *et al*: Functional gene group analysis identifies synaptic gene groups as risk factor for schizophrenia. *Mol Psychiatry* 2012; **17**: 996–1006.
- Zeidán-Chuliá F, Salmina AB, Malinovskaya NA, Noda M, Verkhatsky A, Moreira JCF: The glial perspective of autism spectrum disorders. *Neurosci Biobehav Rev* 2014; **38**: 160–172.
- Zeidán-Chuliá F, de Oliveira BHN, Casanova MF *et al*: Up-regulation of oligodendrocyte lineage markers in the cerebellum of autistic patients: evidence from network analysis of gene expression. *Mol Neurobiol* 2015; **53**: 4019–4025.
- Chavarría-Siles I, White T, de Leeuw C *et al*: Myelination-related genes are associated with decreased white matter integrity in schizophrenia. *Eur J Hum Genet* 2016; **24**: 381–386.
- Verkerk AJ, Pieretti M, Sutcliffe JS *et al*: Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* 1991; **65**: 905–914.
- Pinto D, Delaby E, Merico D *et al*: Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *Am J Hum Genet* 2014; **94**: 677–694.
- Darnell JC, Van Driesche SJ, Zhang C *et al*: FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* 2011; **146**: 247–261.
- Ascano M, Mukherjee N, Bandaru P *et al*: FMR1 targets distinct mRNA sequence elements to regulate protein expression. *Nature* 2012; **492**: 382–386.
- Rojas DC: The role of glutamate and its receptors in autism and the use of glutamate receptor antagonists in treatment. *J Neural Transm* 2014; **121**: 891–905.
- Manji H, Kato T, Di Prospero NA *et al*: Impaired mitochondrial function in psychiatric disorders. *Nat Rev Neurosci* 2012; **13**: 293–307.
- Rosignol DA, Frye RE: Mitochondrial dysfunction in autism spectrum disorders: a systematic review and meta-analysis. *Mol Psychiatry* 2012; **17**: 290–314.
- de Leeuw CA, Mooij JM, Heskes T, Posthuma D: MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* 2015; **11**: e1004219.
- de Leeuw CA, Neale BM, Heskes T, Posthuma D: The statistical properties of gene-set analysis. *Nat Rev Genet* 2016; **17**: 353–364.
- Lee PH, O'Dushlaine C, Thomas B, Purcell SM: INRICH: interval-based enrichment analysis for genome-wide association studies. *Bioinformatics* 2012; **28**: 1797–1799.
- Purcell S, Neale B, Todd-Brown K *et al*: PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559–575.
- De Rubeis S, Bagni C: Regulation of molecular pathways in the fragile X syndrome: insights into autism spectrum disorders. *J Neurodev Disord* 2011; **3**: 257–269.
- Doll CA, Broadie K: Impaired activity-dependent neural circuit assembly and refinement in autism spectrum disorder genetic models. *Front Cell Neurosci* 2014; **8**: 30.
- Fernández E, Rajan N, Bagni C: The FMRP regulon: from targets to disease convergence. *Front Neurosci* 2013; **7**: 191.
- Lozano R, Rosero CA, Hagerman RJ: Fragile X spectrum disorders. *Intractable Rare Dis Res* 2014; **3**: 134–146.
- Purcell SM, Moran JL, Fromer M *et al*: A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* 2014; **506**: 185–190.

- American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*. Washington, DC, USA: American Psychiatric Association, 2013, Vol 5.
- Polderman TJC, Benyamin B, de Leeuw CA *et al*: Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat Genet* 2015; **47**: 702–709.
- Hussman JP, Chung R-H, Griswold AJ *et al*: A noise-reduction GWAS analysis implicates altered regulation of neurite outgrowth and guidance in autism. *Mol Autism* 2011; **2**: 1.
- Ma D, Salyakina D, Jaworski JM *et al*: A genome-wide association study of autism reveals a common novel risk locus at 5p14.1. *Ann Hum Genet* 2009; **73**: 263–273.
- Wang K, Zhang H, Ma D *et al*: Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature* 2009; **459**: 528–533.
- Anney R, Klei L, Pinto D *et al*: A genome-wide scan for common alleles affecting risk for autism. *Hum Mol Genet* 2010; **19**: 4072–4082.
- Anney R, Klei L, Pinto D *et al*: Individual common variants exert weak effects on the risk for autism spectrum disorders. *Hum Mol Genet* 2012; **21**: 4781–4792.
- Connolly JJ, Glessner JT, Hakonarson H: A genome-wide association study of autism incorporating autism diagnostic interview-revised, autism diagnostic observation schedule, and social responsiveness scale. *Child Dev* 2013; **84**: 17–33.
- Ronald A, Butcher LM, Docherty S *et al*: A genome-wide association study of social and non-social autistic-like traits in the general population using pooled DNA, 500 K SNP microarrays and both community and diagnosed autism replication samples. *Behav Genet* 2010; **40**: 31–45.

- 48 Fromer M, Pocklington AJ, Kavanagh DH *et al*: *De novo* mutations in schizophrenia implicate synaptic networks. *Nature* 2014; **506**: 179–184.
- 49 Schizophrenia Working Group of the Psychiatric Genomics Consortium: Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014; **511**: 421–427.
- 50 Szatkiewicz JP, O'Dushlaine C, Chen G *et al*: Copy number variation in schizophrenia in Sweden. *Mol Psychiatry* 2014; **19**: 762–773.
- 51 Cervantes PE, Matson JL: Comorbid symptomology in adults with autism spectrum disorder and intellectual disability. *J Autism Dev Disord* 2015; **45**: 3961–3970.
- 52 Chaste P, Klei L, Sanders SJ *et al*: A genome-wide association study of autism using the Simons simplex collection: does reducing phenotypic heterogeneity in autism increase genetic homogeneity? *Biol Psychiatry* 2015; **77**: 775–784.

Supplementary Information accompanies this paper on *European Journal of Human Genetics* website (<http://www.nature.com/ejhg>)