

11-17-2015

# Psychiatric gene discoveries shape evidence on ADHD's biology


Anita Thapar  
*Cardiff University*

J. Martin  
*Cardiff University*

Eric Mick  
*University of Massachusetts Medical School*

*See next page for additional authors*

Follow this and additional works at: [https://escholarship.umassmed.edu/umccts\\_pubs](https://escholarship.umassmed.edu/umccts_pubs)

 Part of the [Bioinformatics Commons](#), [Computational Biology Commons](#), [Computational Neuroscience Commons](#), [Genetics Commons](#), [Medical Genetics Commons](#), [Medical Molecular Biology Commons](#), [Mental Disorders Commons](#), [Molecular and Cellular Neuroscience Commons](#), [Molecular Biology Commons](#), [Molecular Genetics Commons](#), [Psychiatry Commons](#), and the [Translational Medical Research Commons](#)

---

## Repository Citation

Thapar, Anita; Martin, J.; Mick, Eric; Arias Vasquez, A.; Langley, Kate; Scherer, Stephen W.; Schachar, Russell; Crosbie, Jennifer; Williams, N.; Franke, Barbara; Elia, J.; Glessner, J.; Hakonarson, H.; Owen, M. J.; Faraone, S. V.; O'Donovan, Michael C.; and Holmans, Peter, "Psychiatric gene discoveries shape evidence on ADHD's biology" (2015). *UMass Center for Clinical and Translational Science Supported Publications*. 66.  
[https://escholarship.umassmed.edu/umccts\\_pubs/66](https://escholarship.umassmed.edu/umccts_pubs/66)

---

# Psychiatric gene discoveries shape evidence on ADHD's biology

## Authors

Anita Thapar, J. Martin, Eric Mick, A. Arias Vasquez, Kate Langley, Stephen W. Scherer, Russell Schachar, Jennifer Crosbie, N. Williams, Barbara Franke, J. Elia, J. Glessner, H. Hakonarson, M. J. Owen, S. V. Faraone, Michael C. O'Donovan, and Peter Holmans

## Keywords

UMCCTS funding

## Creative Commons License



This work is licensed under a [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/).

## Rights and Permissions

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

## ORIGINAL ARTICLE

## Psychiatric gene discoveries shape evidence on ADHD's biology

A Thapar<sup>1</sup>, J Martin<sup>1</sup>, E Mick<sup>2,3</sup>, A Arias Vázquez<sup>4,5</sup>, K Langley<sup>1,6</sup>, SW Scherer<sup>7</sup>, R Schachar<sup>8</sup>, J Crosbie<sup>8</sup>, N Williams<sup>1</sup>, B Franke<sup>4</sup>, J Elia<sup>9,10,11</sup>, J Glessner<sup>12</sup>, H Hakonarson<sup>13</sup>, IMAGE 2 Consortium<sup>16</sup>, MJ Owen<sup>1</sup>, SV Faraone<sup>14,15</sup>, MC O'Donovan<sup>1</sup> and P Holmans<sup>1</sup>

A strong motivation for undertaking psychiatric gene discovery studies is to provide novel insights into unknown biology. Although attention-deficit hyperactivity disorder (ADHD) is highly heritable, and large, rare copy number variants (CNVs) contribute to risk, little is known about its pathogenesis and it remains commonly misunderstood. We assembled and pooled five ADHD and control CNV data sets from the United Kingdom, Ireland, United States of America, Northern Europe and Canada. Our aim was to test for enrichment of neurodevelopmental gene sets, implicated by recent exome-sequencing studies of (a) schizophrenia and (b) autism as a means of testing the hypothesis that common pathogenic mechanisms underlie ADHD and these other neurodevelopmental disorders. We also undertook hypothesis-free testing of all biological pathways. We observed significant enrichment of individual genes previously found to harbour schizophrenia *de novo* non-synonymous single-nucleotide variants (SNVs;  $P = 5.4 \times 10^{-4}$ ) and targets of the Fragile X mental retardation protein ( $P = 0.0018$ ). No enrichment was observed for activity-regulated cytoskeleton-associated protein ( $P = 0.23$ ) or *N*-methyl-D-aspartate receptor ( $P = 0.74$ ) post-synaptic signalling gene sets previously implicated in schizophrenia. Enrichment of ADHD CNV hits for genes impacted by autism *de novo* SNVs ( $P = 0.019$  for non-synonymous SNV genes) did not survive Bonferroni correction. Hypothesis-free testing yielded several highly significantly enriched biological pathways, including ion channel pathways. Enrichment findings were robust to multiple testing corrections and to sensitivity analyses that excluded the most significant sample. The findings reveal that CNVs in ADHD converge on biologically meaningful gene clusters, including ones now established as conferring risk of other neurodevelopmental disorders.

*Molecular Psychiatry* advance online publication, 17 November 2015; doi:10.1038/mp.2015.163

## INTRODUCTION

Attention-deficit hyperactivity disorder (ADHD), a childhood-onset neurodevelopmental disorder,<sup>1</sup> is highly heritable.<sup>2,3</sup> Despite this strongly consistent finding, some remain sceptical about the diagnosis of ADHD, the biological validity of such a construct, and its neurodevelopmental origins.<sup>4</sup> Public misunderstanding is further fuelled by the relative scarcity of knowledge regarding its pathophysiology. The knowledge gap on pathophysiology is not straightforward to address given practical limitations to directly assess biological and molecular systems in those who are affected. Genetic findings offer one non-invasive approach to providing clues about the biology and pathogenesis of neuro-psychiatric disorders.<sup>5–7</sup>

In the last 5 years, one class of genetic variant (large, rare chromosomal deletions and duplications (copy number variants; CNVs)) has been found to contribute to ADHD risk across multiple studies.<sup>8–12</sup> Common genetic variants also contribute to ADHD when considered 'en masse' as a composite risk.<sup>13–19</sup> One

important observation from the original ADHD CNV studies was that the CNVs spanned chromosomal and gene regions that overlapped with schizophrenia and autism CNV loci.<sup>9,11</sup> Since then, exome-sequencing investigations of schizophrenia and autism have been published and have highlighted additional novel genomic regions that harbour neurodevelopmental disorder risk variants (single-nucleotide variants; SNVs).<sup>20–23</sup> Exome-sequencing results for ADHD are awaited. Unlike CNVs, which generally span multiple genes, SNVs are located in individual genes and thus offer improved resolution for testing cross-disorder pathogenic mechanisms.

Schizophrenia and autism genetic findings have also indicated specific biological mechanisms; these involve targets of the Fragile X Mental Retardation protein (FMRP) in schizophrenia and autism<sup>20–24</sup> and in schizophrenia, genes involved in glutamatergic post-synaptic processes: ARC (activity-regulated cytoskeleton-associated protein) and NMDAR (*N*-methyl-D-aspartate receptor) complexes.<sup>20,21,25</sup> Additional pathways have been reported to be enriched for SNVs (for example, calcium channel complexes in

<sup>1</sup>Institute of Psychological Medicine and Clinical Neurosciences and MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University School of Medicine, Cardiff, UK;

<sup>2</sup>Department of Quantitative Health Sciences, Department of Psychiatry, University of Massachusetts Medical School, Worcester, MA, USA; <sup>3</sup>Department of Psychiatry, University of Massachusetts Medical School, Worcester, MA, USA; <sup>4</sup>Radboud University Medical Center, Donders Institute for Brain, Cognition and Behaviour, Departments of Psychiatry and Human Genetics, Nijmegen, The Netherlands; <sup>5</sup>Radboud University Medical Center, Donders Institute for Brain, Cognition and Behaviour, Department of Cognitive Neuroscience, Nijmegen, The Netherlands; <sup>6</sup>School of Psychology, Cardiff University, Cardiff, UK; <sup>7</sup>Centre for Applied Genomics, Hospital for Sick Children, Toronto, Ontario, Canada; <sup>8</sup>Department of Psychiatry, Neurosciences and Mental Health, Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada; <sup>9</sup>Department of Pediatrics and Psychiatry, Sidney Kimmel College of Medicine, Thomas Jefferson University, Philadelphia, PA, USA; <sup>10</sup>Nemours Neuroscience Center, Wilmington, DE, USA; <sup>11</sup>Department of Psychiatry, University of Pennsylvania, Pennsylvania, PA, USA; <sup>12</sup>Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA; <sup>13</sup>Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA, USA; <sup>14</sup>Departments of Psychiatry and of Neuroscience and Physiology, SUNY Upstate Medical University, Syracuse, NY, USA and <sup>15</sup>K.G. Jebsen Centre for Research on Neuropsychiatric Disorders, University of Bergen, Bergen, Norway. Correspondence: Professor A Thapar or Professor P Holmans, Institute of Psychological Medicine and Clinical Neurosciences, MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University School of Medicine, Hadyn Ellis Building, Maindy Road, Cathays, Cardiff CF24 4HQ, UK.

E-mail: thapar@cf.ac.uk or holmanspa@cf.ac.uk

<sup>16</sup>Members of IMAGE 2 Consortium are listed before references.

Received 2 April 2015; revised 8 July 2015; accepted 8 September 2015

schizophrenia<sup>21</sup>), but such findings are not yet replicated across multiple studies and research designs.

In the current study, we assemble five ADHD and control CNV data sets from the United Kingdom, Ireland, Northern Europe, United States of America and Canada.<sup>10</sup> Our aim was to test the hypothesis that ADHD-associated CNVs show enrichment for gene sets that have been: (a) implicated by SNVs in the most recent, largest schizophrenia and autism *de novo* exome-sequencing studies<sup>20–23</sup> and (b) selected for *function* involving ARC, NMDAR and FMRP targets. We also undertook a hypothesis-free meta-analysis of all biological pathways across the five ADHD data sets.

## MATERIALS AND METHODS

### Participant and control data sets

The recruitment, assessment processes and clinical description of ADHD case subjects and controls have been described in detail previously.<sup>10</sup> Case–control CNV data sets were provided by the CNV lead for each study (SS, NW, JG, BF, EM). The five data sets came from: (a) Canada,<sup>11</sup> (b) Cardiff, UK,<sup>9,17</sup> (c) the Children's Hospital of Philadelphia, USA (CHOP),<sup>8</sup> (d) the International Multi-Center ADHD Genetics (IMAGE) 2 Project<sup>10,16</sup> and (e) the Pfizer-funded study from UCLA, Washington University and Massachusetts General Hospital (PUWMA).<sup>26</sup> All affected subjects from each of the five studies were children aged 5–18 years, had an IQ  $\geq 70$ , were of European descent, free of psychosis, epilepsy, serious neurological impairment and met Diagnostic and Statistical Manual of Mental Disorders (DSM)-III-R or DSM-IV criteria for ADHD as confirmed by semi-structured research diagnostic interviews. Collection and analysis of case–control data had been approved at each site by each Institutional Review Board or ethics committee. Case data were collected with informed consent from parents and assent from children.

After quality control exclusions (see later), the Canadian study included 247 DSM-IV ADHD cases and 2357 comparison subjects genotyped on the Affymetrix 6.0 array (Affymetrix, Santa Clara, CA, USA). The UK Cardiff study (excluding those that were included in IMAGE 2) included 603 DSM-IV ADHD cases genotyped on Illumina Human 660Q-Quad Beadchip (Illumina, San Diego, CA, USA) and 1047 comparison subjects genotyped on HumanHap550Beadchip. The CHOP study included 1013 participants with DSM-IV ADHD and 4105 comparison children genotyped on the Illumina Infinium HumanHap550K Beadchip. IMAGE 2 included data from 732 affected children collected in the United Kingdom, Ireland, Germany, Switzerland, the Netherlands and the United States. Control data came from 2010 subjects collected for a genome-wide association study of schizophrenia, described elsewhere.<sup>27</sup> A total of 692 ADHD cases from PUWMA and 1101 controls were genotyped on the Illumina 1M BeadChip.

Quality control and procedures for CNV calling are described in each of the original manuscripts (see Williams *et al.*<sup>10</sup> for a summary). CNVs were filtered for frequency ( $< 1\%$ ), and also for overlapping ( $> 50\%$  length) sites of known common CNVs defined by the Genome Structural Variation Consortium ([http://projects.tcag.ca/variation/ng42m\\_cnv.php](http://projects.tcag.ca/variation/ng42m_cnv.php)) or segmental duplications present in the March, 2006, human reference sequence (National Center for Biotechnology Information reference build 36.1, hg18). To allow for differences in CNV detection between the arrays, we used controls genotyped on platforms that matched the platforms used for the cases. The primary analyses used all CNVs  $> 500$  kb as these are the most reliably called across different genotyping platforms, have consistently been found to show increased burden in ADHD cases vs controls and have withstood experimental validation.<sup>10</sup> Pathways highlighted by the primary analysis were also tested for enrichment with all CNVs  $> 100$  kb (results available from last author).

### Statistical method for enrichment testing

The method is adapted from that proposed by Raychaudhuri *et al.*<sup>28</sup> and has been successfully applied to *de novo* schizophrenia CNVs.<sup>25</sup> Briefly, each CNV was labelled 'case/control' according to whether it occurs in a case or a control. A 'study' covariate for each data set was included as this corrects for variation in the genotyping assays and CNV estimation algorithms used across the different studies.

The following two logistic regression models were fitted to the sample of CNVs:

- (1) Case–control study+CNV length+number of genes hit outside pathway
- (2) Case–control study+CNV length+number of genes hit outside pathway +hit gene in pathway (yes/no)

and the deviances of the two models compared. If there is no enrichment of case CNV hits on pathway genes, then the difference in deviances should be distributed asymptotically as a  $\chi^2$  on one degree of freedom.<sup>29</sup> CNV length was fitted in the model because long CNVs are more likely to hit any set of genes than small ones, and CNV length may differ systematically between cases and controls. The 'number of genes hit outside pathway' was fitted to allow for case CNVs influencing disease status by hitting genes other than those in the pathway being tested. A binary variable (yes/no) is used for whether a CNV 'hits gene(s) in a pathway' rather than the number of genes in the pathway hit by the CNV (which is also a possibility) to allow for some pathways having several genes that are physically close together (thus, likely to be hit by the same CNV). The same analysis method was used to determine gene-specific enrichments, by defining the 'pathway' as the gene.

For all analyses, we tested pathways showing significant enrichment in the combined data set of five studies and then undertook tests of sensitivity by examining each of the five data sets separately and rerunning analysis after excluding the most significant single sample. Additional analyses were stratified by CNV type (deletions and duplications). As the number of duplications was greater than the number of deletions,<sup>10</sup> we tested for differential strengths of enrichment by comparing case duplications to case deletions in the logistic regression framework described above.

### Specific gene sets

**Genes selected by location.** Four sets of genes were defined: (a) non-synonymous and (b) loss-of-function *de novo* SNVs in schizophrenia were taken from the most recent published *de novo* exome-sequencing study of schizophrenia, which also catalogued and annotated all SNVs in previous studies in a consistent way.<sup>20</sup> There were a total of 611 genes containing at least one non-synonymous *de novo* SNV, with 87 of these containing a *de novo* loss-of-function SNV. We also took sets of genes containing: (c) non-synonymous and (d) loss-of-function *de novo* SNVs from the two largest, recent exome-sequencing studies of autism.<sup>22,23</sup> Combining the autism SNV sets from these studies, there were 2726 unique genes containing at least one non-synonymous *de novo* SNV, of which 538 contained a loss-of-function *de novo* SNV.

**Genes selected by function: FMRP targets, ARC and NMDAR.** Next, we examined a further three gene sets by function, selecting those implicated in schizophrenia and autism by CNV and more recent exome-sequencing studies. These were genes involved in FMRP targets (840 genes) as defined previously<sup>30</sup> and the ARC (28 genes) and NMDAR (61 genes) complexes, as defined by Kirov *et al.*<sup>25</sup> Bonferroni corrections were used to adjust for testing seven gene sets.

**Hypothesis-free analysis of all pathways.** This final analysis was undertaken using a large, unbiased, general set of pathways comprising:

- (a) Gene Ontology<sup>31</sup> (<http://www.geneontology.org/>, accessed on July 2013)
- (b) KEGG<sup>32</sup> (<http://www.genome.jp/kegg/>, accessed on June 2013)
- (c) PANTHER pathways version 3.1 (<http://www.pantherdb.org/pathway/>)
- (d) Mouse Genome Informatics database<sup>33</sup> (<http://www.informatics.jax.org/>, accessed on August 2013)
- (e) BioCarta (<http://www.biocarta.com>, accessed on June 2013)
- (f) Reactome<sup>34</sup> (<http://www.reactome.org>, accessed on June 2013)
- (g) NCI<sup>35</sup> (<http://pid.nci.nih.gov>, accessed on June 2013)

Pathways containing between 3 and 1500 genes were used in the analysis (15 111 in total). To increase the accuracy of the asymptotic *P*-values described above and to reduce the chance of a small pathway being falsely declared to be enriched based on a few gene hits, analysis was restricted to pathways with at least 10 gene hits in the total sample (7034 in total). Correction for multiple testing of pathways was performed by calculating *q*-values.<sup>36</sup>

## RESULTS

Table 1 shows the number of large, rare CNVs ( $> 500$  kb) for each of the previously published five case control samples and the total number of deletions and duplications. The rate and burden of CNVs for each sample have been published previously.<sup>9–11,26,37</sup>

**Table 1.** Burden and type of CNVs >500 kb in cases and controls from each study

CNV sample	CNVs	Number of CNVs				Number of deletions				Number of duplications			
		ADHD cases	Controls	Ratio	P	ADHD cases	Controls	Ratio	P	ADHD cases	Controls	Ratio	P
Canada <sup>11</sup>	N	22	217	0.97	0.88	4	52	0.73	0.55	18	165	1.04	0.87
	Rate	0.089	0.092			0.016	0.022			0.073	0.070		
Cardiff <sup>9,17</sup>	N	65	78	1.45	0.021	13	13	1.74	0.15	52	65	1.39	0.066
	Rate	0.108	0.074			0.022	0.012			0.086	0.062		
CHOP <sup>8</sup>	N	220	382	2.33	$4.7 \times 10^{-28}$	28	80	1.42	0.11	192	302	2.58	$4.4 \times 10^{-29}$
	Rate	0.217	0.093			0.028	0.019			0.190	0.074		
IMAGE 2 (refs 10,16)	N	89	191	1.28	0.042	22	47	1.29	0.32	67	144	1.28	0.084
	Rate	0.122	0.095			0.030	0.023			0.092	0.072		
PUWMA <sup>26</sup>	N	59	86	1.09	0.59	25	28	1.42	0.19	34	58	0.93	0.74
	Rate	0.085	0.078			0.036	0.025			0.049	0.053		

Abbreviations: ADHD, attention-deficit hyperactivity disorder; CHOP, Children's Hospital of Philadelphia; CNV, copy number variant; IMAGE, International Multi-Center ADHD Genetics; PUWMA, Pfizer-funded study from UCLA, Washington University and Massachusetts General Hospital. N: the number of CNVs observed in the sample; rate: the average number of CNVs per person.

### Gene sets selected by location of schizophrenia and autism *de novo* SNVs

The genes containing non-synonymous schizophrenia *de novo* SNVs<sup>20</sup> were significantly enriched for case ADHD CNV hits ( $P=5.4 \times 10^{-4}$  for CNVs >500 kb, see Table 2), which is significant after Bonferroni correction. Findings remained significant even after the most significant single sample was removed from the analysis (Table 2). The significant enrichment was observed for duplications ( $P=5.6 \times 10^{-4}$ ) but not for deletions ( $P=0.37$ ). However, there was no significant difference between ADHD duplication and deletions in terms of the rate of hits for genes previously found to carry non-synonymous schizophrenia SNVs ( $P=0.142$ ). For sample-specific enrichments, see Supplementary Table 1, and see Supplementary Table 2 for gene-wide *P*-values. Restricting analysis to genes hit by loss-of-function *de novo* schizophrenia SNVs showed no significant enrichment for case CNV hits.

Table 2 shows a nominally significant enrichment of CNVs among autism *de novo* SNV genes (non-synonymous or loss-of-function), although these do not survive Bonferroni correction.

### Gene sets selected by function

There was significant enrichment of ADHD case CNVs in FMRP targets ( $P=0.0018$ ), which remained significant after Bonferroni correction (see Table 3 for details). The enrichments also remained significant when the most significant single sample was removed. Significant enrichment was observed for duplications ( $P=0.005$ ) but not for deletions ( $P=0.18$ ), although there was no significant difference between the rate of case duplication and deletion hits for FMRP target genes ( $P=0.247$ ). See Supplementary Table 3 for sample-specific enrichments and Supplementary Table 4 for FMRP target genes that showed significant enrichment. There was no evidence of enrichment of ARC complex or NMDAR gene sets.

### Hypothesis-free analysis of all pathways

The most significantly enriched pathways are shown in Table 4, with those remaining significant in the sensitivity analysis highlighted in bold. As can be seen from the *q*-values, many of these were highly significant even after correction for multiple testing of pathways. Much of the enrichment appeared to come from duplications, with the exception of the ion channel pathways where enrichment was observed in both deletions and duplications but the strength of enrichment did not differ significantly for duplications and deletions (see Supplementary Table 5 for details). Ion channel pathways (ligand gated ion channel activity and ion

**Table 2.** Enrichment of ADHD case CNV hits in genes containing schizophrenia and autism *de novo* SNVs (non-synonymous and loss-of-function)

Gene set	<sup>a</sup> Genes	CNVs >500 kb			
		P	P (minus best) <sup>b</sup>	P (del)	P (dup)
SCZ (NS)	611	$5.4 \times 10^{-4}$	0.015	0.37	$5.6 \times 10^{-4}$
SCZ (LoF)	87	0.33	0.40	0.50	0.27
AUT (NS)	2726	0.019	0.054	0.31	0.024
AUT (LoF)	538	0.026	0.54	0.16	0.072

Abbreviations: ADHD, attention-deficit hyperactivity disorder; AUT, autism; CNV, copy number variant; Del, deletion; Dup, duplication; LoF, loss of function; NS, non-synonymous; SCZ, schizophrenia; SNV, single-nucleotide variant. <sup>a</sup>Number. <sup>b</sup>*P*-value for pathway enrichment after omitting the most significant sample.

**Table 3.** Enrichment of FMRP, ARC and NMDAR gene sets for ADHD case CNV hits

Gene set	# Genes	CNVs >500 kb			
		P	P (minus best) <sup>a</sup>	P (del)	P (dup)
FMRP	840	0.0018	0.032	0.18	0.0050
ARC	28	0.23	0.44	0.12	0.40
NMDAR	61	0.74	0.94	0.92	0.55

Abbreviations: ADHD, attention-deficit hyperactivity disorder; ARC, activity-regulated cytoskeleton-associated protein; CNV, copy number variant; Del, deletion; Dup, duplication; FMRP, Fragile X mental retardation protein; NMDAR, N-methyl-D-aspartate receptor. <sup>a</sup>*P*-value for pathway enrichment after omitting the most significant sample.

gated channel activity), transmembrane transport and organonitrogen compound catabolic process were robust to sensitivity analysis (see Supplementary Table 6 for enrichment effect sizes and the number of gene hits for pathways listed in Table 4; see Supplementary Tables 7-9 for most significant ion channel pathway genes, organonitrogen compound catabolic process and carbohydrate derivative catabolic process genes and transmembrane transport genes).



**Table 4.** Top pathways in pooled meta-analysis with most significant enrichment for ADHD case CNV hits among CNVs >500 kb

Pathway name	Pathway ID	# Genes	P	q-value	P (minus best) <sup>a</sup>
IL6-mediated signalling events	NCI:62	82	$1.33 \times 10^{-11}$	$9.17 \times 10^{-8}$	0.256
TGF- $\beta$ receptor signalling	NCI:95	44	$4.90 \times 10^{-11}$	$1.69 \times 10^{-7}$	0.508
Defence response to virus	GO:51607	147	$7.02 \times 10^{-8}$	$1.62 \times 10^{-4}$	0.31
Respiratory electron transport	<b>REACT:1019</b>	<b>81</b>	<b><math>1.98 \times 10^{-7}</math></b>	<b><math>3.41 \times 10^{-4}</math></b>	<b>0.02</b>
Organonitrogen compound catabolic process	<b>GO:1901565</b>	<b>893</b>	<b><math>9.08 \times 10^{-7}</math></b>	<b><math>1.05 \times 10^{-3}</math></b>	<b><math>3.13 \times 10^{-5}</math></b>
Transmembrane transporter activity	<b>GO:22857</b>	<b>902</b>	<b><math>9.10 \times 10^{-7}</math></b>	<b><math>1.05 \times 10^{-3}</math></b>	<b><math>8.85 \times 10^{-4}</math></b>
Citric acid (TCA) cycle and respiratory electron transport	REACT:1240	118	$1.16 \times 10^{-6}$	$1.15 \times 10^{-3}$	0.057
Carbohydrate derivative catabolic process	<b>GO:1901136</b>	<b>747</b>	<b><math>2.19 \times 10^{-6}</math></b>	<b><math>1.61 \times 10^{-3}</math></b>	<b><math>3.34 \times 10^{-4}</math></b>
Ligand-gated ion channel activity	<b>GO:15276</b>	<b>136</b>	<b><math>2.33 \times 10^{-6}</math></b>	<b><math>1.61 \times 10^{-3}</math></b>	<b><math>3.41 \times 10^{-4}</math></b>
Methyltransferase activity	<b>GO:8168</b>	<b>201</b>	<b><math>3.19 \times 10^{-6}</math></b>	<b><math>2.00 \times 10^{-3}</math></b>	<b>0.022</b>
Small thymus	MGI:706	199	$4.15 \times 10^{-6}$	$2.39 \times 10^{-3}$	0.488
Transmembrane transport	<b>GO:55085</b>	<b>1124</b>	<b><math>5.22 \times 10^{-6}</math></b>	<b><math>2.44 \times 10^{-3}</math></b>	<b>0.0028</b>
Ion-gated channel activity	<b>GO:22839</b>	<b>300</b>	<b><math>5.29 \times 10^{-6}</math></b>	<b><math>2.44 \times 10^{-3}</math></b>	<b><math>4.18 \times 10^{-4}</math></b>

Abbreviations: ADHD, attention-deficit hyperactivity disorder; CNV, copy number variant; GO, Gene Ontology; IL, interleukin; MGI, Mouse Genome Informatics; TCA, tricarboxylic acid; TGF, transforming growth factor. Pathways that are robustly enriched (that is, those with *P* (minus best) < 0.05) are shown in bold. <sup>a</sup>*P*-value for pathway enrichment after omitting the most significant sample.

Secondary analyses of CNVs >100 kb showed that observed pathways were generally significantly enriched in these analysis (results available from last author).

## DISCUSSION

In these analyses of ADHD case-control CNV data, the largest to date, we found highly significant enrichment of CNVs in many, although not all, hypothesised neurodevelopmental gene sets. Hypothesis-free testing revealed additional biological pathways enriched for CNVs in those with ADHD. Genes spanned by the ADHD CNVs were enriched for those that have recently been found to harbour schizophrenia-associated *de novo* SNVs. Previous work demonstrated overlap of ADHD CNVs with schizophrenia and autism CNVs,<sup>9–11</sup> however, because we were now able to define our gene sets using SNVs, which impact on single genes rather than chromosomal regions encompassing multiple genes, the current findings more precisely suggest overlap at the level of genes. These findings therefore extend, refine and are independent of previous studies that suggest biological overlap across these clinically very different disorders.

The clinical co-morbidity and co-heritability of ADHD is much more strongly established with autism than with schizophrenia.<sup>38</sup> Although previous studies, using some of the data sets in the present paper, found ADHD and autism overlap at the level of CNVs<sup>9–11</sup> and CNV-associated biological pathways,<sup>39</sup> we did not observe this extended to autism *de novo* SNV genes.

Targets of the FMRP have been implicated previously in both schizophrenia and autism.<sup>21–23</sup> Complete expression failure of FMRP itself, which characterises Fragile X syndrome, is known to be associated with elevated rates of ADHD, autism and other neurodevelopmental disorders.<sup>40</sup> Indeed, the majority of males with Fragile X syndrome show ADHD. Previous work has shown that the protein FMRP regulates activity of 842 biological targets and suggested that some of these likely underlie the manifestation of autistic features in Fragile X syndrome.<sup>30</sup> Our findings show that genes encoding FMRP targets are also enriched in ADHD CNVs, which could help explain why children with Fragile X syndrome show such high rates of ADHD. The results suggest that FMRP-mediated biology may be relevant across multiple neuropsychiatric disorders that include ADHD, as well as autism and schizophrenia. Although a recent report suggests that what has been considered as the specific contribution of FMRP targets to the pathogenesis of autism might simply reflect involvement of long, highly brain-expressed genes,<sup>41</sup> here large gene size is controlled for. Nevertheless, further work is needed to

understand how FMR1 is involved in the pathogenesis of different neurodevelopmental phenotypes, that is, ADHD, autism and schizophrenia, each of which has very different phenotype features and treatments.

Although schizophrenia and autism genetic findings have all strongly implicated the involvement of synaptic functions in pathogenesis,<sup>20,21,23</sup> we did not observe that to be the case for ADHD. Nor did we replicate prior pathway analysis implicating neurite outgrowth at synapses in ADHD, although previous analyses have used SNP or candidate gene data or smaller data sets.<sup>42–45</sup> We cannot be certain if this is a genuine point of difference or whether larger studies of ADHD, which consider additional types of mutations, for example, through exome sequencing, will yield a different pattern of findings.

The hypothesis-free analysis of all ADHD CNV pathways yielded strongly significant enrichment for multiple biological pathways. Many but not all of these were robust to analysis after excluding the most significant study. As expected for this final, hypothesis-free set of analyses, the significance of the enrichment was reduced by the removal of the most significant study, even when there was no significant difference in the strength of the enrichment between that study and the remainder of the sample, as was the case for most of the pathways highlighted here (see Supplementary Tables 1 for details). Thus, this analysis should be regarded as a sensitivity analysis testing the robustness of the pathway enrichment, rather than a measure of its overall significance. The fact that the enrichment strengths were not significantly different for most of the pathways of interest gives further evidence for consistency of enrichment across studies.

Ion channel pathways (ligand gated ion channel activity and ion gated channel activity) and especially involvement of *CHRNA7* and associated pathways have been implicated in previous analyses of a subset of the present pooled analysis.<sup>10,17,39</sup> Although some enrichment of immune-related pathways was also observed in the current study, this finding was not robust to the sensitivity analysis and thus requires cautious interpretation until more data become available. Regardless, the present study demonstrates that most findings were robust to sensitivity testing, whereby we excluded the most significant sample, despite potential variation in case mix severity and ascertainment across research centres and concerns from some quarters about the validity of ADHD.

Although our study involves analysis of the largest ADHD CNV data set to date, there are several limitations. First, analyses of this type for any disorder are restricted by the quality of gene and pathway annotations and the fact that the same genes belong to multiple different functional gene sets. Second, genotyped ADHD

sample sizes have markedly lagged behind those of many other neurodevelopmental and psychiatric disorders and results of exome sequencing are awaited. Third, ADHD-associated CNVs do not capture all possible forms of genetic risk, or risk loci; however, previous findings from smaller, individual studies do suggest biological convergence of CNVs and common gene variants.<sup>7,17,42,44,45</sup> However, the biology and validity of ADHD remain poorly understood and it is still widely considered to be primarily a catecholaminergic disorder.<sup>43</sup> Findings from the present study highlight the consistent and robust neurodevelopmental nature of ADHD and provide novel insights about its biological underpinnings.

In conclusion, in an international, cross-centre analysis of ADHD CNV data, we find evidence of biological overlap with schizophrenia at multiple levels; previous findings of overlap with schizophrenia CNVs have now been extended to SNVs. Furthermore, FMRP target enrichment appears to characterise ADHD as well as schizophrenia and autism. Ion channel pathway involvement in ADHD was robust to type of CNV and across samples. The findings reveal that CNVs in children with ADHD, from across multiple centres, converge on biologically meaningful gene clusters that are now robustly established as involved in neurodevelopmental disorder risk.

## CONFLICT OF INTEREST

Dr R Schachar has served as an advisor or consultant to Eli Lilly, Purdue Pharma and Highland Therapeutics and is named on institutional patents for several genes involved in ADHD and for a stop-task software programme. Dr SV Faraone has received research support from, served as consultant or adviser to or participated in CME programmes sponsored by Alcobra, Janssen, Eli Lilly, NIH, Novartis, McNeil, Pfizer and Shire; he receives royalties from Guilford Press and Oxford University Press. In the past year, Dr SV Faraone received income, potential income, travel expenses and/or research support from Pfizer, Ironshore, Shire, Akili Interactive Labs, CogCubed, Alcobra, VAYA Pharma, Neurovance, Impax and NeuroLifeSciences. With his institution, he has US patent US20130217707 A1 for the use of sodium-hydrogen exchange inhibitors in the treatment of ADHD. The remaining authors declare no conflict of interest.

## ACKNOWLEDGMENTS

We thank the patients and the family members who provided data for this project and research staff who helped collect and manage the data. They also thank Benjamin Neale for the original data retrieval. The Wellcome Trust, MRC and Action Medical Research have provided ADHD research support for AT, PH, JM, NW, MJO, MCO; we also acknowledge support from NIH grants R1 3MH059126, R0 1MH62873 and R0 1MH081803 to Dr SV Faraone. Dr E Mick received funding through the UMass Center for Clinical and Translational Science (P30HD004147) supported by the NIH.

## IMAGE 2 CONSORTIUM

Christine M Freitag, MD, MA, Michael Gill, MD, MRC Psych, Ric Anney, PhD, Lindsey Kent, MD, PhD, Frank Middleton, PhD, Yanli Zhang-James, MD, PhD, Lu Liu, MD, Jobst Meyer, PhD, Thuy Trang Nguyen, Dipl.Math.oec, Jasmin Romanos, MD, Marcel Romanos, MD, Christiane Seitz, MD, Tobias J Renner, MD, Susanne Walitza, MD, Andreas Warnke, MD, PhD, Haukur Palmason, PhD, Jan Buitelaar, MD, PhD, Barbara Franke, PhD, Nanda Rommelse, PhD, Ziarhi Hawi, MPhil, PhD, Joseph Sergeant, PhD, Hans-Christoph Steinhausen, MD, PhD, Herbert Roeyers, MD, PhD, Joseph Biederman, MD, Irina Zaharieva, BSc, Anath C Lionel, MSc, Christian R Marshall, PhD, Alexandre Todorov, PhD, Susan L Smalley, PhD, Sandra Loo, PhD, Stanley Nelson, MD, Corina Shtir, PhD, Philip Asherson, MRC Psych, PhD, Jonna Kuntsi, PhD, Andreas Reif, MD, Klaus-Peter Lesch, MD, James McGough, MD.

## IMAGE 2 CONFLICT OF INTEREST

Dr Christine M Freitag has served as speaker or advisor to Desitin, Eli Lilly, and Novartis. Dr Susanne Walitza has served on speakers bureaus for Eli Lilly, Janssen and AstraZeneca. Dr Jan Buitelaar has served as a consultant, advisory

board member or speaker for Bristol-Myers Squibb, Janssen Cilag BV, Eli Lilly, Novartis, Sche ring-Plough, Shire, Servier and UCB. Dr Joseph Sergeant has received an educational grant from Novartis and has served as speaker or adviser to Janssen Cilag, Lilly and Shire. Dr Herbert Roeyers is an advisory board member for Shire and has received research funding and conference attendance support from Shire and Eli Lilly. Dr Joseph Biederman has received research support from Elminda, Janssen, McNeil and Shire; speaking fees from Fundacion Areces (Spain), Fundacion Dr Manuel Camelo AC, Medice Pharmaceuticals and the Spanish Child Psychiatry Association; consulting fees from Shionogi Pharma and Cipher Pharmaceuticals (honoraria were paid to the Department of Psychiatry at Massachusetts General Hospital (MGH)); honoraria from the MGH Psychiatry Academy for a tuition funded CME course; an honorarium from Cambridge University Press for a chapter publication; and departmental royalties for a rating scale used for ADHD diagnosis (paid by Eli Lilly, Shire and AstraZeneca to the Department of Psychiatry at MGH). Dr Philip Asherson received funds on behalf of Kings College London for consultancy to Shire, Eli Lilly and Novartis; educational/research awards from Shire, Lilly, Novartis, Janssen, Vifor Pharma, GW Pharma and QbTech; speaker at sponsored events for Shire, Lilly and Novartis. Dr James McGough has received grant or research support from the National Institutes of Health (NIH), NeuroSigma, Purdue Pharma and Shire Pharmaceuticals. He has served as consultant to Merck and Sunovion Pharmaceuticals and provided expert witness testimony on behalf of Shire.

## IMAGE 2 ACKNOWLEDGMENTS

Supported by an Affymetrix Power Award (2007) to Dr Barbara Franke; a grant from Wellcome Trust, UK, for sample collection to Dr Lindsey Kent; an Internal Grant of Radboud University Nijmegen Medical Center to Dr Jan Buitelaar and grants from the Deutsche Forschungsgemeinschaft (KFO 125, SFB 581, GRK 1156 to Dr Klaus-Peter Lesch; ME 1923/5-1, ME 1923/5-3 to Drs Jobst Meyer and Christine M Freitag; and GRK 1389 to Dr Jobst Meyer) and the Bundesministerium für Bildung und Forschung (BMBF 0 1GV 0605 to Dr Klaus-Peter Lesch). The authors thank the Wellcome Trust and the Health Research Board Ireland for generous support to Drs Gill, Anney and Hawi.

## REFERENCES

- 1 Sonuga-Barke EJS, Taylor E. ADHD and Hyperkinetic Disorder. In: Thapar A, Pine D, Leckman JF, Scott S, Snowling MJ, Taylor E (eds). *Rutter's Child and Adolescent Psychiatry*, 6th (edn). John Wiley and Sons Limited: Chichester, UK, 2015.
- 2 Thapar A, Cooper M, Eyre O, Langley K. What have we learnt about the causes of ADHD? *J Child Psychol Psychiatry* 2013; **54**: 3–16.
- 3 Faraone SV, Mick E. Molecular genetics of attention deficit hyperactivity disorder. *Psychiatr Clin North Am* 2010; **33**: 159–180.
- 4 Timimi S, Taylor E. ADHD is best understood as a cultural construct. *Br J Psychiatry* 2003; **184**: 8–9.
- 5 The Network and Pathway Analysis Subgroup of the Psychiatric Genomics Consortium. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nat Neurosci* 2015; **18**: 199–209.
- 6 Nurnberger JI, Koller DL, Jung J, Edenberg HJ, Foroud T, Guella I *et al*. Identification of pathways for bipolar disorder: a meta-analysis. *JAMA Psychiatry* 2014; **71**: 657–664.
- 7 Hawi Z, Cummins TDR, Tong J, Johnson B, Lau R, Samraoui W *et al*. The molecular genetic architecture of attention deficit hyperactivity disorder. *Mol Psychiatry* 2015; **20**: 289–297.
- 8 Elia J, Glessner JT, Wang K, Takahashi N, Shtir CJ, Hadley D *et al*. Genome-wide copy number variation study associates metabotropic glutamate receptor gene networks with attention deficit hyperactivity disorder. *Nat Genet* 2012; **44**: 78–84.
- 9 Williams NM, Zaharieva I, Martin A, Langley K, Mantripragada K, Fossdal R *et al*. Rare chromosomal deletions and duplications in attention-deficit hyperactivity disorder: a genome-wide analysis. *Lancet* 2010; **376**: 1401–1408.
- 10 Williams NM, Franke B, Mick E, Anney RJL, Freitag CM, Gill M *et al*. Genome-wide analysis of copy number variants in attention deficit hyperactivity disorder: the role of rare variants and duplications at 15q13.3. *Am J Psychiatry* 2012; **169**: 195–204.
- 11 Lionel AC, Crosbie J, Barbosa N, Goodale T, Thiruvahindrapuram B, Rickaby J *et al*. Rare copy number variation discovery and cross-disorder comparisons identify risk genes for ADHD. *Sci Transl Med* 2011; **3**: 95ra75.

- 12 Jarick I, Volckmar AL, Pütter C, Pechlivanis S, Nguyen TT, Dauvermann MR *et al.* Genome-wide analysis of rare copy number variations reveals PARK2 as a candidate gene for attention-deficit/hyperactivity disorder. *Mol Psychiatry* 2012; **19**: 115–121.
- 13 Lee SH, Ripke S, Neale BM, Faraone SV, Purcell SM, Perlis RH *et al.* Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet* 2013; **45**: 984–994.
- 14 Hamshere ML, Langley K, Martin J, Agha SS, Stergiakouli E, Anney RJ *et al.* High loading of polygenic risk for ADHD in children with comorbid aggression. *Am J Psychiatry* 2013; **170**: 909–916.
- 15 Yang L, Neale BM, Liu L, Lee SH, Wray NR, Ji N *et al.* Polygenic transmission and complex neuro developmental network for attention deficit hyperactivity disorder: genome-wide association study of both common and rare variants. *Am J Med Genet Part B Neuropsychiatr Genet* 2013; **162B**: 419–430.
- 16 Neale BM, Medland SE, Ripke S, Asherson P, Franke B, Lesch K-PP *et al.* Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry* 2010; **49**: 884–897.
- 17 Stergiakouli E, Hamshere M, Holmans P, Langley K, Zaharieva I, Hawi Z *et al.* Investigating the contribution of common genetic variants to the risk and pathogenesis of ADHD. *Am J Psychiatry* 2012; **169**: 186–194.
- 18 Martin J, Hamshere ML, Stergiakouli E, O'Donovan MC, Thapar A. Genetic risk for attention-deficit/hyperactivity disorder contributes to neurodevelopmental traits in the general population. *Biol Psychiatry* 2014; **76**: 664–671.
- 19 Groen-Blokhuis MM, Middeldorp CM, Kan K-J, Abdellaoui A, Beijsterveldt CEM van, Ehli EA *et al.* Attention deficit hyperactivity disorder polygenic risk scores predict attention problems in a population-based sample of children. *J Am Acad Child Adolesc Psychiatry* 2014; **53**: 1123–1129.
- 20 Fromer M, Pocklington AJ, Kavanagh DH, Williams HJ, Dwyer S, Gormley P *et al.* De novo mutations in schizophrenia implicate synaptic networks. *Nature* 2014; **506**: 179–184.
- 21 Purcell SM, Moran JL, Fromer M, Ruderfer D, Solovieff N, Roussos P *et al.* A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* 2014; **506**: 185–190.
- 22 Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D *et al.* The contribution of de novo coding mutations to autism spectrum disorder. *Nature* 2014; **515**: 216–221.
- 23 De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Ercument Cicek A *et al.* Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* 2014; **515**: 209–215.
- 24 Szatkiewicz JP, O'Dushlaine C, Chen G, Chambert K, Moran JL, Neale BM *et al.* Copy number variation in schizophrenia in Sweden. *Mol Psychiatry* 2014; **19**: 762–773.
- 25 Kirov G, Pocklington AJ, Holmans P, Ivanov D, Ikeda M, Ruderfer D *et al.* De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Mol Psychiatry* 2011; **17**: 142–153.
- 26 Mick E, Todorov A, Smalley S, Hu X, Loo S, Todd RD *et al.* Family-based genome-wide association scan of attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry* 2010; **49**: 898–905.
- 27 Korn JM, Kuruvilla FG, McCarroll SA, Wysoker A, Nemesh J, Cawley S *et al.* Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat Genet* 2008; **40**: 1253–1260.
- 28 Raychaudhuri S, Korn JM, McCarroll SA, Altshuler D, Sklar P, Purcell S *et al.* Accurately assessing the risk of schizophrenia conferred by rare copy-number variation affecting genes with brain function. *PLoS Genet* 2010; **6**: e1001097.
- 29 Collett D. *Modelling Binary Data*. 2nd (edn). Chapman and Hall/CRC Press: London, UK, 2002.
- 30 Darnell JC, Van Driesche SJ, Zhang C, Hung KYS, Mele A, Fraser CE *et al.* FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* 2011; **146**: 247–261.
- 31 Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R *et al.* The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* 2004; **32**: D258.
- 32 Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* 2012; **40**: D109–D114.
- 33 Bult CJ, Eppig JT, Kadin JA, Richardson JE, Blake JA. The Mouse Genome Database (MGD): mouse biology and model systems. *Nucleic Acids Res* 2008; **36**: D724.
- 34 Croft D, O'Kelly G, Wu G, Haw R, Gillespie M, Matthews L *et al.* Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res* 2011; **39**: D691–D697.
- 35 Schaefer CF, Anthony K, Krupa S, Buchoff J, Day M, Hannay T *et al.* PID: the Pathway Interaction Database. *Nucleic Acids Res* 2009; **37**: D674–D679.
- 36 Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci USA* 2003; **100**: 9440–9445.
- 37 Elia J, Gai X, Xie HM, Perin JC, Geiger E, Glessner JT *et al.* Rare structural variants found in attention-deficit hyperactivity disorder are preferentially associated with neurodevelopmental genes. *Mol Psychiatry* 2010; **15**: 637–646.
- 38 Lichtenstein P, Carlström E, Råstam M, Gillberg C, Anckarsäter H. The genetics of autism spectrum disorders and related neuropsychiatric disorders in childhood. *Am J Psychiatry* 2010; **167**: 1357–1363.
- 39 Martin J, Cooper M, Hamshere ML, Pocklington A, Scherer SW, Kent L *et al.* Biological Overlap of Attention-Deficit/Hyperactivity Disorder and Autism Spectrum Disorder: Evidence From Copy Number Variants. *J Am Acad Child Adolesc Psychiatry* 2014; **53**: 761–770.
- 40 Lozano R, Rosero CA, Hagerman RJ. Fragile X spectrum disorders. *Intractable Rare Dis Res* 2014; **3**: 134–146.
- 41 Ouwenga RL, Dougherty J. Fmrp targets or not: long, highly brain-expressed genes tend to be implicated in autism and brain disorders. *Mol Autism* 2015; **6**: 16.
- 42 Poelmans G, Pauls DL, Buitelaar JK, Franke B. Integrated genome-wide association study findings: identification of a neurodevelopmental network for attention deficit hyperactivity disorder. *Am J Psychiatry* 2011; **168**: 365–377.
- 43 Bralten J, Franke B, Waldman I, Rommelse N, Hartman C, Asherson P *et al.* Candidate genetic pathways for attention-deficit/hyperactivity disorder (ADHD) show association to hyperactive/impulsive symptoms in children with ADHD. *J Am Acad Child Adolesc Psychiatry* 2013; **52**: 1204–1212.
- 44 Cristino AS, Williams SM, Hawi Z, An JY, Bellgrove MA, Schwartz CE *et al.* Neurodevelopmental and neuropsychiatric disorders represent an interconnected molecular system. *Mol Psychiatry* 2014; **19**: 294–301.
- 45 Yang L, Neale BM, Liu L, Lee SH, Wray NR, Ji N. Polygenic transmission and complex neuro developmental network for attention deficit hyperactivity disorder: genome-wide association study of both common and rare variants. *Am J Med Genet B Neuropsychiatr Genet* 2013; **162B**: 419–430.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

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