

# Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <http://orca.cf.ac.uk/71977/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Martin, Joanna, O'Donovan, Michael Conlon, Thapar, Anita, Langley, Kate and Williams, Nigel Melville 2015. The relative contribution of common and rare genetic variants to ADHD. *Translational Psychiatry* 5 (2) , e506. 10.1038/tp.2015.5 file

Publishers page: <http://dx.doi.org/10.1038/tp.2015.5> <<http://dx.doi.org/10.1038/tp.2015.5>>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



## ORIGINAL ARTICLE

## The relative contribution of common and rare genetic variants to ADHD

J Martin<sup>1</sup>, MC O'Donovan<sup>1</sup>, A Thapar<sup>1</sup>, K Langley<sup>1,2</sup> and N Williams<sup>1</sup>

Attention deficit hyperactivity disorder (ADHD) is highly heritable. Genome-wide molecular studies show an increased burden of large, rare copy-number variants (CNVs) in children with ADHD compared with controls. Recent polygenic risk score analyses have also shown that *en masse* common variants are enriched in ADHD cases compared with population controls. The relationship between these common and rare variants has yet to be explored. In this study, we tested whether children with ADHD with ( $N = 60$ ) a large ( $> 500$  kb), rare ( $< 1\%$  frequency) CNV differ by polygenic risk scores for ADHD to children with ADHD without such CNVs ( $N = 421$ ). We also compared ADHD polygenic scores in ADHD children with and without CNVs with a group of population controls ( $N = 4670$ ; of whom  $N = 397$  had CNVs). The results show that children with ADHD with large, rare CNVs have lower polygenic scores than children without such CNVs (odds ratio (OR) = 0.73,  $P = 0.023$ ). Although ADHD children without CNVs had higher scores than controls (OR = 1.18,  $P = 0.0031$ ), this difference was not observed for ADHD children with CNVs (OR = 0.86,  $P = 0.27$ ). These results are consistent with a polygenic liability threshold model of ADHD with both common and rare variants involved.

*Translational Psychiatry* (2015) 5, e506; doi:10.1038/tp.2015.5; published online 10 February 2015

## INTRODUCTION

Attention deficit hyperactivity disorder (ADHD) is a childhood-onset neurodevelopmental disorder characterized by developmentally inappropriate levels of hyperactivity, impulsivity and inattention. Twin studies consistently indicate that it is highly heritable.<sup>1</sup> Molecular genetic studies are beginning to suggest that the genetic architecture of ADHD is complex, with multiple common and rare genetic variants involved.<sup>2,3</sup> Recent molecular genetic studies using a method called 'genomic-relationship-matrix restricted maximum likelihood' as implemented by the software Genome-wide Complex Trait Analysis confirm that additive common genetic variants contribute to estimates of heritability.<sup>4,5</sup> Although, genome-wide association studies (GWAS) thus far have not reported any significantly associated single-nucleotide polymorphisms (SNPs), these studies have been relatively small and underpowered for detecting individual loci significant at a genome-wide level,<sup>2,6</sup> compared with GWAS of adult psychiatric conditions such as schizophrenia.<sup>7</sup> Other factors, including use of 'pseudo-controls' in trio-based analyses<sup>4</sup> and sample heterogeneity (in terms of phenotype, as well as ancestry), may have also had a role. However, an aggregate score (that is, a 'polygenic risk' score) of common genetic variants was found to be, on average, higher in ADHD cases than in controls.<sup>8</sup> For now, the power of this analysis is weak, though it is likely to improve with score alleles derived from future larger discovery GWAS samples, as has been seen in genetic studies of other psychiatric conditions, for example schizophrenia.<sup>7</sup>

In terms of rare variants, several studies have implicated an excess of large ( $> 100$  kb or  $> 500$  kb), rare ( $< 1\%$  population frequency) copy-number variants (CNVs) in ADHD cases compared with controls.<sup>3,5,9</sup> However, only a small proportion of affected

children have such large deletions or duplications and only a few specific loci (for example, duplications at 15q13.3 and 16p13.11) have been found to be significantly enriched in ADHD cases relative to controls.<sup>3,9</sup> Nevertheless, there is evidence that CNVs in children with ADHD are enriched in gene sets that are also enriched for common variant associations, suggesting that common and rare variants disrupt similar biological pathways in ADHD, such as those involved in cholesterol homeostasis or central nervous system development.<sup>2</sup> Also, a pathway analysis of the top SNPs identified in ADHD GWAS has implicated a protein network related to neurite outgrowth, which also appears to be disrupted by several CNVs previously reported in ADHD cases.<sup>10</sup>

A question that has not yet been addressed is how common and rare genetic variants jointly contribute to individual liability. A polygenic liability threshold model of disorder implies that affected individuals with large, rare CNVs (which, in theory, might confer a more substantial amount of trait liability at the level of an individual carrier than can be seen at a population level) should, on average, have lower polygenic risk loading than affected individuals without such CNVs. To address this question, we analysed previously derived<sup>8</sup> polygenic risk scores in a sample of children with ADHD for whom we also had CNV data. Risk scores were based on association data from an independent ADHD case-control GWAS.<sup>6</sup> We hypothesized that children without a CNV would have a higher polygenic risk score for ADHD than those carrying a large and rare deletion or duplication.

## MATERIALS AND METHODS

Our cases were UK-based children with a research diagnosis of ADHD.<sup>2</sup> ADHD was assessed using the Child and Adolescent Psychiatric Assessment,<sup>11</sup> a semi-structured interview with parents, administered by

<sup>1</sup>MRC Centre for Neuropsychiatric Genetics and Genomics, Institute of Psychological Medicine and Clinical Neurosciences, Cardiff University School of Medicine, Cardiff, UK and <sup>2</sup>School of Psychology, College of Biomedical and Life Sciences, Cardiff University, Cardiff, UK. Correspondence: Dr J Martin or Professor A Thapar, MRC Centre for Neuropsychiatric Genetics and Genomics, Institute of Psychological Medicine and Clinical Neurosciences, Cardiff University School of Medicine, Hadyri Ellis Building, Maindy Road, Cardiff CF24 4HQ, UK.

E-mail: martinjm1@cardiff.ac.uk or thapar@cardiff.ac.uk

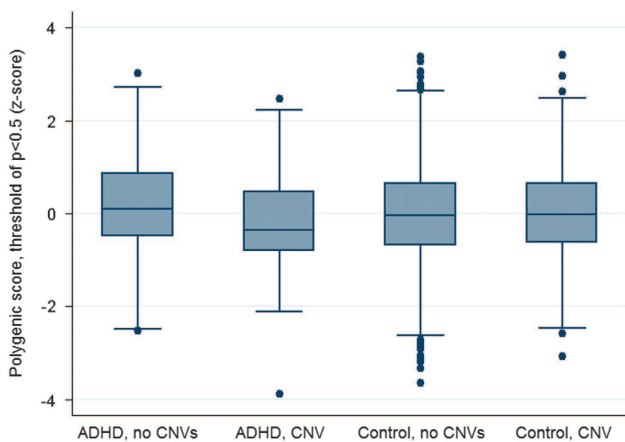
Received 8 December 2014; accepted 19 December 2014

trained research psychologists, under the weekly supervision of a child psychiatrist. Children with comorbid epilepsy, autism spectrum disorder (ASD), neurological or genetic disorders, Tourette's syndrome or bipolar disorder were excluded. Children with comorbid intellectual disability (ID), as defined by an IQ < 70, were included in the sample ( $N=45$ ). Ethical approval was obtained from the North West England and Wales Multicentre Research Ethics Committees and written informed consent/assent was obtained from parents and children. Controls were provided by the Wellcome Trust Case Control Consortium-Phase 2, consisting of 3000 individuals born in the UK during 1 week in 1958 (that is, the 1958 British Birth Cohort) and 3000 individuals from the UK Blood Services collection.<sup>12</sup> Controls were unscreened for ADHD or other psychiatric problems and were not matched with the cases on the basis of any phenotypes as such data were not available.

Quality control of SNP data has been described previously.<sup>2</sup> Analyses were based on the 502 542 SNPs, which passed quality control. The polygenic score and CNV data have been used for previous analyses in this sample; as such, the method is presented in brief, with further information on quality control, CNV calling procedures and polygenic score generation available from the previously published work.<sup>2,3,8</sup> Polygenic scores were calculated in this sample using the Psychiatric Genomics Consortium ADHD GWAS meta-analysis (consisting of 2064 trios, 896 cases and 2455 controls) as the discovery sample.<sup>6</sup> The polygenic scores for our primary analysis were calculated in PLINK<sup>13</sup> for LD-pruned SNPs reaching a  $P$ -value threshold of < 0.5 in the main discovery sample, as previously published.<sup>8</sup> Polygenic scores were additionally calculated at several other thresholds (0.4, 0.3, 0.2, 0.1 and 0.05) using the same method.

As reported previously,<sup>2,3</sup> CNVs > 500 kb were called in the sample using PennCNV and limited to those CNVs which had a population (cases and controls) frequency of < 1%. Primary analyses were limited to these CNVs because they have better concordance across different genotyping platforms, and are determined with greater accuracy and are more robustly associated with neurodevelopmental disorders.<sup>3</sup> Data on CNVs and polygenic scores were available for  $N=481$  children with ADHD, aged 6–17 years (mean = 10.9, s.d. = 2.8), of whom  $N=67$  (13.9%) were female. Control samples consisted of  $N=5081$  individuals with SNP data, of whom  $N=4670$  ( $N=2375$  males and  $N=2295$  females) had both polygenic score and CNV data available. Of note, 12.5% ( $N=60$ ) of the children with ADHD and 8.5% ( $N=397$ ) of the controls had a rare CNV > 500 kb.

The polygenic scores were based on a set of  $N=46\,156$  risk alleles (at threshold  $P < 0.5$ ) and were standardized using z-score transformations. A logistic regression was used to determine whether polygenic risk scores were higher in children with ADHD with CNVs than in those without. Following on from previous work,<sup>8</sup> logistic regressions were also used to test whether higher polygenic scores predicted case or control status, separately for children with ADHD with and without CNVs. In line with previous analyses, two principal components (calculated using the EIGENSTRAT software package and found to have maximum impact on the genomic control inflation factor in the original GWAS analysis) were entered into all the regression analyses as covariates to control for population stratification.<sup>2,8</sup> All analyses were also co-varied for gender. For each association test, the amount of variance explained was calculated as



**Figure 1.** Box plot of mean standardized polygenic score in ADHD cases and controls, with and without CNVs > 500 kb. ADHD, attention deficit hyperactivity disorder; CNV, copy-number variant.

the difference of Nagelkerke pseudo- $R^2$  in the full regression model (including all covariates and polygenic risk score), as compared with the reduced model (containing only the population covariates and gender). All analyses were performed in Stata version 13.1.

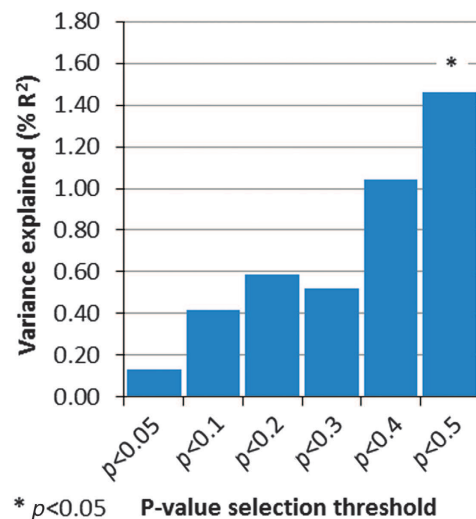
### Secondary analyses

The main analyses (comparing children with ADHD with and without CNVs) were re-run using polygenic scores based on the different thresholds (0.4, 0.3, 0.2, 0.1 and 0.05). The number of SNPs in each of these sets was: 37 078, 28 005, 18 723, 9519 and 4845, respectively. The analyses were also re-run expanding the list of CNVs in cases to any that were > 200 kb ( $N$  children with = 139;  $N$  children without = 342). Additive and multiplicative models of CNV presence and polygenic risk scores were also tested.

### RESULTS

Children with a CNV > 500 kb had a significantly lower polygenic risk score than those without such a CNV (odds ratio (OR) = 0.73, 95% confidence intervals (CI) = 0.55–0.96,  $P=0.023$ ,  $R^2=0.015$ ). A two-tailed permutation test of CNV status in the case group ( $N=481$ ) using 10 000 replicates supported these results ( $P=0.0224$ ,  $SE(P)=0.0015$ ). Figure 1 displays box plots of the polygenic scores of children with ADHD and controls, split by those with and without a CNV > 500 kb. Although we have previously shown that children with ADHD have higher polygenic risk scores for ADHD than population controls,<sup>8</sup> limiting the analysis to children with CNVs ( $N=60$ ) shows no difference between cases and controls (OR = 0.86, 95% CI = 0.67–1.12,  $P=0.27$ ,  $R^2=0.002$ ), although this may be due to the small sample size of children with ADHD with large, rare CNVs. On the other hand, restricting the analysis to children with ADHD without large CNVs ( $N=421$ ) confirms that these children have higher polygenic scores as compared with controls (OR = 1.18, 95% CI = 1.06–1.31,  $P=0.0031$ ,  $R^2=0.003$ ).

Secondary analyses using different  $P$ -value thresholds for calculating polygenic scores demonstrated that significant effects are seen only at the least stringent  $P$ -value threshold ( $P < 0.5$ ), a finding that is consistent with other work based on polygenic score analysis when risk alleles are derived from modest sample sizes.<sup>14</sup> See Figure 2 for the results using all the thresholds. Analyses using the primary threshold for selecting risk alleles ( $P < 0.5$ ) and including smaller CNVs > 200 kb also showed higher polygenic scores in those without CNVs (OR = 0.77, 95% CI = 0.63–0.95,



**Figure 2.** Variance explained ( $R^2$ ) in analysis comparing ADHD children with and without CNVs > 500 kb, using polygenic scores calculated at a variety of different thresholds. ADHD, attention deficit hyperactivity disorder; CNV, copy-number variant.

$P=0.012$ ,  $R^2=0.011$ ). Significant results were additionally attained at more of the risk allele thresholds (see Figure 3). Excluding children with ID ( $N=45$ , of whom  $N=10$  had a large, rare CNV) did not alter the main results ( $OR=0.66$ , 95%  $CI=0.49-0.90$ ,  $P=0.0086$ ,  $R^2=0.023$ ).

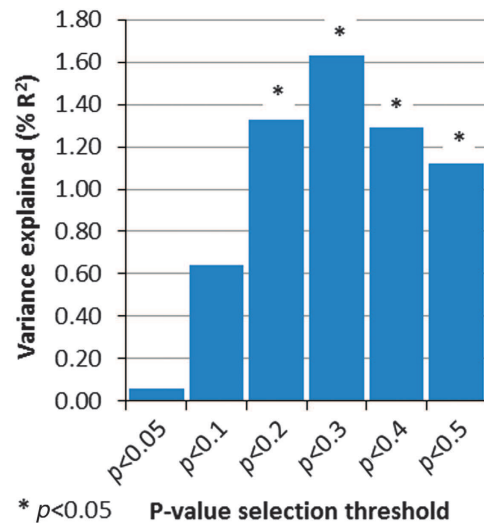
To test an additive model including both common and rare variants, a multivariate logistic regression was run with both ADHD polygenic risk scores and CNV presence as predictors, all the covariates, and with case status as the outcome. Both polygenic scores ( $OR=1.15$ , 95%  $CI=1.04-1.27$ ,  $P=0.009$ ,  $R^2=0.002$ ) and CNV presence ( $OR=1.63$ , 95%  $CI=1.17-2.26$ ,  $P=0.004$ ,  $R^2=0.003$ ) showed independent associations with case status. A multiplicative model adding an interaction term of 'CNV presence  $\times$  polygenic risk score' showed a significant interaction term ( $OR=0.71$ , 95%  $CI=0.51-0.98$ ,  $P=0.036$ ,  $R^2=0.001$ ).

## DISCUSSION

These results imply that children with ADHD with a large, rare CNV (>500 kb or >200 kb) require less of a polygenic burden of common alleles to manifest the disorder. The polygenic risk score loading was relatively low in ADHD children with large, rare CNVs. However, the difference in polygenic loading between them and controls was not significant, although this may have resulted from low power. The results of this study are consistent with a polygenic liability threshold model of ADHD, with both common SNPs and rare CNVs involved in ADHD disease risk. Additive and multiplicative models support these results. Both CNV presence and higher polygenic risk scores were found to be independently associated with case status. Furthermore, the significant interaction term in that multiplicative model supports the main result of this study in that those with a CNV and lower polygenic score are more likely to be cases.

Although these results could be interpreted to suggest that there may be common and rare variant forms of ADHD, it is important to bear in mind that the effect sizes of this and previous studies are quite small and it is not yet clear whether these are truly distinct variants of the disorder. We have previously shown that children with ADHD with a large, rare CNV do not differ clinically (beyond a greater likelihood of ID) or in terms of family history of psychiatric disorders or parental ADHD symptoms from children with ADHD without such CNVs.<sup>15</sup> This suggests that there may be no clear and distinct forms of ADHD but further work examining this issue is needed. Interestingly, CNVs found in children with ADHD appear to disrupt similar biological processes to those hit by top GWAS SNPs,<sup>2,10</sup> which suggests that even if there are different aetiological routes to ADHD, they may act on similar biological mechanisms.

Given the known pleiotropy of genetic risks for various neurodevelopmental, psychiatric and neurological outcomes,<sup>3,4,16,17</sup> it is possible that some individuals with both a pathogenic CNV and a greater than average polygenic loading were not recruited to the ADHD sample by virtue of having another primary neurodevelopmental disorder, which may share aetiology with ADHD (for example, ASD, severe ID or epilepsy) and would have meant that they did not meet study inclusion criteria. These inclusion criteria are typical of ADHD studies and are intended to increase sample homogeneity. The recent publication of the DSM-5 saw the removal of previous diagnostic exclusions relevant to ADHD (for example, an ASD diagnosis excluding a diagnosis of ADHD in DSM-IV). As use of the DSM-5 becomes more widespread, research samples for genetic analyses of childhood disorders, such as ADHD may become more clinically representative. Although studies do find an increased burden of potentially pathogenic CNVs in children with ID and ASD<sup>18</sup> and another study has reported an increased polygenic loading of ASD common variants in an independent sample of children with ASD,<sup>19</sup> it is not yet known whether children with more complex



**Figure 3.** Variance explained ( $R^2$ ) in analysis comparing ADHD children with and without CNVs >200 kb, using polygenic scores calculated at a variety of different thresholds. ADHD, attention deficit hyperactivity disorder; CNV, copy-number variant.

neurodevelopmental problems (for example, those with multiple comorbidities) are more likely to have both types of risk variants (common and rare) than children with ADHD without these comorbidities.

Children with CNVs and high polygenic scores may have been less likely to be ascertained for other reasons. For example, those who are most severely affected may have not agreed to participate, whereas those families in which the genetic loading is highest may also have more chaotic family structures, which is likely to be a barrier for study participation.

Future ADHD studies should aim to determine whether the pattern of results observed in the current study is driven by particular CNV risk loci confirmed to be relevant to ADHD. Current ADHD sample sizes are too small to implicate many specific CNV loci in the disorder.<sup>3,9</sup> Similarly, as ADHD GWAS sample sizes increase, so will the power to robustly detect associated common variants. The current analyses are limited by existing small sample sizes and would therefore benefit from replication in larger future samples.

Like other psychiatric conditions (for example, ASD or schizophrenia), ADHD is an aetiologically complex and highly heritable condition. The results of the present study provide the first empirical molecular genetic support for the hypothesis that risk for ADHD follows a polygenic liability threshold model, in which individuals with large, rare CNVs require less loading of multiple common genetic risk variants for developing ADHD. A similar pattern, where the liability threshold for disorder is shifted downwards, might also be seen for other types of rare genetic mutations and will require future investigation.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## ACKNOWLEDGMENTS

We are very grateful to the children and families and clinicians who participated in this study. This work has been supported by the Medical Research Council (UK), Wellcome Trust and Action Research. We acknowledge the investigators in the original genetic discovery study: Marian Hamshere, Evangelia Stergiakouli, Peter Holmans, Michael Owen, Lindsey Kent and Michael Gill. We also acknowledge the Psychiatric Genomics Consortium (PGC) ADHD group for the published ADHD meta-analysis data used to derive polygenic risk scores.

## REFERENCES

- 1 Faraone SV, Perlis RH, Doyle AE, Smoller JW, Goralnick JJ, Holmgren MA et al. Molecular genetics of attention-deficit/hyperactivity disorder. *Biol Psychiatry* 2005; **57**: 1313–1323.
- 2 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.
- 3 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.
- 4 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.
- 5 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 B Neuropsychiatr Genet* 2013; **162B**: 419–430.
- 6 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.
- 7 Ripke S, Neale BM, Corvin A, Walters JTR, Farh K-H, Holmans PA et al. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014; **511**: 421–427.
- 8 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.
- 9 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.
- 10 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.
- 11 Angold A, Prendergast M, Cox A, Harrington R, Simonoff E, Rutter M. The Child and Adolescent Psychiatric Assessment (CAPA). *Psychol Med* 1995; **25**: 739–753.
- 12 The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007; **447**: 661–678.
- 13 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.
- 14 Purcell S, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 2009; **460**: 748–752.
- 15 Langley K, Martin J, Agha SS, Davies C, Stergiakouli E, Holmans P et al. Clinical and cognitive characteristics of children with attention-deficit hyperactivity disorder, with and without copy number variants. *Br J Psychiatry* 2011; **199**: 398–403.
- 16 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.
- 17 Smoller JW, Craddock N, Kendler K, Lee PH, Neale BM, Nurnberger JI et al. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* 2013; **381**: 1371–1379.
- 18 Girirajan S, Brkanac Z, Coe BP, Baker C, Vives L, Vu TH et al. Relative burden of large CNVs on a range of neurodevelopmental phenotypes. *PLoS Genet* 2011; **7**: e1002334.
- 19 Anney R, Klei L, Pinto D, Almeida J, Bacchelli E, Baird G et al. Individual common variants exert weak effects on risk for autism spectrum disorders. *Hum Mol Genet* 2012; **21**: 4781–4792.



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/>