

## ORIGINAL ARTICLE

## Exome chip analyses in adult attention deficit hyperactivity disorder

T Zayats<sup>1</sup>, KK Jacobsen<sup>1</sup>, R Kleppe<sup>1</sup>, CP Jacob<sup>2</sup>, S Kittel-Schneider<sup>3</sup>, M Ribasés<sup>4,5,6</sup>, JA Ramos-Quiroga<sup>4,5,6,7</sup>, V Richarte<sup>4,5,6,7</sup>, M Casas<sup>4,5,6,7</sup>, NR Mota<sup>8</sup>, EH Grevet<sup>9</sup>, M Klein<sup>10</sup>, J Corominas<sup>10,11</sup>, J Bralten<sup>10</sup>, T Galesloot<sup>12</sup>, AA Vasquez<sup>10,13,14</sup>, S Herms<sup>15,16,17</sup>, AJ Forstner<sup>15,16</sup>, H Larsson<sup>18</sup>, G Breen<sup>19,20</sup>, P Asherson<sup>19,20</sup>, S Gross-Lesch<sup>2</sup>, KP Lesch<sup>2</sup>, S Cichon<sup>16,17,21</sup>, MB Gabrielsen<sup>22,23</sup>, OL Holmen<sup>22</sup>, CHD Bau<sup>8</sup>, J Buitelaar<sup>13</sup>, L Kiemeneij<sup>12</sup>, SV Faraone<sup>1,24</sup>, B Cormand<sup>25,26,27,28</sup>, B Franke<sup>10,14</sup>, A Reif<sup>3</sup>, J Haavik<sup>1,29</sup> and S Johansson<sup>30,31</sup>

Attention-deficit/hyperactivity disorder (ADHD) is a highly heritable childhood-onset neuropsychiatric condition, often persisting into adulthood. The genetic architecture of ADHD, particularly in adults, is largely unknown. We performed an exome-wide scan of adult ADHD using the Illumina Human Exome Bead Chip, which interrogates over 250 000 common and rare variants. Participants were recruited by the International Multicenter persistent ADHD CollaboraTion (IMpACT). Statistical analyses were divided into 3 steps: (1) gene-level analysis of rare variants (minor allele frequency (MAF) < 1%); (2) single marker association tests of common variants (MAF ≥ 1%), with replication of the top signals; and (3) pathway analyses. In total, 9365 individuals (1846 cases and 7519 controls) were examined. Replication of the most associated common variants was attempted in 9847 individuals (2077 cases and 7770 controls) using fixed-effects inverse variance meta-analysis. With a Bonferroni-corrected significance level of 1.82E−06, our analyses of rare coding variants revealed four study-wide significant loci: 6q22.1 locus ( $P=4.46E-08$ ), where *NTSDC1* and *COL10A1* reside; the *SEC23IP* locus ( $P=6.47E-07$ ); the *PSD* locus ( $P=7.58E-08$ ) and *ZCCHC4* locus ( $P=1.79E-06$ ). No genome-wide significant association was observed among the common variants. The strongest signal was noted at rs9325032 in *PPP2R2B* (odds ratio = 0.81,  $P=1.61E-05$ ). Taken together, our data add to the growing evidence of general signal transduction molecules (*NTSDC1*, *PSD*, *SEC23IP* and *ZCCHC4*) having an important role in the etiology of ADHD. Although the biological implications of these findings need to be further explored, they highlight the possible role of cellular communication as a potential core component in the development of both adult and childhood forms of ADHD.

*Translational Psychiatry* (2016) 6, e923; doi:10.1038/tp.2016.196; published online 18 October 2016

## INTRODUCTION

Attention-deficit/hyperactivity disorder (ADHD) is the most commonly diagnosed childhood-onset neuropsychiatric condition,<sup>1</sup> persisting into adulthood in up to 60% of the cases.<sup>2</sup> The adult form of ADHD presents a high risk for developing co-morbid psychiatric conditions, increasing the burden of disease and,

consequently, global impairment, resistance to treatment and costs of illness.<sup>3,4</sup>

ADHD in children and adolescents is highly heritable with an estimated mean heritability of ~76%.<sup>5</sup> Although genome-wide association studies of ADHD show that ~40% of ADHD's heritability can be accounted for by common variants,<sup>6</sup> no

<sup>1</sup>K.G. Jebsen Centre for Neuropsychiatric Disorders, Department of Biomedicine, University of Bergen, Bergen, Norway; <sup>2</sup>Section of Molecular Psychiatry, Clinical Research Unit on Disorders of Neurodevelopment and Cognition Center of Mental Health, University of Würzburg, Würzburg, Germany; <sup>3</sup>Department of Psychiatry, Psychosomatic Medicine and Psychotherapy, University Hospital Frankfurt, Frankfurt am Main, Germany; <sup>4</sup>Psychiatric Genetics Unit, Group of Psychiatry, Mental Health and Addictions, Vall d'Hebron Research Institute, Universitat Autònoma de Barcelona, Barcelona, Spain; <sup>5</sup>Department of Psychiatry, Hospital Universitari Vall d'Hebron, Barcelona, Spain; <sup>6</sup>Biomedical Network Research Centre on Mental Health (CIBERSAM), Madrid, Spain; <sup>7</sup>Departments of Psychiatry and Legal Medicine, Universitat Autònoma de Barcelona, Barcelona, Spain; <sup>8</sup>Department of Genetics, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil; <sup>9</sup>Department of Psychiatry, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil; <sup>10</sup>Department of Human Genetics, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, The Netherlands; <sup>11</sup>Department of Ophthalmology, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, Nijmegen, The Netherlands; <sup>12</sup>Department for Health Evidence, Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, The Netherlands; <sup>13</sup>Department of Cognitive Neuroscience, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, The Netherlands; <sup>14</sup>Department of Psychiatry, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, The Netherlands; <sup>15</sup>Institute of Human Genetics, University of Bonn, Bonn, Germany; <sup>16</sup>Department of Genomics, Life and Brain Center, Bonn, Germany; <sup>17</sup>Department of Biomedicine, University of Basel, Basel, Switzerland; <sup>18</sup>Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; <sup>19</sup>NIHR BRC for Mental Health, Institute of Psychiatry, Psychology and Neuroscience and SLAM NHS Trust, King's College London, London, UK; <sup>20</sup>MRC Social Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK; <sup>21</sup>Institute of Neuroscience and Medicine, Structural and Functional Organization of the Brain (INM-1), Research Center Juelich, Juelich, Germany; <sup>22</sup>K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health, NTNU, Norwegian University of Science and Technology, Trondheim, Norway; <sup>23</sup>Department of Laboratory Medicine, Children's and Women's Health, Norwegian University of Science and Technology, Trondheim, Norway; <sup>24</sup>Departments of Psychiatry and of Neuroscience and Physiology, SUNY Upstate Medical University, Syracuse, NY, USA; <sup>25</sup>Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Universitat de Barcelona, Barcelona, Spain; <sup>26</sup>Centro de Investigación Biomédica Anchor en Red de Enfermedades Raras, Barcelona, Spain; <sup>27</sup>Institut de Biomedicina de la Universitat de Barcelona, Barcelona, Spain; <sup>28</sup>Institut de Recerca Pediàtrica HosAnchorpital Sant Joan de Déu, Barcelona, Spain; <sup>29</sup>Division of Psychiatry, Haukeland University Hospital, Bergen, Norway; <sup>30</sup>K.G. Jebsen Centre for Neuropsychiatric Disorders, Department of Clinical Science, University of Bergen, Bergen, Norway and <sup>31</sup>Center for Medical Genetics and Molecular Medicine, Haukeland University Hospital, Bergen, Norway. Correspondence: Dr T Zayats or Professor S Johansson, K.G. Jebsen Centre for Neuropsychiatric Disorders, Department of Biomedicine, University of Bergen, Jonas Lies vei 91, Bergen 5009, Norway. E-mail: tetyana.zayats@uib.no or stefan.johansson@uib.no

genome-wide significant ( $P < 5.00E-08$ ) single-nucleotide variants (SNVs) have yet been identified. Meta-analyses of some candidate genes and genome-wide association results have yielded interesting hypotheses for this disorder, with pathway analyses pointing towards neurodevelopmental networks.<sup>1</sup>

The mixed results of the previous association studies may be partially due to the polygenic and multifactorial nature of ADHD, with both common and rare variants likely contributing to its etiology.<sup>7,8</sup> Phenotypic heterogeneity could be another explanation for the multifarious findings in ADHD genetics. Such heterogeneity could occur when various smaller sets are combined to increase sample size and it may lead to the loss of a real association signal.<sup>9</sup> Examination of a refined phenotype may reduce phenotypic heterogeneity and increase the probability of detecting a true association signal, although adequately large sample size will still be required to detect associations of small effects.

Given that ADHD is a developmentally heterogeneous condition that covers both a remitting and a persistent adult type of ADHD,<sup>1</sup> symptom persistence may represent one promising refined phenotype.<sup>10</sup> As most molecular genetic studies on ADHD so far have been limited to its childhood form only, this might constrain the possibilities to identify genetics of symptom persistence.

In this study, we investigated the role of common and rare exonic polymorphisms in the persistent, adult form of ADHD (aADHD). The focus on low frequency protein-coding variants in addition to the genome-wide common polymorphisms provides an additional source of potentially relevant variation that has received little attention in aADHD genetic studies so far.

## MATERIALS AND METHODS

### Subjects

All adult ADHD cases examined in this study were volunteers enrolled through the International Multicenter persistent ADHD CollaboraTion (IMpACT). All patients were diagnosed with ADHD according to DSM-IV criteria, meaning the presence of ADHD symptoms before 7 years of age. Controls were recruited either at an IMpACT site (Brazil, Germany, The Netherlands, UiB Norway and Spain) or through population studies (Germany, The Netherlands, HUNT Norway and the United Kingdom). All subjects were of European descent who was ensured by means of self-report together with multi-dimensional scaling (Supplementary Appendix 1). In Brazil, only Native Brazilians of European descent were recruited by including only those who reported having grandparents of European descent and by performing a morphological classification based on skin color and morphological traits rather than relying on self-classification only.<sup>11</sup> The population in Southern Brazil where this study was performed is mainly of European descent and population stratification was not found in the European-derived population of Rio Grande do Sul.<sup>12,13</sup> Individuals from Southern Brazil show predominantly European ancestry (94%) according to estimates of interethnic admixture.<sup>13,14</sup>

All participants provided signed informed consent in accordance with the Declaration of Helsinki. The study was approved by ethics committees in each collaborating country at the corresponding recruitment center: in Germany, by the Ethics Committee at the University of Würzburg (Würzburg, Germany); in Norway, by the regional Ethics Committee for medical and health research ethics, western Norway at the University of Bergen (Bergen, Norway) and the Norwegian Data Inspectorate, the Regional and National Committees for Medical and Health Research Ethics and the Norwegian Directorate of Health at the University of Trondheim (Trondheim, Norway); in Spain, by the Ethics Committee of the Hospital Universitari Vall d'Hebron (HUVH, Barcelona, Spain) at the Department of Psychiatry from the Hospital Universitari Vall d'Hebron (HUVH, Barcelona, Spain); in the Netherlands, by the regional Ethics Committee (Centrale Commissie Mensgebonden Onderzoek: CMO Regio Arnhem—Nijmegen; Protocol number Ill.04.0403), the Institutional Review Board of the Radboud University Medical Center (Nijmegen, the Netherlands) and by the regional Ethics Committee (Centrale Commissie Mensgebonden Onderzoek: CMO Regio Arnhem Nijmegen; 2008/163; ABR: NL23894.091.08) at the outpatient clinic of GGZ Delfland in

Delft, to Parnassia, psycho-medical centre in The Hague or at the department of Psychiatry at the Radboud University Nijmegen Medical Centre in Nijmegen, the Netherlands.; in Brazil, by the Ethics Committee of the Hospital de Clínicas de Porto Alegre (Porto Alegre, Brazil) at the ADHD Outpatient Program at the Hospital de Clínicas de Porto Alegre (HCPA); in UK, by the National Research Ethics Committee at the National Adult ADHD Outpatient Clinic at the South London and Maudsley NHS Trust, (London, UK) and the Wellcome Trust Case Control Consortium (WTCCC).

For the discovery phase of the study, samples were available from four European IMpACT sites: Germany, The Netherlands, Norway and Spain. A replication sample, to examine the most significant SNVs from the common variant analysis identified in the discovery stage, was recruited from the following four IMpACT sites: Brazil, Germany, The Netherlands and the UK. Additional samples were recruited through the WTCCC. A detailed description of all samples is provided in the Supplementary Appendix 1.

### Genotyping, genotype calling and quality control

All subjects included in the discovery stage were genotyped on the InfiniumHumanExome array (Illumina, San Diego, CA, USA). The German cases, the full Dutch sample and the full Spanish sample were genotyped on HumanExome-12v1-1\_A, the German controls and the HUNT Norwegian sample of controls were genotyped on HumanExome-12v1\_A and the UiB Norwegian sample was genotyped on HumanExome-12v1\_B version of the chip. All genotypes were processed using the Illumina GenomeStudio V2011.1 software, with additional genotype assignments implemented in the zCall.<sup>15</sup> Best practice guidelines were used to perform quality control (QC) of genotype calls in GenomeStudio.<sup>16</sup> Further QC was carried out on all data sets using PLINK,<sup>17</sup> with the following steps: (1) genotyping rate threshold was set to 98% both for individuals and for SNVs, (2) Hardy-Weinberg test threshold was set to  $P < 1.00E-05$ , (3) overall heterozygosity of individuals was screened based on common (minor allele frequency (MAF)  $\geq 1\%$ ) and rare (MAF  $< 1\%$ ) SNVs separately, with outliers defined as those outside the range of  $\text{mean} \pm 3\text{s.d.}$  of the total heterozygosity observed in a sample, (4) relatedness (PI\_HAT) threshold was set to 10% and (5) ethnic homogeneity was ensured by means of multi-dimensional scaling with HapMap3 populations. Genotype calling and all QC steps were performed for each data set individually. Those samples that were collected in the same country (namely German cases and German controls, Dutch cases and Dutch controls, Norwegian UiB and HUNT samples) were merged in PLINK and additional QC steps were implemented. Specifically, the screening for heterozygosity, cryptic relatedness and population outliers was performed once more as described above. As the end result, four data sets were produced: a combined German sample, a combined Dutch sample, a combined Norwegian sample and a Spanish sample.

For the replication stage, the most significant three common (MAF  $> 1\%$ ) SNVs were genotyped by KASP assay (LGC Genomics, UK) in IMpACT samples (Brazil, Germany, The Netherlands and the UK). The Dutch replication sample was also genotyped on Perlegen (Perlegen Sciences) and HumanCytoSNP-12 array (Illumina).

### Statistical analyses

Statistical analyses were divided into three main stages: (1) examination of rare coding SNVs (MAF  $< 1\%$ ), (2) examination of common SNVs (MAF  $\geq 1\%$ ) with replication in an independent sample and (3) exploration of biological pathways possibly involved in the development of adult ADHD based on the results from the previous two steps. Prior to the analyses, each subject's genetic substructure characteristics were estimated with principal components analysis implemented in EIGENSTRAT software for each data set individually.<sup>18</sup> Calculation of principal components was performed based only on individuals and SNVs revealing high genotyping rate ( $\geq 99\%$ ) and common variants (MAF  $\geq 1\%$ ) after removal of strand ambiguous SNVs and those in high LD ( $r^2 < 0.2$ ). Long stretches of LD were also removed prior to calculation of principal components.<sup>19</sup>

### Examination of rare coding variants

Rare variants were defined as those with MAF  $< 1\%$ . The variants were combined per gene and tested for association with adult ADHD in RAREMETAL.<sup>20</sup> The software first analyzes each data set individually and then combines the generated summary statistics across all the data (meta-analysis). The individual analyses generated single-variant score statistics

as well as covariance matrices to reflect linkage disequilibrium (LD) among the variants. We performed two meta-analyses at the gene level: the MAF-weighted burden test and the variable threshold test (VT) across all data sets. Only those genes were tested in which at least two variants were observed across all data sets. All individual tests were adjusted for the number of significant principal components as calculated in EIGENSTRAT. Variants of those genes that reached a  $P$ -value below  $1.00E-05$  in the individual association tests were manually evaluated in GenomeStudio in the German, Dutch and Norwegian data sets. GenomeStudio data was not available for the Spanish data set. Single-nucleotide polymorphisms displaying less than optimal clustering were excluded. Bonferroni correction was applied to adjust for multiple testing in the final step of the analyses.

Significant association signals, revealed by these analyses of rare variants, were also examined in the context of association among common variants in meta-analysis of this study as well as in that of Psychiatric Genomics Consortium (PGC).

#### Examination of common variants

Common SNVs were defined as those with  $MAF \geq 1\%$ . The variants were tested in PLINK software, assuming an additive model. Each data set was first analyzed separately, including significant EIGENSTRAT principal components as covariates. The results were then combined with the use of fixed-effects inverse variance meta-analysis in METAL,<sup>21</sup> applying genomic control.<sup>22</sup> Only those SNVs present in at least two data sets were meta-analyzed. QQ-plots were constructed to ensure proper distribution of observed statistics.

Replication was attempted for the top three SNVs detected in the discovery phase. In the Brazilian, Dutch and German samples, the association of each top variant was assessed in PLINK with logistic regression assuming an additive model. The association in the British sample was assessed by the chi-square test after constructing  $2 \times 2$  contingency tables of allele counts, as we did not have access to raw genotype data of the controls. Each data set was analyzed individually. The results were combined in the fixed-effects inverse variance meta-analysis implemented in PLINK. The three top common variants were also looked up in the largest published ADHD genome-wide association meta-analysis of PGC.

#### In silico functional analyses of the significant loci

We performed a comprehensive assessment of all rare SNVs contributing to the association signals that survived the study-wide correction for multiple testing to explore their potential functional impact. Such effect was evaluated in two ways: (1) influence of the genetic variation on the encoded protein and (2) regulatory DNA effects (sequences which regulate gene transcription, gene expression and/or DNA replication). The severity of missense SNVs on the function of the encoded proteins was assessed with sequence alignment, using protein blast and protein homology modeling. Applying homology models of human *NTSDC1*, the structural details surrounding the sites of missense mutations were investigated in Discovery studio Visualizer (v 4.1; Illumina). In addition, SIFT and PolyPhen predictions were also carried out in Ensembl Variant Effect Predictor (VEP).

The possible regulatory effects were assessed in RegulomeDB, utilizing genomic annotation databases that employ the data curated by the Encyclopedia of DNA Elements (ENCODE)<sup>24</sup> and provide information on functionally important elements. RegulomeDB identifies DNA features and regulatory elements of the human genome.<sup>23</sup>

#### Exploration of biological pathways: GO terms

To evaluate whether any biological processes may be implicated by our results, we performed gene-set analyses using gene ontology (GO) terms. The analyses were performed in MAGMA.<sup>24</sup> First, a degree of association was calculated for each gene based on both common and rare variants by the use of principal component regression. Rare variants were defined as those with  $MAF < 1\%$  and a burden score was created for each gene, computed as weighted sum of all rare variants in that gene. All data sets were analyzed individually, adjusted for significant EIGENSTRAT components. Fixed effects meta-analysis was then performed, using the square root of the sample sizes as weights. To evaluate each gene's contribution to the examined gene-sets (GO term), the  $P$ -value of each gene from the meta-analysis was converted to a  $Z$ -value and used as an outcome variable for a regression model with gene-set membership as a predictor. Gene size, gene-sets' gene density as well as global LD were added as covariates

to adjust for possible confounding effects and prevent spurious association. We restricted our pathway analysis only to those sets that contained SNVs in at least 10 genes in our total aADHD data (all four data sets). Bonferroni correction was applied to adjust for multiple testing.

#### Gene expression in brain tissue

Gene expression in the human brain across the lifetime was assessed for the loci surviving the Bonferroni correction. Data was accessed from the Human Brain Transcriptome project (<http://hbatlas.org>), the Expression Atlas (<https://www.ebi.ac.uk/gxa/home>) and Allen Brain Atlas (<http://www.brain-map.org/>).

## RESULTS

### Subjects and genotyping

After QC, 9365 individuals (1846 adult ADHD cases and 7519 controls) were available for the analyses. Supplementary Appendix 1 and Table 1 present the details of the four discovery data sets.

### Examination of rare coding variants

As it is still unknown what is the most powerful statistical approach for testing whether a mixture of neutral, risk and protective variants is present within a single gene,<sup>25</sup> we leaned on the view of rare functional variants being under selective pressure and utilized two methods to explore such scenario: (1) the MAF-weighted burden test that defines rare variants with a fixed threshold and places weights inverse to the MAFs of the variants within a gene and (2) the VT that empirically determines the MAF threshold and makes no prior assumption about the relationship between MAFs and effect sizes of the variants within a gene. Out of 14 431 genes with observed variants, 13 715 genes contained at least two variants. We adjusted our significance level for the use of two methods and the 13 715 genes containing more than two rare SNVs (27 430 tests in total), which yielded a Bonferroni-corrected study-wide significance threshold of  $1.82E-06$ . Four loci survived this correction for multiple testing. The most significant association was observed using the VT method, present at 6q22.1 ( $P = 4.46E-08$ , MAF threshold = 0.00027), where the 5'-nucleotidase domain containing 1 (*NTSDC1*) gene resides along with the collagen type X alpha 1 (*COL10A1*) gene. This locus was also the most significant finding using the MAF-weighted burden test (ST1). Supplementary Table 2 presents details of the variants contributing to association signal in 6q22.1 locus.

The three additional study-wide significant associations were observed for the SEC23 interacting protein gene (*SEC23IP*;  $P = 6.47E-07$ , VT method), the pleckstrin and Sec7 domain containing gene (*PSD*;  $P = 7.58E-07$ , MAF-weighted burden test)

**Table 1.** Properties of the IMpACT discovery samples examined in this study

<i>Individuals</i>			
IMpACT site	Cases	Controls	Total
Germany	340	2.286	2.626
Netherlands	294	1.703	1.997
Norway	597	2.598	3.195
Spain	615	932	1.547
Total	1.846	7.519	9.365
<i>Variants</i>			
IMpACT site	Rare (MAF < 1%)	Common (MAF ≥ 1%)	Total
Germany	71.350	33.053	104.403
Netherlands	59.415	33.336	92.751
Norway	60.649	33.870	94.519
Spain	62.559	31.564	94.123

Abbreviation: MAF, minor allele frequency.

and zinc finger, CCHC domain containing 4 gene (*ZCCHC4*,  $P=1.79E-06$ , VT method). ST1 summarizes the most significant genes ( $P \leq 1.00E-05$ ) observed in the performed gene-based analyses of rare coding variants.

These four study-wide significant loci revealed no compelling association signals among single point tests of common variants in this study nor in PGC (ST2 and ST3).

#### Examination of common variants

The most significant signal in our meta-analysis of common variants was observed at rs9325032 in the *PPP2R2B* gene (odds ratio=0.81,  $P=1.61E-05$ ). No variant reached genome-wide significance ( $P < 5.00E-08$ ). ST3 details the top SNVs with association  $P$ -values  $< 1.00E-03$  in the meta-analysis. SF1 depicts the QQ-plots.

Replication was attempted for the top three common SNVs (ST4). No variant reached genome-wide significance (SF2). However, rs9325032 (*PPP2R2B* gene) showed a modest trend of significance ( $P=0.033$ , ST4A), although significant heterogeneity (for example, differences in estimated MAFs) and discordant direction of effect was observed among samples (ST4A). Comparisons using the data for childhood ADHD collected by the PGC<sup>26</sup> revealed no significant findings for the top SNVs in the current study (ST4B).

#### In silico functional analyses of the significant loci

In total, 32 rare coding SNVs contributed to the gene-based association signals passing study-wide Bonferroni correction in the four loci (6q22.1, *SEC23IP*, *PSD* and *ZCCHC4*). No single common variant reached genome-wide significance and, thus, common variants were not assessed further for functionality. First, we evaluated the effect of rare missense variants on the encoded protein in the most significant locus of *NT5DC1*. The *NT5DC1* protein contains a known 5'-nucleotidase domain (aa11–337) and a haloacid dehalogenase-like domain (aa198–324). Table 2 summarizes the results of sequence alignment using protein blast and protein homology modeling (template NP\_689942.2) to assess the severity of the observed *NT5DC1* missense variants in this locus. Structural inspection of the SNVs in the *NT5DC1* locus revealed their potential to form a possible phosphorylation site (rs150257749) and to induce conformational change (rs150293032) (Table 2 and SF3B), both of which may influence the function of the protein. A homology model of human *NT5DC1* and structural inspection of the three missense mutation sites in human *NT5DC1* are presented in SF3. The *NT5DC1* sequence alignment is shown in SF4.

While examining the impact on their respective encoded proteins across all the 32 variants, six were predicted to be deleterious by both PolyPhen and SIFT: one in *NT5DC1* (rs150293032), two in *PSD* (rs142273937 and rs200819772), two in *SEC23IP* (rs142665854 and rs142266445) and one in *ZCCHC4* (rs151252286). Detailed features of these missense variants are presented in ST5.

Apart from surveying the possible direct effects of the variants on their encoded proteins, we also explored their regulatory potential using RegulomeDB. These analyses revealed that the SNVs with the most likely regulatory effect reside within the *PSD* gene (rs147203944, rs148732359 and rs142273937) and probably affect the binding properties of polymerase (RNA) II (DNA directed) polypeptide A (POLR2A) as well as may alter CTCF regulatory element binding (rs140739855 and rs200819772) (ST6). In addition, the binding of POLR2A and CTCF may also be affected by the variants in the other two study-wide significant loci: rs73357833 in *SEC23IP*, rs201763036 in *ZCCHC4* (ST6). No significant eQTL effects were observed. Detailed RegulomeDB features of all variants are presented in ST6.

**Table 2.** Summary of effect severity of the rare coding missense variants in the *NT5DC1* gene with observed minor allele frequency below the established critical threshold

SNV	Mutation	AA sequence Conservation	Predicted structure	Comments
rs201990931	V84I	Conserved	$\beta$ -sheet (end)	Part of a hydrophobic cluster, but Ile is likely to fit well
rs150257749	N151S	N(59)/S(7)/D(9)/-(3)	Loop	Possible H-bond with K154, conserved with S. Possible formation of phosphorylation site
rs150293032	G302D	G(73)/S(6)/N(2)	Loop	Neighboring basic amino acid—D may cause conf. change. T301 plausible phosphorylation site (PKA/PKG/MSK)

Abbreviation: SNV, single-nucleotide variants. The severity of missense SNVs on the function of the encoded protein was assessed with sequence alignment, using protein BLAST and protein homology modeling (SF3). The mutations investigated are shown in column two along with the amino acid (AA) conservation based on the sequence alignment (SF4). Single letter amino acid codes are used; valine (V), isoleucine (I), asparagine (N), serine (S), aspartic acid (D), glycine (G) and threonine (T).

### Exploration of biological pathways: GO terms

To evaluate whether certain biological processes may be implicated by our results, we performed gene-set analyses using GO terms. Among the GO terms, 1,844 terms contained data for 10 or more genes in our data set. This brings the Bonferroni-corrected significance threshold to a  $P$ -value of  $2.17E-05$ . The strongest association was observed for 'mRNA 3'-end processing (GO:0031124 term,  $P=1.07E-04$ ). ST7 summarizes the top GO terms ( $P \leq 1.00E-03$ ) observed in the meta-analysis of both common and rare variants.

### Gene expression in brain tissue

The expression of the four study-wide significant loci (*NT5DC1*, *SEC23IP*, *PSD* and *ZCCHC4*) was evaluated throughout life in the Human Brain Transcriptome atlas. The transcriptional trajectories of all four loci revealed variability between pre- and post-natal stages, in line with the general patterns of gene expression in the brain across the lifetime.<sup>27</sup> Supplementary Figure 6 depicts the expression levels of these loci throughout life.

We have also examined the expression patterns of the four aforementioned loci with regards to the brain regions. *NT5DC1* revealed the highest expression in substantia nigra as well as other midbrain and hindbrain structures (FANTOM5 and GTEx projects, adult expression levels). The other three loci showed high expression in the cortex and basal ganglia input nuclei (*ZCCHC4*, FANTOM5 data), and throughout brain (*SEC23IP*, *PSD*, mouse data in Allen Brain Atlas).

## DISCUSSION

This study aimed to shed light to the genetic architecture of adult ADHD, the form of this disorder that has received relatively little attention so far. Our main findings are the novel, study-wide significant candidate loci for adult ADHD at 6q22.1, where *NT5DC1* and *COL10A1* reside, as well as the *SEC23IP*, *PSD* and *ZCCHC4* loci.

*NT5DC1* encodes a member of the haloacid dehalogenase superfamily of enzymes, the exact physiological role of which is still largely unknown.<sup>28</sup> On the basis of its sequence similarity to other haloacid dehalogenases, *NT5DC1* may be involved in the dephosphorylation of intracellular signaling molecules and our *in silico* analyses of the SNVs contributing to its association signal suggest that these variants may alter its function (Table 2). The *NT5DC1* locus, however, also contains a shorter *COL10A1* gene that is embedded in reverse orientation in *NT5DC1* intron 6 (SF5). With our sample size it is not possible to statistically distinguish the effects of these two genes from each other, but the strongest association is seen when testing the larger *NT5DC1* gene ( $P=4.48E-08$ ).

Similarly to 6q22.1 locus, the protein encoded by *SEC23IP* belongs to a family of enzymes as well: intracellular phospholipase A<sub>1</sub> family that degrades phospholipids and is involved in membrane trafficking.<sup>29</sup> Altogether with actin cytoskeleton organization, membrane trafficking is also a mechanism through which *PSD*-encoded protein (a guanine nucleotide exchange factor) mediates a number of neuronal functions,<sup>30</sup> including dendritic spine formation and stabilization.<sup>31</sup> Furthermore, *ZCCHC4*-encoded zinc-finger DNA-binding protein may also be linked to cell signal transduction as suggested epigenetic silencer of *RAS* gene family.<sup>32,33</sup> *RAS* genes reportedly inhibit apoptosis,<sup>33</sup> a physiological cell death by which unsuccessful neurons are deleted from the central nervous system in the developing brain.<sup>34</sup>

Thus, in this study, we noted association signals in several genes involved in signal transduction, highlighting the possible role of cellular communication in the development of adult ADHD. In line with these observations, previous studies have implicated all four of our study-wide significant loci in disorders of the central

nervous system, including ADHD and related neurodevelopmental disorders, ADHD co-morbid conditions as well as syndromes where ADHD symptoms are present.<sup>35-51</sup> In addition, paralogs of the *NT5DC1* gene have also been reported to have role in the development of ADHD and other neuropsychiatric disorders.<sup>52-54</sup> These findings fit with the higher transcript level of *NT5DC1* reported in substantia nigra, a major localization of dopaminergic neurons.

Despite being the largest genetic study on adult ADHD, it is still likely to be underpowered to detect variants with very small effect sizes. Thus, no genome-wide significant signal was observed for common polymorphisms (MAF  $\geq 1\%$ ). However, gene-based statistical methods to analyze rare variants combined could potentially improve power.<sup>55</sup> These aggregation tests are based on various assumptions about the underlying genetic architecture and their power depends on the true, and, in case of ADHD, unknown, disease model.<sup>55</sup> In this study, we used two methods to detect gene-based association: (1) the VT method that leans on the view of functional allelic variants being a subject to purifying selection pressure and empirically calculates a threshold below which the variants are more likely to be functional, based on the observed data; and (2) MAF-weighted burden test, which makes the same assumption of functional variants being under selective pressure as the VT method. As all the four study-wide significant loci reveal association signals with both methods (ST1B), it may suggest that the results appear to be robust with regard to different gene-based tests operating under the same hypothesis and that such applications may help us in uncovering underlying genetic architecture of such complex disorders as ADHD.

Taken together, all four study-wide significant loci have been implicated in the function of neural circuitry and communication, cellular mechanisms previously linked to the development of mental disorders,<sup>56</sup> including childhood ADHD.<sup>1</sup> Thus, dysregulation of cellular communication could be a core component in the development of both adult and childhood forms of ADHD. Nonetheless, these findings should be subjected to further examination in larger samples before their role in ADHD can be firmly established.

## CONFLICT OF INTEREST

JH has received speaker fees from Lilly, Novartis and Janssen-Cilag. HL has served as a speaker for Eli-Lilly and Shire and has received a research grant from Shire. BF has received a speaker fee from Merz. In the past year, SVF received consulting income, travel expenses and/or research support from Ironshore, Arbor, Shire, Akili Interactive Labs, Alcobra, VAYA, and Neurovance and research support from the National Institutes of Health (NIH). His institution is seeking a patent for the use of sodium-hydrogen exchange inhibitors in the treatment of ADHD. In previous years, he received consulting fees or was on Advisory Boards or participated in continuing medical education programs sponsored by: Shire, Alcobra, Otsuka, McNeil, Janssen, Novartis, Pfizer and Eli-Lilly. SVF receives royalties from books published by Guilford Press: Straight Talk about Your Child's Mental Health and Oxford University Press: Schizophrenia: The Facts. The remaining authors declare no conflict of interest.

## ACKNOWLEDGMENTS

We are grateful to the participants of the study. Special acknowledgment is due to Benjamin Neale and Christiaan de Leeuw for invaluable insight and suggestions regarding the presented data. We thank the UK Exome Chip consortium for access to control the data. This work was supported by the K.G. Jebsen Foundation for Medical Research, University of Bergen, the Western Norwegian Health Authorities (Helse Vest), as well as European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no 602805 and from the European Community's Horizon 2020 Programme (H2020/2014-2020) under grant agreement no 643051 (MiND). The Nord-Trøndelag Health Study (The HUNT Study) is a collaboration between the HUNT Research Centre (Faculty of Medicine, Norwegian University of Science and Technology NTNU), the Nord-Trøndelag County Council, the Central Norway Health Authority and the Norwegian Institute of Public Health. This work was supported by the European College of Neuropsychopharmacology (Grant reference:

ECNP Networks). This study is part of the International Multicentre persistent ADHD Collaboration (IMPACT). IMPACT unites major research centers working on the genetics of ADHD persistence across the lifespan and has participants in The Netherlands, Germany, Spain, Norway, the United Kingdom, the United States, Brazil and Sweden. Principal investigators of IMPACT are: Barbara Franke (chair), Andreas Reif (co-chair), Stephen V. Faraone, Jan Haavik, Bru Cormand, Antoni Ramos-Quiroga, Philip Asherson, Klaus-Peter Lesch, Jonna Kuntsi, Claiton Bau, Jan Buitelaar, Stefan Johansson, Henrik Larsson, Alys Doyle and Eugenio Grevet. IMPACT-NL: The Dutch IMPACT study is supported by grants from the Netherlands Organization for Scientific Research (NWO), i.e. the NWO Brain & Cognition Excellence Program (grant 433-09-229) and a Vici grant to BF (grant 016-130-669), and by grants from the Netherlands Brain Foundation (grant 15F07[2]27) and BBMRI-NL (grant CP2010-33). The research leading to these results also received funding from the European Community's Seventh Framework Programme (FP7/2007 – 2013) under grant agreements no. 602805 (Aggrosotype) and no. 602450 (IMAGEMEND), and from the European Community's Horizon 2020 Programme (H2020/2014–2020) under grant agreement no. 643051 (MiND). In addition, the work was supported by a grant for the ENIGMA Consortium (grant number U54 EB020403) from the BD2K Initiative of a cross-NIH partnership. NeuroIMAGE: This study used the sample from the NeuroIMAGE project. NeuroIMAGE was performed between 2009 and 2012, and is the follow-up study of the Dutch part of the International Multisite ADHD Genetics (IMAGE) project, a multi-site, international effort. Funding support for the IMAGE project was provided by NIH grants R01MH62873 and R01MH081803 to Dr Faraone. This work was further supported by an NWO Large Investment Grant 1750102007010 and NWO Brain & Cognition an Integrative Approach grant (433-09-242) (to Dr Buitelaar), and grants from Radboud University Nijmegen Medical Center, University Medical Center Groningen and Accare, and VU University Amsterdam. NeuroIMAGE also receives funding from the European Community's Seventh Framework Programme (FP7-/2007–2013) under grant agreement no. 278948 (TACTICS), no. 602805 (Aggrosotype) and no. 602450 (IMAGEMEND). Nijmegen Biomedical Study (NBS): The NBS is a population-based survey conducted at the Department for Health Evidence, and the Department of Laboratory Medicine of the Radboud university medical center. Principal investigators of the Nijmegen Biomedical Study are Lambertus Kiemeneij, André Verbeek, Dorine Swinkels and Barbara Franke. The NBS data are managed by the NBS project team and are available upon request; see [www.nijmegenbiomedischestudie.nl](http://www.nijmegenbiomedischestudie.nl) for an overview of the data available in this study. Current practical coordinator of the NBS is Dr T.E. Galesloot. Readers can contact her to request the data ([Tessel.Galesloot@radboudumc.nl](mailto:Tessel.Galesloot@radboudumc.nl)). This research was financially supported by BBMRI-NL, a Research Infrastructure financed by the Dutch government (NWO 184.021.007). SVF is supported by the K.G. Jebsen Centre for Neuropsychiatric Disorders, University of Bergen, Bergen, Norway, the European Community's Seventh Framework Programme (FP7/2007–2013) under grant agreement no. 602805 and NIMH grant R01MH094469. BC was supported by the Spanish Ministerio de Economía y Competitividad (SAF2012-33484, SAF2015-68341-R), AGAUR (2014SGR932), European Community's Seventh Framework Programme (under grant agreement 602805, Aggrosotype) and European Community's H2020 Programme (under grant agreements 643051, MiND, and 667302, CoCA). MR is a recipient of a Miguel de Servet contract from the Instituto de Salud Carlos III, Spain (CP09/00119). This work was funded by Instituto de Salud Carlos III (PI11/00571, PI11/01629, PI12/01139, PI14/01700), and co-financed by the European Regional Development Fund (ERDF), Agència de Gestió d'Ajuts Universitaris i de Recerca-AGAUR, Generalitat de Catalunya (2014SGR1357) and European Community's H2020 Programme (under grant agreement no. 643051, MiND).

## REFERENCES

- Faraone SV, Asherson P, Banaschewski T, Biederman J, Buitelaar J, Ramos-Quiroga JA et al. Attention-deficit/hyperactivity disorder. *Nat Rev Dis Primers* 2015; **1**: 15020.
- Faraone SV, Biederman J, Mick E. The age-dependent decline of attention deficit hyperactivity disorder: a meta-analysis of follow-up studies. *Psychol Med* 2006; **36**: 159–165.
- Biederman J. Impact of comorbidity in adults with attention-deficit/hyperactivity disorder. *J Clin Psychiatry* 2004; **65**: 3–7.
- Asherson P. Clinical assessment and treatment of attention deficit hyperactivity disorder in adults. *Expert Rev Neurother* 2005; **5**: 525–539.
- Nikolas MA, Burt SA. Genetic and environmental influences on ADHD symptom dimensions of inattention and hyperactivity: a meta-analysis. *J Abnorm Psychol* 2010; **119**: 1–17.
- Cross-Disorder Group of the Psychiatric Genomics C, Lee SH, Ripke S, Neale BM, Faraone SV, Purcell SM et al. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet* 2013; **45**: 984–994.
- 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.
- Williams NM, Franke B, Mick E, Anney RJ, 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.
- 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.
- Faraone SV, Biederman J, Spencer T, Wilens T, Seidman LJ, Mick E et al. Attention-deficit/hyperactivity disorder in adults: an overview. *Biol Psychiatry* 2000; **48**: 9–20.
- Polina ER, Rovaris DL, de Azeredo LA, Mota NR, Vitola ES, Silva KL et al. ADHD diagnosis may influence the association between polymorphisms in nicotinic acetylcholine receptor genes and tobacco smoking. *Neuromol Med* 2014; **16**: 389–397.
- Salzano FM, Bortolini MC. *The Evolution and Genetics of Latin American Populations*. Cambridge University Press: Cambridge, UK, 2002.
- Zembrzusi VM, Callegari-Jacques SM, Hutz MH. Application of an African Ancestry Index as a genomic control approach in a Brazilian population. *Ann Hum Genet* 2006; **70**: 822–828.
- Santos NP, Ribeiro-Rodrigues EM, Ribeiro-Dos-Santos AK, Pereira R, Gusmao L, Amorim A et al. Assessing individual interethnic admixture and population substructure using a 48-insertion-deletion (INSEL) ancestry-informative marker (AIM) panel. *Hum Mutat* 2010; **31**: 184–190.
- Goldstein JI, Crenshaw A, Carey J, Grant GB, Maguire J, Fromer M et al. zCall: a rare variant caller for array-based genotyping: genetics and population analysis. *Bioinformatics* 2012; **28**: 2543–2545.
- Grove ML, Yu B, Cochran BJ, Haritunians T, Bis JC, Taylor KD et al. Best practices and joint calling of the HumanExome BeadChip: the CHARGE Consortium. *PLoS ONE* 2013; **8**: e68095.
- 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.
- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006; **38**: 904–909.
- Price AL, Weale ME, Patterson N, Myers SR, Need AC, Shianna KV et al. Long-range LD can confound genome scans in admixed populations. *Am J Hum Genet* 2008; **83**: 132–135, author reply 135–139.
- Feng S, Liu D, Zhan X, Wing MK, Abecasis GR. RAREMETAL: fast and powerful meta-analysis for rare variants. *Bioinformatics* 2014; **30**: 2828–2829.
- Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genome-wide association scans. *Bioinformatics* 2010; **26**: 2190–2191.
- Devlin B, Roeder K. Genomic control for association studies. *Biometrics* 1999; **55**: 997–1004.
- Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* 2012; **22**: 1790–1797.
- de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* 2015; **11**: e1004219.
- Morris AP, Zeggini E. An evaluation of statistical approaches to rare variant analysis in genetic association studies. *Genet Epidemiol* 2010; **34**: 188–193.
- Neale BM, Medland SE, Ripke S, Asherson P, Franke B, Lesch KP et al. Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry* 2010; **49**: 884–897.
- Pletikos M, Sousa AM, Sedmak G, Meyer KA, Zhu Y, Cheng F et al. Temporal specification and bilaterality of human neocortical topographic gene expression. *Neuron* 2014; **81**: 321–332.
- Seiffers A, Schultz J, Gohla A. Human HAD phosphatases: structure, mechanism, and roles in health and disease. *FEBS J* 2013; **280**: 549–571.
- Tani K, Kogure T, Inoue H. The intracellular phospholipase A1 protein family. *Biomol Concepts* 2012; **3**: 471–478.
- Fukaya M, Fukushima D, Hara Y, Sakagami H. EFA6A, a guanine nucleotide exchange factor for Arf6, interacts with sorting nexin-1 and regulates neurite outgrowth. *J Neurochem* 2014; **129**: 21–36.
- Choi S, Ko J, Lee JR, Lee HW, Kim K, Chung HS et al. ARF6 and EFA6A regulate the development and maintenance of dendritic spines. *J Neurosci* 2006; **26**: 4811–4819.
- Barbacid M. ras genes. *Annu Rev Biochem* 1987; **56**: 779–827.
- Gazin C, Wajapeyee N, Gobeil S, Virbasius CM, Green MR. An elaborate pathway required for Ras-mediated epigenetic silencing. *Nature* 2007; **449**: 1073–1077.
- Ikonomidou C. Triggers of apoptosis in the immature brain. *Brain Dev* 2009; **31**: 488–492.
- Shaffer JR, Feingold E, Wang X, Lee M, Tcuenko K, Weeks DE et al. GWAS of dental caries patterns in the permanent dentition. *J Dent Res* 2013; **92**: 38–44.

- 36 Vieira AR, McHenry TG, Daack-Hirsch S, Murray JC, Marazita ML. Candidate gene/loci studies in cleft lip/palate and dental anomalies finds novel susceptibility genes for clefts. *Genet Med* 2008; **10**: 668–674.
- 37 Wang Q, Jia P, Cuenco KT, Zeng Z, Feingold E, Marazita ML *et al*. Association signals unveiled by a comprehensive gene set enrichment analysis of dental caries genome-wide association studies. *PLoS ONE* 2013; **8**: e72653.
- 38 Bigdeli TB, Maher BS, Zhao Z, Sun J, Medeiros H, Akula N *et al*. Association study of 83 candidate genes for bipolar disorder in chromosome 6q selected using an evidence-based prioritization algorithm. *Am J Med Genet B Neuropsychiatr Genet* 2013; **162B**: 898–906.
- 39 Fan J, Ionita-Laza I, McQueen MB, Devlin B, Purcell S, Faraone SV *et al*. Linkage disequilibrium mapping of the chromosome 6q21-22.31 bipolar I disorder susceptibility locus. *Am J Med Genet B Neuropsychiatr Genet* 2010; **153B**: 29–37.
- 40 Halmoy A, Klungsoyr K, Skjaerven R, Haavik J. Pre- and perinatal risk factors in adults with attention-deficit/hyperactivity disorder. *Biol Psychiatry* 2012; **71**: 474–481.
- 41 Landaas ET, Johansson S, Halmoy A, Oedegaard KJ, Fasmer OB, Haavik J. Bipolar disorder risk alleles in adult ADHD patients. *Genes Brain Behav* 2011; **10**: 418–423.
- 42 Rosenberg SS, Kumar S, Williams NJ. Attention deficit/hyperactivity disorder medication and dental caries in children. *J Dent Hyg* 2014; **88**: 342–347.
- 43 Nopoulos P, Boes AD, Jabines A, Conrad AL, Canady J, Richman L *et al*. Hyperactivity, impulsivity, and inattention in boys with cleft lip and palate: relationship to ventromedial prefrontal cortex morphology. *J Neurodev Disord* 2010; **2**: 235–242.
- 44 Persico AM, Bourgeron T. Searching for ways out of the autism maze: genetic, epigenetic and environmental clues. *Trends Neurosci* 2006; **29**: 349–358.
- 45 Bagni C, Greenough WT. From mRNP trafficking to spine dysmorphogenesis: the roots of fragile X syndrome. *Nat Rev Neurosci* 2005; **6**: 376–387.
- 46 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.
- 47 Farooqui AA, Horrocks LA, Farooqui T. Glycerophospholipids in brain: their metabolism, incorporation into membranes, functions, and involvement in neurological disorders. *Chem Phys Lipids* 2000; **106**: 1–29.
- 48 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, e726.
- 49 Must A, Koks S, Vasar E, Tasa G, Lang A, Maron E *et al*. Common variations in 4p locus are related to male completed suicide. *Neuromolecular Med* 2009; **11**: 13–19.
- 50 Hudson C, Schwanke C, Johnson JP, Elias AF, Phillips S, Schwalbe T *et al*. Confirmation of 6q21-6q22.1 deletion in acro-cardio-facial syndrome and further delineation of this contiguous gene deletion syndrome. *Am J Med Genet A* 2014; **164A**: 2109–2113.
- 51 Palacios-Reyes C, Espinosa A, Contreras A, Ordonez R, Hidalgo-Miranda A, Rubio-Gayosso I *et al*. Williams' neural stem cells: new model for insight into microRNA dysregulation. *Front Biosci (Elite Ed)* 2013; **5**: 1057–1073.
- 52 Schizophrenia Working Group of the Psychiatric Genomics C. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014; **511**: 421–427.
- 53 Psychiatric GCBWDG. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat Genet* 2011; **43**: 977–983.
- 54 Lesch KP, Timmesfeld N, Renner TJ, Halperin R, Roser C, Nguyen TT *et al*. Molecular genetics of adult ADHD: converging evidence from genome-wide association and extended pedigree linkage studies. *J Neural Transm* 2008; **115**: 1573–1585.
- 55 Lee S, Abecasis GR, Boehnke M, Lin X. Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet* 2014; **95**: 5–23.
- 56 Hyman S. Mental illness: genetically complex disorders of neural circuitry and neural communication. *Neuron* 2000; **28**: 321–323.



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/>

© The Author(s) 2016

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