

# SCIENTIFIC REPORTS



OPEN

## Gene-wide Association Study Reveals *RNF122* Ubiquitin Ligase as a Novel Susceptibility Gene for Attention Deficit Hyperactivity Disorder

Iris Garcia-Martínez<sup>1,2</sup>, Cristina Sánchez-Mora<sup>1,2,3</sup>, María Soler Artigas<sup>1,2,3</sup>, Paula Rovira<sup>1,2</sup>, Mireia Pagerols<sup>1,2</sup>, Montse Corrales<sup>2,4</sup>, Eva Calvo-Sánchez<sup>1,2</sup>, Vanesa Richarte<sup>2,3,4</sup>, Mariona Bustamante<sup>5,6,7,8</sup>, Jordi Sunyer<sup>5,7,8,9</sup>, Bru Cormand <sup>10,11,12,13</sup>, Miquel Casas<sup>1,2,3,4</sup>, Josep Antoni Ramos-Quiroga<sup>1,2,3,4</sup> & Marta Ribasés <sup>1,2,3</sup>

Attention Deficit Hyperactivity Disorder (ADHD) is a common childhood-onset neurodevelopmental condition characterized by pervasive impairment of attention, hyperactivity, and/or impulsivity that can persist into adulthood. The aetiology of ADHD is complex and multifactorial and, despite the wealth of evidence for its high heritability, genetic studies have provided modest evidence for the involvement of specific genes and have failed to identify consistent and replicable results. Due to the lack of robust findings, we performed gene-wide and pathway enrichment analyses using pre-existing GWAS data from 607 persistent ADHD subjects and 584 controls, produced by our group. Subsequently, expression profiles of genes surpassing a follow-up threshold of P-value < 1e-03 in the gene-wide analyses were tested in peripheral blood mononucleated cells (PBMCs) of 45 medication-naïve adults with ADHD and 39 healthy unrelated controls. We found preliminary evidence for genetic association between *RNF122* and ADHD and for its overexpression in adults with ADHD. *RNF122* encodes for an E3 ubiquitin ligase involved in the proteasome-mediated processing, trafficking, and degradation of proteins that acts as an essential mediator of the substrate specificity of ubiquitin ligation. Thus, our findings support previous data that place the ubiquitin-proteasome system as a promising candidate for its involvement in the aetiology of ADHD.

Attention Deficit Hyperactivity Disorder (ADHD) is a common childhood-onset neurodevelopmental disorder with a high estimated prevalence of 5.3% among children and of 2.5% in adulthood<sup>1</sup>. Family and twin studies have

<sup>1</sup>Psychiatric Genetics Unit, Group of Psychiatry, Mental Health and Addiction, Vall d'Hebron Research Institute (VHIR), Universitat Autònoma de Barcelona, Barcelona, Catalonia, Spain. <sup>2</sup>Department of Psychiatry, Hospital Universitari Vall d'Hebron, Barcelona, Catalonia, Spain. <sup>3</sup>Biomedical Network Research Centre on Mental Health (CIBERSAM), Instituto de Salud Carlos III, Madrid, Spain. <sup>4</sup>Department of Psychiatry and Legal Medicine, Universitat Autònoma de Barcelona, Barcelona, Catalonia, Spain. <sup>5</sup>ISGlobal, Centre for Research in Environmental Epidemiology (CREAL), Barcelona, Spain. <sup>6</sup>Genomics and Disease Group, Bioinformatics and Genomics Program, Centre for Genomic Regulation (CRG), Barcelona, Spain. <sup>7</sup>Universitat Pompeu Fabra (UPF), Barcelona, Spain. <sup>8</sup>CIBER Epidemiología y Salud Pública (CIBERESP), Madrid, Spain. <sup>9</sup>IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain. <sup>10</sup>Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Universitat de Barcelona, Barcelona, Catalonia, Spain. <sup>11</sup>Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Instituto de Salud Carlos III, Madrid, Spain. <sup>12</sup>Institut de Biomedicina de la Universitat de Barcelona (IBUB), Barcelona, Catalonia, Spain. <sup>13</sup>Institut de Recerca Sant Joan de Déu (IR-SJD), Esplugues, Catalonia, Spain. Iris Garcia-Martínez and Cristina Sánchez-Mora contributed equally to this work. Correspondence and requests for materials should be addressed to M.R. (email: [marta.ribases@vhir.org](mailto:marta.ribases@vhir.org))

shown that genetic factors play a crucial role in ADHD susceptibility and have estimated the heritability of the disorder to be around 76–80% both in children and in adults<sup>1</sup>.

In spite of this high heritability, genome-wide linkage studies or hypothesis-driven candidate gene association analyses in ADHD have failed to identify consistent and replicable genetic factors, and provide modest evidence for the involvement of some specific genes on the basis of meta-analyses<sup>2–5</sup>. Aiming to overcome these issues, and along with advances in high-throughput technologies, a number of genome-wide association studies (GWAS) has been performed in ADHD in the last few years. It has been reported that around 28% of the total variance in the liability to ADHD may be explained by common nucleotide polymorphisms (SNPs), and that a considerable part of this estimated SNP-based heritability might be shared with different psychiatric disorders<sup>5,7</sup>. GWAS in ADHD have shown suggestive evidence for association of cadherin 13 (*CDH13*) and other candidate genes, including monoamine system-related genes (e.g. *SLC9A9* and *SNAP25*), the glutamate metabotropic receptor 5 (*GRM5*), the glucose-fructose oxidoreductase domain containing 1 (*GFOD1*), the cannabinoid receptor 1 (*CNR1*), the protein kinase CGMP-dependent 1 (*PRKG1*) or the nicotinic acetylcholine receptor alpha 7 (*CHRNA7*), among others<sup>2–4,8</sup>. However, no genome-wide significant findings have been identified [ $P \leq 5e-8$ ], there is limited overlap between results from different GWAS, and none of the classical candidate genes for ADHD has been found among the top findings. These results may reflect differences in study design or phenotypes analyzed in the different GWAS performed so far, and suggest that larger samples, more homogeneous populations and more complex statistical strategies are required to identify genetic variants displaying low to moderate effects in ADHD<sup>4,8</sup>.

Considering the absence of genome-wide significant associations and the limited overlap between top hits from SNP-based GWAS on ADHD, analyses focusing on gene-set enrichment or pathway-based approaches raise as promising strategies to address the genetic complexity of the disorder<sup>9</sup>. These strategies allow combining effects of multiple SNPs (gene-wide studies) or multiple genes (gene-set analyses), while reducing multiple testing comparisons and providing insights into the involvement of specific biological functions or pathways<sup>9</sup>. In this line, Mick *et al.* performed gene-wide analyses and provided additional evidence for the association between ADHD and the solute carrier *SLC9A9*, a candidate gene for the disorder that had been found nominally associated with ADHD in other SNP-based GWAS<sup>2–4,8</sup>. Gene-set enrichment approaches or pathway analyses on ADHD also highlighted genes involved in the regulation of gene expression, cell adhesion and inflammation, ubiquitin-mediated proteasomal degradation, neurodegenerative disorders, axon guidance, neuron projections or synaptic components, and revealed significant overlap between pathways enriched for SNP association and those enriched for rare copy number variants (CNV)<sup>2–4,8</sup>. To integrate findings from GWAS and provide knowledge about potential molecular processes underlying ADHD, Poelmans *et al.* performed network analysis considering top findings from five GWAS and revealed that 45 of the 85 top-ranked ADHD candidate genes encoded proteins that fitted into a neurodevelopmental network involved in directed neurite outgrowth<sup>10</sup>. Recently, the Psychiatric Genomics Consortium (PGC) combined GWAS signals from over 60,000 subjects to identify common network and biological pathways across three adult psychiatric conditions: schizophrenia, major depression and bipolar disorder. This study highlighted an important overlap across disorders, with histone methylation processes showing the strongest association, as well as multiple immune and neuronal signaling pathways<sup>11</sup>.

Due to the lack of robust findings in ADHD through SNP-based GWAS, which may be explained by the lack of power of the studies conducted so far and the polygenic, multifactorial nature of the disorder -with common and rare variants likely contributing small to moderate effects to its etiology-, we performed gene-wide and pathway enrichment analyses using GWAS data from 607 persistent ADHD subjects and 584 controls. Subsequently, expression profiles of genes surpassing a follow-up significance threshold of  $P$ -value  $< 1e-03$  in the gene-wide analyses were tested in peripheral blood mononucleated cells (PBMCs) of 45 medication-naïve adults with ADHD and 39 healthy unrelated controls.

## Materials and Methods

**Subjects and Clinical Assessment.** The clinical sample consisted of 607 adult ADHD subjects and 584 unrelated healthy individuals, all Spanish and Caucasian. Detailed information regarding the sample is described elsewhere<sup>12</sup>.

The evaluation of the ADHD diagnosis was carried out with the Structured Clinical Interview for DSM-IV Axis I and II Disorders (SCID-I and SCID-II) and with the Conners' Adult ADHD Diagnostic Interview for DSM-IV (CAADID parts I and II)<sup>13</sup>. Severity of ADHD symptoms in adulthood was assessed with the long version of the Conners' ADHD Rating Scale (self-report [CAARS-S:L] and observer [CAARS-O:L])<sup>14</sup>, the ADHD Rating Scale (ADHD-RS)<sup>15</sup> and the Wender Utah Rating Scale (WURS) for retrospective symptomatology in childhood<sup>16</sup>. The level of impairment was measured with the Clinical Global Impression (CGI) and the Sheehan Disability Inventory<sup>17,18</sup>. Additional tests used for clinical assessment are available in Ribasés *et al.*<sup>19</sup>. Exclusion criteria were IQ  $< 70$ ; lifelong and current history of mood, psychotic, anxiety, substance abuse, and DSM-IV axis II disorders; pervasive developmental disorders; a history or the current presence of a condition or illness, including neurologic, metabolic, cardiac, liver, kidney, or respiratory disease; a chronic medication of any kind; birth weight  $\leq 1.5$  kg; and other neurological or systemic disorders that might explain ADHD symptoms.

The control sample consisted of unrelated healthy individuals matched for sex with the clinical group. ADHD symptomatology was excluded retrospectively under the following criteria: 1) not having been diagnosed with ADHD previously and 2) answering negatively to the life-time presence of the following ADHD symptoms: a) often has trouble in keeping attention on tasks; b) usually loses things needed for tasks; c) often fidgets with hands or feet or squirms in seat and d) often gets up from seat when remaining in seat is expected.

All subjects were evaluated and recruited at Hospital Universitari Vall d'Hebron of Barcelona (Spain) and diagnosis was blind to genotype. The study was approved by the Clinical Research Ethics Committee (CREC) of Hospital Universitari Vall d'Hebron, all methods were performed in accordance with the relevant guidelines and regulations and written informed consent was obtained from all subjects before the inclusion into the study.

**Gene-wide Analysis.** Genome-wide genotyping of 607 adults with ADHD and 584 healthy controls was performed with the Illumina HumanOmni1-Quad BeadChip platform [Illumina Inc., San Diego, California, USA]. Quality control assessment was implemented at the individual and SNP level using PLINK v1.07<sup>20</sup> and included filtering subjects with low call rate (<98%) or gender discrepancy, followed by filtering SNPs with minor allele frequency (MAF) < 0.01, Hardy-Weinberg equilibrium test P-values ( $P_{HWE}$ ) < 1e-06 or call rate < 0.99 in either cases or controls. After stringent quality control assessment, five samples were excluded from the analysis due to low call rate and a total of 794,090 SNPs with a mean call rate of 0.9994 for the remaining 603 cases and 583 controls were included in the study.

Genome-wide association analysis was performed using the Cochran-Armitage trend test with PLINK v1.07<sup>20</sup> and the genomic inflation factor ( $\lambda$ ) of 1.031 was used to correct for the degree of inflation<sup>12</sup>. Gene-wide association analyses were conducted with the VEGAS software [Versatile Gene-Based Association Study; <http://gump.qimr.edu.au/VEGAS/>], which combines the effect of all SNPs in a gene locus and corrects for linkage disequilibrium (LD)<sup>21</sup>. Regions were defined using VEGAS default option as  $\pm 50$  kb upstream and downstream from each gene locus (UCSC Genome Browser; NCBI36/hg18 assembly), and the HapMap CEU population was used to estimate patterns of LD for each gene. Two different strategies were followed for gene-wide analyses by (i) considering all SNPs that mapped within a gene locus and (ii) restricting the analysis to the 10% most significant SNPs from each gene locus. Multiple-testing adjustment was addressed by the Bonferroni correction, and the significance threshold was set at P-value < 1.4e-06, taking into account 17,787 autosomal genes and the two gene-wide strategies (0.05/(17,787\*2)). Positive signals surpassing a follow-up threshold of P-value < 1e-03 were tested in an independent dataset of 2,064 ADHD trios, 896 ADHD cases and 2,455 controls, following the same strategy described for the discovery sample<sup>7,22</sup>.

**Gene-Set and Pathway Analysis.** Gene-set and pathway analyses were restricted to genes surpassing a follow-up significance threshold set at P-value < 1e-03 in the gene-wide analyses when either all SNPs or the 10% most significant SNPs in gene locus were considered. The definition of the follow-up threshold was based on previous evidence supporting that VEGAS software provides an accurate performance for genes with P-value < 1e-03, offering a sizeable sensitivity, with less than 1% false positives and specificities ranging from 98% to 100%, while being able to distinguish between multiple independent causal loci and multiple signals due to linkage disequilibrium<sup>23</sup>.

The gene-set analyses were performed with MAGMA software [Generalized Gene-Set Analysis of GWAS Data; <http://ctg.cncr.nl/software/magma>]<sup>24</sup>, a statistical method for analyzing multiple genetic variants simultaneously to determine their joint effect while correcting for LD (gene-wide), and subsequently, assembling individual genes into groups of genes sharing biological or functional characteristics (gene-set). Multiple-testing correction was assessed by 10,000 permutations as implemented in the MAGMA software and the significance threshold was set at adjusted P-value < 0.05.

Functional and pathway enrichment analyses were conducted with IPA software [Ingenuity Pathway Analysis; Ingenuity Systems<sup>®</sup>, Redwood City, California, USA; <http://www.ingenuity.com>]. Fisher's exact test was used to calculate P-values, based on the number of genes/molecules that map to a biological function, disease, pathway, or network. A Benjamini-Hochberg's (BH) threshold of 0.05 was applied for multiple comparison correction in the functional, disease and pathway enrichment analyses. Gene networks were considered of relevance when the network score (P-score =  $-\log_{10}(P\text{-value})$ ) was over 8 (P-value < 1e-08)<sup>25,26</sup>.

**Gene expression analysis in peripheral blood mononuclear cells (PBMCs) with microarrays.** Expression levels of genes identified in the gene-wide study (follow-up P-value threshold < 1e-03) were tested in PBMCs of 45 medication-naïve adults with a new clinical diagnosis of ADHD (69% male; mean age = 38 years (SD = 9.9)) and 39 healthy unrelated controls (60% males, mean age = 35 years (SD = 12.1)).

Briefly, PBMCs were isolated using the Ficoll density gradient method, and total RNA was extracted using Qiazol Lysis reagent and the RNeasy Midi kit [Qiagen, Hilden, Germany]. The quality of the samples was assayed by 2100 Bioanalyzer [Agilent Technologies Inc. Santa Clara, California, USA]. RNA was reverse transcribed using the Ambion WT Expression Kit [Life technologies, Massachusetts, USA]. The cRNA was subsequently fragmented, labelled and hybridized with the GeneChip WT Terminal Labeling and Hybridization Kit to the Genechip Human Gene 1.1 ST 96-Array plate [Affymetrix, Santa Clara, California, USA]. The experiment was performed in 9 different amplification rounds, introducing a batch factor that was taken into consideration in the following statistical analyses. The array processing and data generation were assessed using the Gene Titan Affymetrix microarray platform. Background correction, normalization and summarization of probes values was performed using the Robust Multichip Average (RMA) algorithm implemented in the *oligo* R library<sup>27</sup>. Expression patterns for genes of interest in ADHD subjects and controls were contrasted using the *limma* R library<sup>28</sup>, including batch and gender as covariates. Bonferroni correction was applied and the significance threshold was set at P-value = 1.5e-03, taking into account the 33 genes displaying P-values below the follow-up threshold in the gene-wide analyses and with microarray data available (0.05/33).

**Validation of gene expression differences with reverse transcription real-time quantitative polymerase chain reaction (RT-qPCR).** Validation of gene expression differences was performed on genes showing tentative evidence for differential expression in the microarray analysis, which included *MAB21L2*, *STXBP3* and *RNF122*. Gene expression validation assays were conducted with reverse transcription real-time quantitative polymerase chain reaction (RT-qPCR) in the same clinical sample of 45 ADHD cases and 39 controls. First, 2  $\mu$ g of total RNA was reverse transcribed using the High Capacity cDNA Reverse Transcription kit according to manufacturer's protocol in a final reaction volume of 20  $\mu$ L [Applied Biosystems; Foster City, California, USA]. Following, RT-qPCRs were run in triplicate using 2  $\mu$ L of cDNA, 2  $\mu$ L of RNase free water,

5  $\mu$ L of TaqMan<sup>®</sup> Gene Expression Master Mix and 1  $\mu$ L of TaqMan<sup>®</sup> Gene Expression Assay, in a final reaction volume of 10  $\mu$ L [Applied Biosystems; Foster City, California, USA]. Reactions were measured in an Applied Biosystems 7900HT Fast Real-Time PCR system using the default thermal cycling conditions specified by the manufacturer [Applied Biosystems; Foster City, California, USA]. The threshold cycle ( $C_T$ ) was defined as the fractional cycle number at which the fluorescence exceeded the threshold of 0.2. The relative quantification of mRNA expression was calculated by the  $2^{-\Delta\Delta C_T}$  method<sup>29</sup>, considering only those samples showing standard deviations values  $\leq 0.3$  among triplicates and the *GADD45A* gene as an endogenous control, after checking its stability and linearity across all samples. Using the *Stats* R package [<https://www.R-project.org/>]<sup>30</sup>, generalized linear models (GLM) were applied to compare gene expression levels between ADHD cases and controls, including gender as covariate in the fitted model. The statistical test was one-sided and the Bonferroni correction was applied for multiple-testing control, setting the statistical significance threshold at P-value  $< 0.017$  when taking into account three genes (0.05/3).

### Imputation, *cis*-expression Quantitative Trait Loci (*cis*-eQTL) Analyses and Prediction of Functional Effects.

To better delineate the involvement of *RNF122* in ADHD and to detect potential functional variants, markers at this locus were imputed in the original dataset of 603 subjects with ADHD and 583 healthy controls. Pre-imputation quality control of the GWAS dataset at the individual and SNP level was implemented in accordance to the QC module instructions from the Ricopili pipeline considering default settings [<https://sites.google.com/a/broadinstitute.org/ricopili/>]. Screening for cryptic relatedness and population stratification was performed by Principal Components Analysis (PCA). Markers at the *RNF122* gene region plus 10 kb upstream and 5 kb downstream from the locus (chr8:33519815–33554185; NCBI36/hg18) were imputed in the GWAS sample through the pre-phasing and imputation strategies implemented by SHAPEIT and IMPUTE2, respectively<sup>31, 32</sup>, using the Ricopili pipeline [<https://sites.google.com/a/broadinstitute.org/ricopili/>] and data from the 1000 Genomes Project as the reference panel [<http://www.1000genomes.org/>]<sup>33</sup>. After filtering SNPs with MAF  $< 0.01$  and low imputation quality ( $r < 0.4$ ), 138 SNPs were finally considered. We performed the association analysis using logistic regression models with the PLINK v1.07 software<sup>20</sup> and multiple-testing was addressed by the Bonferroni correction, setting the significance threshold at P-value  $< 3.6e-04$  when considering 138 imputed SNPs in the *RNF122* locus (0.05/138). Since this approach may be too conservative, alternative multiple-testing control was assessed using the Single Nucleotide Polymorphism Spectral Decomposition (SNPSpD) software [<http://neurogenetics.qimrberghofer.edu.au/SNPSpDlite/>]<sup>34</sup>, which takes into account patterns of LD (P-value  $< 1.27e-03$ ). Once top signals were identified, in order to uncover additional independent effects and to assess evidence for multi-risk loci in each region, further conditioned analysis was performed with PLINK v1.07 software<sup>20</sup>. To condition the logistic regression analysis on a specific SNP, we tested all markers again but adding the allelic dosage for the conditioned SNP as a covariate<sup>20</sup>.

The *cis*-eQTL analyses were conducted using genotype and expression data from a series of neuropathologically and neuropsychiatrically normal human brain samples from GSE8919<sup>35</sup> and GSE30272<sup>36</sup> datasets, available at the Gene Expression Omnibus site [<http://www.ncbi.nlm.nih.gov/geo>]. Genotype data from both studies were imputed following the methodology described above. Expression levels of the *RNF122* transcript NM\_024787.2 were available for 62 cortical samples of European ancestry in the GSE8919 dataset (probe ID: GI\_38045930-S) and for 94 prefrontal cortex samples of Caucasian origin in the GSE30272 dataset (probe ID:HEEBO-062-HCC62D14)<sup>35, 36</sup>. Rank-based inverse normal transformation of expression data was applied using *Stats* R package [<https://www.R-project.org/>]<sup>30</sup> and additive linear regression models were fitted for eQTL mapping using PLINK v1.07 software<sup>20</sup>, considering covariates showing suggestive association with the outcome (P-value  $< 0.2$ ; gender, age\_at\_death and transcripts\_detected\_rate24354 (average transcript detection rate for sample, out of all 24354 probes) for GSE8919 and sv2 (surrogate variable 2) for GSE30272 datasets). Bonferroni correction was applied for multiple testing and the significance threshold was set at P-value  $< 4.3e-04$  for the GSE8919 dataset and at P-value  $< 3.8e-04$  for the GSE30272 dataset, considering the number of available SNPs.

Potential functional effects of the independent risk alleles associated with ADHD were predicted using the SNPinfo, ESEfinder, and RESCUE\_ESE web softwares<sup>37–40</sup>.

## Results

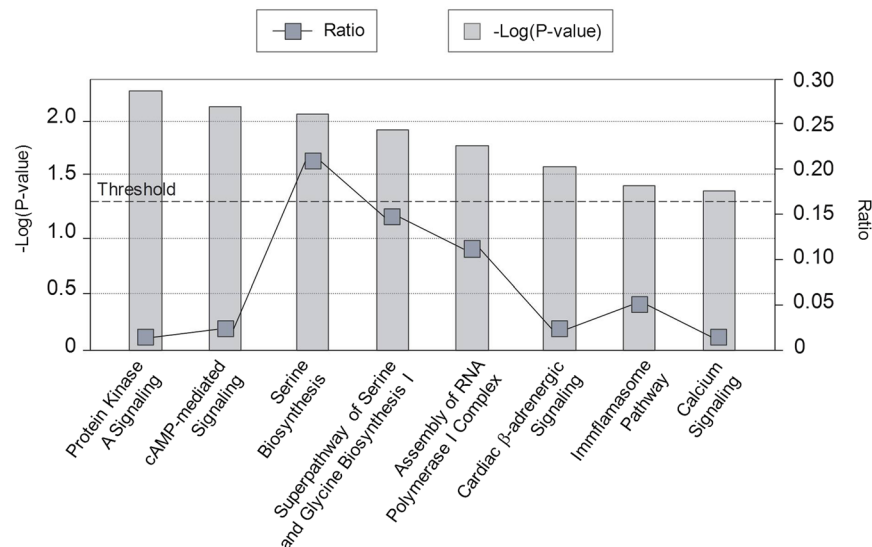
**Gene-Wide and Gene-Set Analysis.** After individual and SNP-based standard quality control filtering, 794,090 autosomal SNPs in 603 adult ADHD cases and 583 healthy controls were included in the GWAS. The quantile–quantile plot showed no departure from the expected P-values distribution, with a genomic control inflation factor of  $\lambda = 1.031$ <sup>12</sup>.

The gene-wide association analysis revealed 20 genes surpassing the follow-up significance threshold (P-value  $< 1e-03$ ) when considering all SNPs located within a gene locus, with *CTAGE5* (P-value  $< 1.0e-06$ ) and *FBXO33* (P-value =  $3.0e-06$ ) as the most associated genes (Table 1). Twenty additional genes were found associated with persistent ADHD when we considered the 10% most significant SNPs of each gene locus, with *KCNG4* (P-value =  $1.0e-04$ ) and *TAF1C* (P-value =  $1.9e-04$ ) as top signals (Table 1). After Bonferroni correction, only *CTAGE5* remained associated with ADHD. Consistently, the gene-set analysis revealed that both gene-sets were significantly associated with persistent ADHD after correcting for multiple comparisons (gene-set corrected P-value =  $1e-03$  when considering all SNPs within a gene locus and gene-set corrected P-value  $< 1e-03$  when considering the top 10% SNPs in each locus) (Table 1), although no overlap was detected between them. However, some genes highlighted in the previous single-marker GWAS were found among top genes surpassing the follow-up significance threshold in the present gene-wide analyses (*FBXO33*, *PEX19*, *COPA* and *KCNG4*)<sup>12</sup>.

We further tested the 40 genes surpassing the follow-up threshold of P-value  $< 1e-03$  in an independent dataset from the first large-scale meta-analysis of ADHD GWAS, consisting of 2,064 ADHD trios, 896 ADHD cases and 2,455 controls from the Psychiatric Genetics Consortium<sup>7, 22</sup>. No signal was associated with ADHD after

Chr	Gene	Number of SNPs	Start Position	End Position	Gene-wide P-value	Loci associated with neuropsychiatric or neurological disorders
<b>Full set of SNPs</b>						
<b>Global gene-set P-value = 2.2e-4 (<math>P_{corrected} = 1.0e-03</math>)</b>						
14	<i>CTAGE5</i>	39	38754226	38940148	0*	Antidepressant efficacy in major depressive disorder, mild intellectual disability, and traits of inattention and hyperactivity <sup>61, 66</sup>
14	<i>FBXO33</i>	17	38886709	39021371	3.00e-06	Autism spectrum disorders, attention deficit, hyperactivity and mild mental retardation <sup>12, 65, 66</sup>
14	<i>MIA2</i>	26	38722875	38842326	4.00e-06	Mild intellectual disability, and traits of inattention and hyperactivity <sup>61</sup>
1	<i>PEX19</i>	10	158463225	158571555	3.40e-05	Major depressive disorder and ADHD <sup>12, 67</sup>
1	<i>COPA</i>	32	158475000	158629978	4.40e-05	Bipolar Disorder and ADHD <sup>12, 68</sup>
14	<i>PNN</i>	24	38664137	38772173	6.70e-05	Mild intellectual disability, and traits of inattention and hyperactivity <sup>61</sup>
1	<i>WDR42A</i>	34	158402128	158548603	8.80e-05	—
1	<i>NCSTN</i>	32	158529686	158645366	1.76e-04	Alzheimer's Disease and Schizophrenia <sup>69, 70</sup>
1	<i>MAST2</i>	34	45991871	46324383	3.35e-04	—
1	<i>CASQ1</i>	46	158376988	158488300	3.56e-04	—
15	<i>CALML4</i>	14	66220096	66335502	4.50e-04	Major depressive disorder and stress <sup>71, 72</sup>
1	<i>ZP4</i>	45	236062332	236170558	5.44e-04	ADHD <sup>73, 74</sup>
1	<i>IPP</i>	22	45886993	46034720	6.75e-04	Cocaine dependence <sup>75</sup>
1	<i>TMEM69</i>	19	45876433	45982695	7.12e-04	Amphetamine effects and aggressiveness <sup>76, 77</sup>
4	<i>MAB21L2</i>	15	151672752	151775295	7.60e-04	Autism spectrum disorders, and mild intellectual disability <sup>78-80</sup>
1	<i>GPBP1L1</i>	29	45815567	45949398	7.75e-04	Autism spectrum disorders, and traits of inattention in Tourette syndrome <sup>81, 82</sup>
1	<i>STXBP3</i>	69	109040807	109203671	8.53e-04	—
13	<i>MLNR</i>	16	48642474	48744514	9.24e-04	—
2	<i>NLRC4</i>	26	32253021	32394305	9.36e-04	Schizophrenia, and bipolar disorder <sup>83, 84</sup>
1	<i>PEA15</i>	37	158391750	158501786	9.42e-04	Major depressive disorder <sup>85</sup>
<b>Top 10% of SNPs</b>						
<b>Global gene-set P-value = 1.12e-10 (<math>P_{corrected} &lt; 1.0e-03</math>)</b>						
16	<i>KCNQ4</i>	118	82763323	82880857	1.00e-04	Alcohol consumption, and bipolar disorder in Amish population <sup>86, 87</sup>
16	<i>TAF1C</i>	86	82718961	82828163	1.90e-04	Autism and schizophrenia <sup>88-90</sup>
22	<i>NCF4</i>	48	35536975	35654005	2.10e-04	Schizophrenia <sup>91</sup>
22	<i>CSF2RB</i>	61	35589620	35716425	3.60e-04	Schizophrenia, major depressive disorder and bipolar disorder <sup>92-94</sup>
17	<i>AKAP10</i>	18	19699341	19871721	3.80e-04	Autism spectrum disorders, schizophrenia and bipolar disorder <sup>95-97</sup>
1	<i>ATP1A1</i>	23	116667358	116798919	4.10e-04	Major depressive disorder, autism spectrum disorders, anxiety and bipolar disorder <sup>98-101</sup>
8	<i>DUSP26</i>	21	33518392	33626981	5.10e-04	Antidepressant Efficacy in Major Depressive Disorder <sup>66</sup>
12	<i>RAP1B</i>	18	67240918	67390641	5.80e-04	Autism spectrum disorders <sup>102, 103</sup>
19	<i>AKAP8</i>	9	15275334	15401603	6.10e-04	Autism spectrum disorders <sup>95, 104</sup>
20	<i>RIN2</i>	113	19768209	19981100	6.40e-04	Bipolar disorder, schizophrenia and stress-induced changes <sup>105-107</sup>
2	<i>SPAST</i>	29	32092183	32286210	6.80e-04	Autism spectrum disorders, and cognitive deficiency <sup>109, 110</sup>
7	<i>CCL26</i>	34	75186777	75307000	7.10e-04	Schizophrenia, and bipolar disorder <sup>111, 112</sup>
20	<i>NAT5</i>	59	19895936	20012269	7.60e-04	Autism spectrum disorders and comorbid anxiety <sup>113</sup>
8	<i>RNF122</i>	31	33474814	33594185	7.70e-04	Antidepressant efficacy in major depressive disorder <sup>66</sup>
15	<i>C15orf53</i>	66	36726090	36829531	7.90e-04	Bipolar disorder, and alcohol dependence <sup>108, 114</sup>
7	<i>GPR146</i>	34	1013666	1115423	8.90e-04	Autism spectrum disorders <sup>115, 116</sup>
8	<i>C8orf41</i>	25	33425777	33540245	9.20e-04	Antidepressant efficacy in major depressive disorder, and cognitive deficiency <sup>66, 117, 118</sup>
2	<i>MEMO1</i>	25	31896397	32139202	9.20e-04	—
12	<i>ALX1</i>	16	84148166	84269692	9.30e-04	Autism spectrum disorders <sup>119, 120</sup>
7	<i>GPRI1</i>	36	1042968	1149977	9.40e-04	Anxiety and stress, bipolar disorder, and antipsychotic treatment for schizophrenia <sup>121-124</sup>

**Table 1.** Gene-wide and gene-set P-values considering all SNPs or the 10% most significant SNPs in gene locus using VEGAS and MAGMA softwares. Genes displaying P-value < 1e-03 are shown. Start and end positions include default flanking regions of  $\pm 50$  Kb from 5' and 3' UTRs of each gene, defined by default by VEGAS software (<http://gump.qimr.edu.au/VEGAS/>), and are based on UCSC annotation, build NCBI36/hg18 (Mar. 2006). \*An empirical P-value of 0 (from  $10^6$  simulations) can be interpreted as P-value <  $10^6$ , which exceeds a Bonferroni-corrected threshold set at P-value <  $2.8e-06$  ( $\sim 0.05/17,787$ ).



**Figure 1.** Enriched canonical pathways ( $P$ -value  $< 0.05$ ) considering 40 genes from the gene-wide analyses surpassing follow-up significance threshold ( $P$ -value  $< 1e-03$ ). The  $-\log(P$ -value) from the Fisher's exact test for each enriched category is indicated by grey bars. The ratio indicates the number of genes tested that map to the canonical pathway divided by the total number of genes that map to that pathway, and is represented by dark grey squares.

Super-category	P-value	Molecules
Embryonic Development	2.58e-04–3.83e-02	NCSTN, ALX1, CSF2RB, PNN, MAB21L2, PEA15
Organismal Development	2.58e-04–4e-02	NCSTN, GPER1, ALX1, COPA, ATP1A1, CSF2RB, AKAP8, RAP1B, MAB21L2, NLRC4, PEA15
Tissue Development	2.58e-04–4.94e-02	NCSTN, GPER1, ALX1, CSF2RB, RAP1B, IPP, MAB21L2
Cell-To-Cell Signaling and Interaction	3.96e-04–3.47e-02	NCF4, CCL26, CSF2RB, STXB3, RAP1B
Cellular Movement	1.28e-03–3.65e-02	NCF4, CCL26, GPER1, CSF2RB, RAP1B, NLRC4
Hematological System Development and Function	1.28e-03–4.89e-02	CCL26, NCF4, NCSTN, CSF2RB, RAP1B, NLRC4
Immune Cell Trafficking	1.28e-03–3.65e-02	NCF4, CCL26, CSF2RB, RAP1B, NLRC4
DNA Replication, Recombination, and Repair	1.51e-03–2.39e-02	GPER1, AKAP8, PEA15
Cardiovascular System Development and Function	1.85e-03–2.39e-02	AKAP10, GPER1, ATP1A1, CSF2RB, RAP1B
Cell Cycle	1.85e-03–3.65e-02	PEX19, PNN, NLRC4, PEA15

**Table 2.** Top-ten enriched super-categories from disease and functional enrichment analyses using IPA software, considering 40 genes identified in the gene-wide analyses surpassing follow-up significance threshold ( $P$ -value  $< 1e-03$ ).

Bonferroni correction. Nominal signals, however, were found for *FBXO33* ( $P$ -value = 0.047 when considering all SNPs in the gene locus and  $P$ -value = 0.01 when considering the top 10% SNPs) and *C15orf53* ( $P$ -value = 0.036 when considering the top 10% SNPs in the gene locus) (see Supplementary Table S1).

**Functional and Pathway Analysis.** Although none of them survived BH multiple-testing correction, enrichment for eight canonical pathways was detected when the 40 genes from the two gene-sets identified in the gene-wide association analysis were considered jointly, including protein kinase A (PKA) and cAMP-mediated signaling ( $P$ -value = 5.75e-03 and  $P$ -value = 7.94e-03, respectively), serine biosynthesis and superpathway of serine and glycine biosynthesis ( $P$ -value = 9.33e-03 and  $P$ -value = 1.29e-02, respectively), assembly of RNA polymerase I complex ( $P$ -value = 1.66e-02), cardiac adrenergic signaling ( $P$ -value = 2.57e-02), inflammasome pathway ( $P$ -value = 3.80e-02) and calcium signaling ( $P$ -value = 4.27e-02) (Fig. 1). Among the top functions and diseases, most enriched categories were mainly related to organismal and embryonic development ( $P$ -value = 2.58e-04–4.00e-02), cell-to-cell signaling and interaction ( $P$ -value = 3.96e-04–3.47e-02) and cellular movement ( $P$ -value = 1.28e-03–3.60e-02) (Table 2). In addition, several nervous system-related terms and psychological traits were also identified among enriched categories, such as relaxation of mice ( $P$ -value = 5.55e-03), swelling of neurites ( $P$ -value = 7.40e-03), axonal transport of vesicles ( $P$ -value = 9.24e-03), abnormal morphology of nervous system ( $P$ -value = 1.10e-02) or addiction behaviour ( $P$ -value = 4.36e-02) (Table 3). Three relevant networks that included genes mainly related to cellular development, organization, function and maintenance, cell death and survival, and cell-to-cell signaling and interaction were generated (Table 4). The two most relevant

Super-Categories	Diseases or Functions Categories	P-value	Molecules
Hereditary Disorder, Neurological Disease, Organismal Injury and Abnormalities	autosomal dominant spastic paraplegia type 4	1.85e-03	<i>SPAST</i>
Developmental Disorder, Hereditary Disorder, Neurological Disease, Organismal Injury and Abnormalities	autosomal recessive mental retardation type 39	1.85e-03	<i>TTI2</i>
Developmental Disorder, Neurological Disease, Organismal Injury and Abnormalities	meroanencephaly	1.85e-03	<i>ALX1</i>
Organismal Development	relaxation of mice	5.55e-03	<i>ATPIA1</i>
Connective Tissue Disorders, Developmental Disorder, Neurological Disease, Skeletal and Muscular Disorders	acrania	7.40e-03	<i>ALX1</i>
Cell Morphology, Cellular Assembly and Organization, Cellular Development, Cellular Function and Maintenance, Cellular Growth and Proliferation, Neurological Disease, Organismal Injury and Abnormalities	swelling of neurites	7.40e-03	<i>SPAST</i>
Nervous System Development and Function	antinociception of spinal cord	9.24e-03	<i>GPRI1</i>
Cellular Assembly and Organization, Cellular Function and Maintenance, Nervous System Development and Function	axonal transport of vesicles	9.24e-03	<i>SPAST</i>
Nervous System Development and Function	abnormal morphology of nervous system	1.10e-02	<i>ALX1, MAB21L2, PEA15, RAP1B, SPAST, STXBP3</i>
Connective Tissue Disorders, Developmental Disorder, Organismal Development, Skeletal and Muscular Disorders	clefing of face	1.29e-02	<i>ALX1</i>
Embryonic Development, Nervous System Development and Function, Organismal Development, Tissue Morphology	abnormal morphology of neural crest	1.47e-02	<i>NCSTN</i>
Cell Morphology, Nervous System Development and Function, Tissue Morphology	abnormal morphology of neurites	3.92e-02	<i>RAP1B, SPAST</i>
Connective Tissue Development and Function, Organ Morphology, Organismal Development, Skeletal and Muscular System Development and Function, Tissue Development	abnormal morphology of supraoccipital bone	4.00e-02	<i>ALX1</i>
Behavior	addiction behaviour	4.36e-02	<i>ATPIA1</i>
Cell Morphology, Nervous System Development and Function, Neurological Disease, Tissue Morphology	loss of axons	4.89e-02	<i>RAP1B</i>

**Table 3.** Significant nervous system-related categories identified in the disease and functional enrichment analyses using IPA software, considering 40 genes identified in the gene-wide analyses surpassing follow-up significance threshold (P-value < 1e-03).

Molecules in Network	P-score*	Focus Molecules	Top Diseases and Functions
<i>ABL1, AKAP8, AKAP10, Akt, ATPIA1, C1QBP, caspase, CASQ1, CCL26, COPA, CSF2RB, DDX3X, EMILIN1, ERK, ERK1/2, Gpcr, GPRI1, GPRI146, Insulin, MAST2, MEMO1, MLNR, NCF4, NCSTN, NFkB (complex), NLRC4, P38 MAPK, PEA15, PI3K (complex), Pka, Pkc(s), RAP1B, STXBP3, TARDBP, TRAF6</i>	44	18	Cell-To-Cell Signaling and Interaction, Cellular Development, Hematological System Development and Function
<i>AES, APP, C3orf33, CEP57, Collagen type VII, CTAGE5, DCAF8, DGC6/LOC102724770, DUSP26, ELAVL1, EMILIN1, FAM20B, FBXO33, GPBP1L1, HIGD2A, HINT3, HSP90A1, MAB21L2, MAPK3, MMP3, PEX19, PEX26, PLD6, PNN, RAF1, RBX1, RIN2, RNF122, SELT, SPAST, TARDBP, TMTC4, TRAF6, TTI2, YBX1</i>	26	12	Cellular Assembly and Organization, Cellular Compromise, Cellular Function and Maintenance
<i>ABL1, Actin, ALB, ALX1, BZW2, CALML4, Cdc2, CDC25A, CREB1, CUL1, CYBA, Cyclin B, EGLN1, EPO, GABBR1, GDF15, HNF1A, IKBKB, IPP, IRF3, KHDRBS1, KRT8, KRT18, LANCL1, MIA2, mir-296, MYB, MYBL2, MYO5B, NAA20, TAF1C, TARDBP, TGFBI, TRAF6, YBX1</i>	11	6	Cell Death and Survival, Digestive System Development and Function, Hepatic System Development and Function

**Table 4.** Significant generated networks (P-score < 8) using IPA software, considering 40 genes identified in the gene-wide analyses surpassing follow-up significance threshold (P-value < 1e-03). Focus molecules from the follow-up gene-set are shown in bold. \*P-score =  $-\log_{10}(P\text{-value})$ .

networks were highly scored (P-score = 44 and P-score = 26, respectively) and included 18 (45%) and 12 (30%) out of the 40 genes considered for this analysis.

**Gene expression analysis in peripheral blood mononuclear cells (PBMCs).** Expression levels of the 40 genes in the two gene-sets identified in the gene-wide association study were explored in microarray data from peripheral blood mononuclear cells (PBMCs) of 45 adults with ADHD and 39 healthy controls. No data were available for two of them (*WDR42A* and *NAT5*) and another five were excluded because the corresponding probes matched to multiple genes (*COPA, TMEM69, TAF1C, MEMO1, C8orf41*). Although no significant expression differences were identified after Bonferroni correction, tentative evidence for overexpression was detected in ADHD subjects for *STXBP3* (P-value = 1.9e-03; log fold change = 0.115) and *RNF122* (P-value = 0.045; log fold change = 0.111) and decreased expression for *MAB21L2* (P-value = 0.036; log fold change = -0.77) when

Probe ID	Log-Fold change	P-value	Gene Symbol	Description
8150186	0.111	0.048	<i>RNF122</i>	Localizes to the endoplasmic reticulum and golgi apparatus and may be associated with cell viability by inducing necrosis and apoptosis. Mediates protein-protein and protein-DNA interactions, and has been identified as a new E3 ubiquitin ligase that can ubiquitinate itself and undergo degradation in a RING finger- and proteasome-dependent manner <sup>41</sup> .
7903541	0.115	2.3e-03	<i>STXBP3</i>	Plays an integral role in vesicle transport through their interaction with SNAREs and could play a positive regulatory role in SNARE assembly <sup>125</sup> .
8097773	-0.77	0.034	<i>MAB21L2</i>	Required for several aspects of embryonic development including normal development of the eye. May be involved in neural development <sup>80</sup> .

**Table 5.** Differentially expressed genes in peripheral blood mononucleated cells (PBMCs) of 45 pharmacologically-naïve ADHD subjects and 39 healthy unrelated controls in the microarray analysis.

Gene name	Gene Symbol	Gene ID	Location	Assay ID	Amplicon Size		
<b>Assay information of target and control genes</b>							
<b>Housekeeping gene</b>							
growth arrest and DNA damage inducible alpha	<i>GADD45A</i>	1647	1p31.2	Hs00169255_m1	123		
<b>Target genes</b>							
mab-21-like 2	<i>MAB21L2</i>	10586	4q31	Hs00740710_s1	93		
syntaxin binding protein 3	<i>STXBP3</i>	6814	1p13.3	HS01029364_m1	149		
ring finger protein 122	<i>RNF122</i>	79845	8p12	Hs00227141	99		
Target gene	Control gene	ODDS Ratio	Standard Error	Z-value	P-value	Number of controls	Number of cases
<b>Results from RT-qPCR validation of <i>MAB21L2</i>, <i>STXBP3</i> and <i>RNF122</i> genes</b>							
<i>RNF122</i>	<i>GADD45A</i>	4.133	0.517	2.744	<b>3.04e-03</b>	34	42
<i>STXBP3</i>	<i>GADD45A</i>	2.361	0.521	1.648	0.049	40	44
<i>MAB21L2</i>	<i>GADD45A</i>	0.866	0.701	-0.206	0.419	35	25

**Table 6.** Validation of *MAB21L2*, *STXBP3* and *RNF122* gene expression differences between pharmacologically-naïve ADHD subjects and healthy unrelated controls by RT-qPCR. Statistically significant P-values after applying a Bonferroni threshold of P-value < 0.017, when taking into account three genes (0.05/3), are shown in bold.

compared to controls (Table 5). Differences in *RNF122*, *STXBP3* and *MAB21L2* expression levels were subsequently tested by RT-qPCR and evidence for significant overexpression was confirmed for *RNF122* (but not for *STXBP3* or *MAB21L2*) after controlling for multiple testing using the Bonferroni correction (P-value = 3.04e-03; OR = 4.13 [2.46–5.81]); Table 6).

**Cis-expression Quantitative Trait Loci (cis-eQTL) Analyses and Prediction of Functional Effects.** To better define the role of *RNF122* in ADHD and to detect potential functional variants, we imputed markers at this locus in the original dataset of 603 subjects with ADHD and 583 healthy controls and found 47 out of 138 SNPs nominally associated with ADHD. Although none of them exceeded the conservative Bonferroni-corrected threshold, three surpassed the SNPSpD multiple-testing correction (rs3735951, rs9297208 and rs9297209) (See Supplementary Figure S1). Suggestive evidence for association was detected along the entire gene, with the top signal at the genotyped variant rs3735951, located in exon 2, being the allele T the risk one (P-value = 8.18e-04; OR = 1.37 [1.14–1.65]) (See Supplementary Figure S1). After conditional analysis, no evidence for additional independent effects was detected along the gene, being the entire association of this locus explained by the rs3735951 marker (See Supplementary Table S2). Although we did not find evidence for rs3735951 acting as cis-eQTL in preexisting datasets of cortical gene expression (GSE8919 and GSE30272)<sup>35,36</sup>, that could explain the *RNF122* expression differences identified in ADHD (See Supplementary Table S3), functional prediction revealed that rs3735951 may lie within a partially overlapped exonic splicing enhancer (ESE) site and an exonic splicing silencer (ESS) site (See Supplementary Table S4). While the rs3735951T risk allele may create an ESE site potentially targeted by the spliceosome factors SRSF2 and SRSF5, the rs3735951C allele may generate an ESS site (See Supplementary Table S4).

## Discussion

With the aim of uncovering new underlying genes involved in persistent ADHD and providing additional evidence for the contribution of previously identified genes, we performed gene-wide and pathway enrichment analyses in a pre-existing GWAS dataset of adult ADHD followed by gene expression profiling. Our results provide



preliminary evidence for genetic association between ADHD and the *RNF122* gene and abnormal *RNF122* expression levels in PBMCs of medication-naïve ADHD subjects. These findings highlight *RNF122* as a strong candidate for ADHD.

*RNF122* (RING Finger Protein 122) is an E3 ubiquitin ligase involved in the proteasome-mediated processing, trafficking, and degradation of proteins that acts as an essential mediator of the substrate specificity of ubiquitin ligation<sup>41</sup>. This finding is in line with previous genome-wide analyses supporting the involvement of genes related to the ubiquitination machinery in the genetic susceptibility to ADHD or attention function, including the *FBXO33* and *PARK2* genes, which also encode components of the E3 ubiquitin-protein ligase complex<sup>12, 42</sup>. Interestingly, the second most significant network identified in the present study revealed an indirect connection between *RNF122* and *FBXO33*, and included other best hits such as *MAB21L2* or *PEX19* (Tables 1 and 3). The ubiquitin-proteasome pathway controls wide-ranging functions in the central nervous system (CNS), including fine-tuning of synaptic connections during development and synaptic plasticity in the adult organism, and has been identified as a well-founded pathway for other psychiatric or neurological conditions, including bipolar disorder, schizophrenia, Alzheimer's disease, Parkinson's disease, Huntington's disease, intellectual disability or autism spectrum disorder<sup>43–50</sup>. In line with these findings, the identification of “Protein Kinase A Signaling” and “cAMP-mediated Signaling” as the two most significantly enriched canonical pathways may also support the role of the identified genes in adult synaptic plasticity mediated by the ubiquitin-proteasome system since ubiquitination, degradation and subsequent removal of regulatory subunits of protein Kinase A in response to cAMP stimulation induce neuronal differentiation and activity, as well as synaptic plasticity (Fig. 1)<sup>43, 44, 52, 53</sup>.

The overexpression of *RNF122* that we found in persistent ADHD is in agreement with previous studies reporting increased expression levels of E3 ubiquitin ligases in neuropsychiatric disorders<sup>51, 54</sup>. Given its role in synaptic surface-protein turnover and in the regulation of the number of scaffolding proteins and neurotransmitter receptors, increased ubiquitin ligation and subsequent enhanced protein removal may compromise neuronal functioning<sup>43, 55, 56</sup>. Although its consequences in the remodelling of the synaptic density should be further investigated, enhanced ubiquitin-proteasome pathway activity may entail downregulation of synaptic receptors, such as gamma-aminobutyric acid (GABA),  $\alpha$ -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid (AMPA), N-methyl-d-aspartate (NMDA) or nicotinic acetylcholine (nACh) receptors; and eventually lead to the neuro-pathological mechanisms that likely underlie neuropsychiatric disorders<sup>51, 54, 56–60</sup>.

We also identified a significant association between *CTAGE5* and persistent ADHD after Bonferroni correction. *CTAGE5*, together with other top hits identified in the gene-wide analyses, namely *FBXO33*, *MIA2* and *PNN*, lies within a genomic segment on chromosome 14 that was found to be deleted in three subjects with different neurologic and/or psychiatric traits, such as mild mental retardation, severe learning difficulties, motor alterations or ADHD symptoms<sup>61</sup>. *FBXO33* gene encodes a member of the F-box protein family which acts as a component of an E3 ubiquitin-protein ligase complex and is involved in targeting substrates for proteasomal degradation, while *CTAGE5* and *MIA2* genes are both involved in the traffic of large cargos from the endoplasmic reticulum to the Golgi apparatus, a process that involves ubiquitination<sup>62</sup>. Thus, together with the involvement of *RNF122* in ADHD, these findings highlight the need for further studies exploring in depth the potential role of the ubiquitin - proteasome pathway in the susceptibility to ADHD.

Additionally, nominal signals were identified for *C15orf53* and *FBXO33* in both the discovery and the replication datasets. Interestingly, *C15orf53* has previously been associated with bipolar disorder and alcohol dependence, and *FBXO33*, which may be involved in autism spectrum disorders, was highlighted in the previous SNP-based GWAS that we performed in the same discovery dataset and provides additional evidence supporting the ubiquitination machinery as a new mechanism for ADHD<sup>7, 22</sup>.

Our study should be viewed in light of several methodological considerations:

First, we attempted to overcome SNP-based GWAS limitations by using a gene-wide approach, a promising complement to GWAS since it considers the combined effects of all genetic variants within a locus and might be more powerful than traditional SNP-based strategies<sup>9</sup>. Our modest sample size and the anticipated small effect of common polymorphisms in complex traits, however, may have prevented us from detecting additional signals, apart from *CTAGE5*, exceeding the conservative Bonferroni correction. Additionally, although controls were screened retrospectively for ADHD symptoms, no specific scale or structured interview was used to discard the presence of other psychiatric disorders and, therefore, certain degree of heterogeneity in the control sample may exist and might influence our results.

Second, the conditional analysis highlighted rs3735951 as the top-ranked variant at the *RNF122* locus and eliminated evidence for association for other SNPs within the region. This sequence variant did not tag any eQTL for *RNF122* in human brain samples from two available pre-existing data sets<sup>35, 36</sup>, but lies within potential exonic splicing regulatory elements. Specifically, the rs3735951T risk allele is predicted to create an exonic splicing enhancer (ESE), potentially targeted by *SRSF2* and *SRSF5* spliceosome factors, which could result in alternative splicing of the *RNF122* transcript and, thus, modulate substrate specificity or ligation function (Supplementary Table S4)<sup>37–40</sup>. Given that rs3735951 lies within a LD block that spans the entire gene, however, we cannot discard additional relevant variants within the locus exerting functional effects.

Third, while we identified genes displaying biologically interesting functions and pathways pointing to the ubiquitin-proteasome pathway as a promising candidate system, no overlap was observed between our gene-wide results and previous findings in ADHD through either individual GWAS or meta-analysis<sup>2–4, 8</sup>. The replication attempt in an independent sample, however, yielded preliminary evidence for nominal association of ADHD with *FBXO33* and *C15orf53*. Heterogeneity between populations and differences in study design or in the proportion of persistent ADHD between datasets may account for discordant results across studies, making it difficult to establish direct comparisons between reports. For this reason, further replication of the gene-wide and gene expression results in independent cohorts are needed to confirm these associations and to estimate the magnitude of their effects.

Fourth, aberrant *RNF122* expression levels were detected in PBMCs of ADHD subjects, naive for pharmacological treatment. Transcriptome analysis in peripheral blood has become an increasingly useful tool in the search for biomarkers in multiple medical fields, including psychiatric disorders, given their great deal of potential for non-invasive screening, diagnosis and prognosis, or for differentiation of biological endophenotypes, development of targeted therapies and anticipation of clinical response or adverse effects<sup>63</sup>. Although whole blood shares substantial transcriptome similarities with different CNS tissues so as to use peripheral expression profiles as a surrogate for gene expression in the CNS, further evidence in brain tissues is required to assert the role of *RNF122* in the pathophysiology of ADHD<sup>64</sup>.

In conclusion, we performed gene-wide and pathway enrichment analyses using data from a pre-existing GWAS dataset of persistent ADHD and provided tentative evidence for the involvement of the *CTAGE5* and *RNF122* genes in the susceptibility to the disorder. We also detected overexpression of the *RNF122* gene in PBMCs of adult ADHD patients, placing this gene as a promising candidate for the disorder. The evidence provided by our findings point to the ubiquitin-proteasome system as a well-founded pathway involved in the etiology of ADHD. Further collaborative efforts are required to disentangle the exact molecular mechanisms by which *CTAGE5*, *RNF122*, and the ubiquitin-proteasome system may contribute to the pathophysiology of ADHD and other neuropsychiatric disorders.

## References

1. Faraone, S. V. *et al.* Attention-deficit/hyperactivity disorder. *Nat Rev Dis Primers*. **1**, 15020, doi:10.1038/nrdp.2015.20 (2015).
2. Gizer, I. R., Ficks, C. & Waldman, I. D. Candidate gene studies of ADHD: a meta-analytic review. *Hum Genet*. **126**, 51–90 (2009).
3. Franke, B. *et al.* The genetics of attention deficit/hyperactivity disorder in adults, a review. *Mol Psychiatry*. **17**, 960–87 (2012).
4. Li, Z., Chang, S. H., Zhang, L. Y., Gao, L. & Wang, J. Molecular genetic studies of ADHD and its candidate genes: a review. *Psychiatry Res*. **219**, 10–24 (2014).
5. Thapar, A. & Cooper, M. Attention deficit hyperactivity disorder. *Lancet*. **387**, 1240–50 (2016).
6. Cross-Disorder Group of the Psychiatric Genomics Consortium. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet*. **45**, 984–94 (2013).
7. Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet*. **381**, 1371–9 (2013).
8. Hawi, Z. *et al.* The molecular genetic architecture of attention deficit hyperactivity disorder. *Mol Psychiatry*. **20**, 289–97 (2015).
9. Wang, L., Jia, P., Wolfinger, R. D., Chen, X. & Zhao, Z. Gene set analysis of genome-wide association studies: methodological issues and perspectives. *Genomics*. **98**, 1–8 (2011).
10. Poelmans, G., Pauls, D. L., Buitelaar, J. K. & Franke, B. Integrated genome-wide association study findings: identification of a neurodevelopmental network for attention deficit hyperactivity disorder. *Am J Psychiatry*. **168**, 365–77 (2011).
11. Network and Pathway Analysis Subgroup of Psychiatric Genomics Consortium. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nat Neurosci*. **18**, 199–209 (2015).
12. Sánchez-Mora, C. *et al.* Case-control genome-wide association study of persistent attention-deficit hyperactivity disorder identifies FBXO33 as a novel susceptibility gene for the disorder. *Neuropsychopharmacology*. **40**, 915–26 (2015).
13. Epstein, J., Johnson, D. & Conners, K. *Adult ADHD Diagnostic Interview for DSM-IV*. (Multi-Healt Systems, 1999).
14. Conners, C. K. The computerized continuous performance test. *Psychopharmacol Bull*. **21**, 891–892 (1985).
15. DuPaul, G., Power, T., Anastopoulos, A. & Reid, R. *ADHD Rating Scales, IV: Checklists, Norms, and Clinical Interpretation*. (Guilford Press, 1998).
16. Ward, M. F., Wender, P. H. & Reimherr, F. W. The Wender Utah Rating Scale: an aid in the retrospective diagnosis of childhood attention deficit hyperactivity disorder. *Am J Psychiatry*. **150**, 885–890 (1993).
17. NIMH, C. G. I. Clinical Global Impression Scale. *Psychopharmacol Bull*. **21**, 839–844 (1985).
18. Sheehan, D. *The Anxiety Disease*. (Charles Scribner's Sons, 1983).
19. Ribases, M. *et al.* Exploration of 19 serotonergic candidate genes in adults and children with attention-deficit/hyperactivity disorder identifies association for 5HT2A, DDC and MAOB. *Mol Psychiatry*. **14**, 71–85 (2009).
20. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. **81**, 559–75 (2007).
21. Liu, J. Z. *et al.* A Versatile Gene-Based Test for Genome-wide Association Studies. *Am J Hum Genet*. **87**, 139–45 (2010).
22. Neale, B. M. *et al.* Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry*. **49**, 884–97 (2010).
23. Wojcik, G. L., Kao, W. H. & Duggal, P. Relative performance of gene- and pathway-level methods as secondary analyses for genome-wide association studies. *BMC Genet*. **16**, 34 (2015).
24. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol*. **11**, e1004219, doi:10.1371/journal.pcbi.1004219 (2015).
25. Ingenuity Systems, Redwood City, California, USA. *IPA Network Generation Algorithm* <https://www.ingenuity.com/wp-content/themes/ingenuity-qiagen/pdf/ipa/IPA-netgen-algorithm-whitepaper.pdf> (2005).
26. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Series B Stat Methodol*. **57**, 289–300 (1995).
27. Carvalho, B. S. & Irizarry, R. A. A framework for oligonucleotide microarray preprocessing. *Bioinformatics*. **26**, 2363–2367 (2010).
28. Ritchie, M. E. *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. **43**, e47, doi:10.1093/nar/gkv007 (2015).
29. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*. **25**, 402–408 (2001).
30. R Core Team. R Foundation for Statistical Computing, Vienna, Austria. *R: A language and environment for statistical computing* <https://www.R-project.org/> (2016)
31. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*. **5**, e1000529, doi:10.1371/journal.pgen.1000529 (2009).
32. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet*. **44**, 955–9 (2012).
33. 1000 Genomes Project Consortium. *et al.* A global reference for human genetic variation. *Nature*. **526**, 68–74 (2015).
34. Nyholt, D. R. A simple correction for multiple testing for SNPs in linkage disequilibrium with each other. *Am J Hum Genet*. **74**, 765–769 (2004).
35. Myers, A. J. *et al.* A survey of genetic human cortical gene expression. *Nat Genet*. **39**, 1494–9 (2007).
36. Colantuoni, C. *et al.* Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature*. **478**, 519–23 (2011).

37. Xu, Z. & Taylor, J. A. SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. *Nucleic Acids Res.* **37**, W600–5, doi:10.1093/nar/gkp290 (2009).
38. Smith, P. J. *et al.* An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers. *Hum. Mol. Genet.* **15**, 2490–2508 (2006).
39. Cartegni, L., Wang, J., Zhu, Z., Zhang, M. Q. & Krainer, A. R. ESEfinder: a web resource to identify exonic splicing enhancers. *Nucleic Acid Research.* **31**, 3568–3571 (2003).
40. Fairbrother, W. G., Yeh, R. F., Sharp, P. A. & Burge, C. B. Predictive identification of exonic splicing enhancers in human genes. *Science.* **297**, 1007–13 (2002).
41. Peng, Z., Shi, T. & Ma, D. RNF122: a novel ubiquitin ligase associated with calcium-modulating cyclophilin ligand. *BMC Cell Biol.* **11**, 41, doi:10.1186/1471-2121-11-41 (2010).
42. Jarick, I. *et al.* Genome-wide analysis of rare copy number variations reveals PARK2 as a candidate gene for attention-deficit/hyperactivity disorder. *Mol Psychiatry.* **19**, 115–21 (2014).
43. Hegde, A. N. The ubiquitin-proteasome pathway and synaptic plasticity. *Learn Mem.* **17**, 314–27 (2010).
44. Tai, H. C. & Schuman, E. M. Ubiquitin, the proteasome and protein degradation in neuronal function and dysfunction. *Nat Rev Neurosci.* **9**, 826–38 (2008).
45. Bousman, C. A. *et al.* Preliminary evidence of ubiquitin proteasome system dysregulation in schizophrenia and bipolar disorder: convergent pathway analysis findings from two independent samples. *Am J Med Genet B Neuropsychiatr Genet.* **153B**, 494–502 (2010).
46. Rubio, M. D., Wood, K., Haroutunian, V. & Meador-Woodruff, J. H. Dysfunction of the ubiquitin proteasome and ubiquitin-like systems in schizophrenia. *Neuropsychopharmacology.* **38**, 1910–20 (2013).
47. Lam, Y. A. *et al.* Inhibition of the ubiquitin-proteasome system in Alzheimer's disease. *Proc Natl Acad Sci USA.* **97**, 9902–6 (2000).
48. Shimura, H. *et al.* Ubiquitination of a new form of alpha-synuclein by parkin from human brain: implications for Parkinson's disease. *Science.* **293**, 263–9 (2001).
49. Rubinsztein, D. C. The roles of intracellular protein-degradation pathways in neurodegeneration. *Nature.* **443**, 780–786 (2006).
50. Glessner, J. T. *et al.* Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature.* **459**, 569–573 (2009).
51. Crider, A., Pandya, C. D., Peter, D., Ahmed, A. O. & Pillai, A. Ubiquitin-proteasome dependent degradation of GABAA $\alpha$ 1 in autism spectrum disorder. *Mol Autism.* **5**, 45, doi:10.1186/2040-2392-5-45 (2014).
52. Gong, B., Radulovic, M., Figueiredo-Pereira, M. E. & Cardozo, C. The Ubiquitin-Proteasome System: Potential Therapeutic Targets for Alzheimer's Disease and Spinal Cord Injury. *Front Mol Neurosci.* **9**, 4, doi:10.3389/fnmol.2016.00004 (2016).
53. Fioravante, D. & Byrne, J. H. Protein degradation and memory formation. *Brain Res Bull.* **85**, 14–20 (2011).
54. Pandya, C. D., Crider, A. & Pillai, A. Glucocorticoid regulates parkin expression in mouse frontal cortex: implications in schizophrenia. *Curr Neuropharmacol.* **12**, 100–7 (2014).
55. Yi, J. J. & Ehlers, M. D. Emerging roles for ubiquitin and protein degradation in neuronal function. *Pharmacol Rev.* **59**, 14–39 (2007).
56. Hamilton, A. M. & Zito, K. Breaking it down: the ubiquitin proteasome system in neuronal morphogenesis. *Neural Plast.* **2013**, 196848, doi:10.1155/2013/196848 (2013).
57. Speese, S. D., Trotta, N., Rodesch, C. K., Aravamudan, B. & Broadie, K. The ubiquitin proteasome system acutely regulates presynaptic protein turnover and synaptic efficacy. *Curr Biol.* **13**, 899–910 (2003).
58. Ferreira, J. S. *et al.* GluN2B-Containing NMDA Receptors Regulate AMPA Receptor Traffic through Anchoring of the Synaptic Proteasome. *J Neurosci.* **35**, 8462–79 (2015).
59. Rezvani, K., Teng, Y. & De Biasi, M. The ubiquitin-proteasome system regulates the stability of neuronal nicotinic acetylcholine receptors. *J Mol Neurosci.* **40**, 177–84 (2010).
60. Yuen, E. Y. *et al.* Repeated stress causes cognitive impairment by suppressing glutamate receptor expression and function in prefrontal cortex. *Neuron.* **73**, 962–77 (2012).
61. de Bruijn, D. R. *et al.* Severe Progressive Autism Associated with Two de novo Changes: A 2.6-Mb 2q31.1 Deletion and a Balanced t(14;21)(q21.1;p11.2) Translocation with Long-Range Epigenetic Silencing of LRFN5 Expression. *Mol Syndromol.* **1**, 46–57 (2010).
62. Jin, L. *et al.* Ubiquitin-dependent regulation of COPII coat size and function. *Nature.* **482**, 495–500 (2012).
63. Strimbu, K. & Tavel, J. A. What are biomarkers? *Curr Opin HIV AIDS.* **5**, 463–6 (2010).
64. Sullivan, P. F., Fan, C. & Perou, C. M. Evaluating the comparability of gene expression in blood and brain. *Am J Med Genet B Neuropsychiatr Genet.* **141B**, 261–268 (2006).
65. Wang, K. *et al.* Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature.* **459**, 528–33 (2009).
66. Uher, R. *et al.* Common genetic variation and antidepressant efficacy in major depressive disorder: a meta-analysis of three genome-wide pharmacogenetic studies. *Am J Psychiatry.* **170**, 207–17 (2013).
67. Ripke, S. *et al.* A mega-analysis of genome-wide association studies for major depressive disorder. *Mol Psychiatry.* **18**, 497–511 (2013).
68. Ryan, M. M. *et al.* Gene expression analysis of bipolar disorder reveals downregulation of the ubiquitin cycle and alterations in synaptic genes. *Mol Psychiatry.* **11**, 965–78 (2006).
69. Sesele, K. *et al.* Conditional inactivation of nicastrin restricts amyloid deposition in an Alzheimer's disease mouse model. *Aging Cell.* **12**, 1032–40 (2013).
70. Hatzimanolis, A. *et al.* Multiple variants aggregate in the neuregulin signaling pathway in a subset of schizophrenia patients. *Transl Psychiatry.* **3**, e264, doi:10.1038/tp.2013.33 (2013).
71. Andrus, B. M. *et al.* Gene expression patterns in the hippocampus and amygdala of endogenous depression and chronic stress models. *Mol Psychiatry.* **17**, 49–61 (2012).
72. Sannino, G. *et al.* Acute stress enhances the expression of neuroprotection- and neurogenesis-associated genes in the hippocampus of a mouse restraint model. *Oncotarget.* **7**, 8455–65 (2016).
73. Williams, N. M. *et al.* Rare chromosomal deletions and duplications in attention-deficit hyperactivity disorder: a genome-wide analysis. *Lancet.* **376**, 1401–8 (2010).
74. Martin, J. *et al.* Biological overlap of attention-deficit/hyperactivity disorder and autism spectrum disorder: evidence from copy number variants. *J Am Acad Child Adolesc Psychiatry.* **53**(761–70), e26, doi:10.1016/j.jaac.2014.03.004 (2014).
75. Fernández-Castillo, N. *et al.* Transcriptomic and genetic studies identify NFAT5 as a candidate gene for cocaine dependence. *Transl Psychiatry.* **5**, e667, doi:10.1038/tp.2015.158 (2015).
76. Datson, N. A. *et al.* Hippocampal CA1 region shows differential regulation of gene expression in mice displaying extremes in behavioral sensitization to amphetamine: relevance for psychosis susceptibility? *Psychopharmacology (Berl).* **217**, 525–38 (2011).
77. Malki, K. *et al.* Transcriptome analysis of genes and gene networks involved in aggressive behavior in mouse and zebrafish. *Am J Med Genet B Neuropsychiatr Genet.* **171**, 827–38 (2016).
78. Smith, M. *et al.* Molecular genetic delineation of a deletion of chromosome 13q12->q13 in a patient with autism and auditory processing deficits. *Cytogenet Genome Res.* **98**, 233–9 (2002).
79. Chow, M. L. *et al.* Age-dependent brain gene expression and copy number anomalies in autism suggest distinct pathological processes at young versus mature ages. *PLoS Genet.* **8**, e1002592, doi:10.1371/journal.pgen.1002592 (2012).

80. Horn, D. *et al.* A Novel Oculo-Skeletal syndrome with intellectual disability caused by a particular MAB21L2 mutation. *Eur J Med Genet.* **58**, 387–91 (2015).
81. Berko, E. R. *et al.* Calder RB1 Mosaic epigenetic dysregulation of ectodermal cells in autism spectrum disorder. *PLoS Genet.* **10**, e1004402, doi:10.1371/journal.pgen.1004402 (2014).
82. Tian, Y. *et al.* Correlations of gene expression with ratings of inattention and hyperactivity/impulsivity in Tourette syndrome: a pilot study. *BMC Med Genomics.* **5**, 49 (2012).
83. Szatkiewicz, J. P. *et al.* Copy number variation in schizophrenia in Sweden. *Mol Psychiatry.* **19**, 762–73 (2014).
84. Ament, S. A. *et al.* Rare variants in neuronal excitability genes influence risk for bipolar disorder. *Proc Natl Acad Sci USA.* **112**, 3576–81 (2015).
85. García-Fuster, M. J. *et al.* FADD adaptor and PEA-15/ERK1/2 partners in major depression and schizophrenia postmortem brains: basal contents and effects of psychotropic treatments. *Neuroscience.* **277**, 541–51 (2014).
86. Adkins, D. E. *et al.* Genome-Wide Meta-Analysis of Longitudinal Alcohol Consumption Across Youth and Early Adulthood. *Twin Res Hum Genet.* **18**, 335–47 (2015).
87. Georgi, B. *et al.* Genomic view of bipolar disorder revealed by whole genome sequencing in a genetic isolate. *PLoS Genet.* **10**, e1004229, doi:10.1371/journal.pgen.1004229 (2014).
88. Anney, R. *et al.* A genome-wide scan for common alleles affecting risk for autism. *Hum Mol Genet.* **19**, 4072–82 (2010).
89. Codina-Solà, M. *et al.* Integrated analysis of whole-exome sequencing and transcriptome profiling in males with autism spectrum disorders. *Mol Autism.* **6**, 21 (2015).
90. Fromer, M. *et al.* Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat Neurosci.* **19**, 1442–53 (2016).
91. Walton, E. *et al.* Correspondence of DNA Methylation Between Blood and Brain Tissue and Its Application to Schizophrenia Research. *Schizophr Bull.* **42**, 406–14 (2016).
92. Chen, Q. *et al.* Association study of CSF2RB with schizophrenia in Irish family and case – control samples. *Mol Psychiatry.* **13**, 930–8 (2008).
93. Chen, P. *et al.* Common SNPs in CSF2RB are associated with major depression and schizophrenia in the Chinese Han population. *World J Biol Psychiatry.* **12**, 233–8 (2011).
94. Moskvina, V. *et al.* Gene-wide analyses of genome-wide association data sets: evidence for multiple common risk alleles for schizophrenia and bipolar disorder and for overlap in genetic risk. *Mol Psychiatry.* **14**, 252–60 (2009).
95. Poelmans, G., Franke, B., Pauls, D. L., Glennon, J. C. & Buitelaar, J. K. AKAPs integrate genetic findings for autism spectrum disorders. *Transl Psychiatry.* **3**, e270, doi:10.1038/tp.2013.48 (2013).
96. Saia-Cereda, V. M. *et al.* Proteomics of the corpus callosum unravel pivotal players in the dysfunction of cell signaling, structure, and myelination in schizophrenia brains. *Eur Arch Psychiatry Clin Neurosci.* **265**, 601–12 (2015).
97. Iwamoto, K., Ueda, J., Bundo, M., Kojima, T. & Kato, T. Survey of the effect of genetic variations on gene expression in human prefrontal cortex and its application to genetics of psychiatric disorders. *Neurosci Res.* **70**, 238–42 (2011).
98. Li, L. *et al.* Decreased Na<sup>+</sup>/K<sup>+</sup> ATPase  $\alpha$ 1 (ATP1A1) gene expression in major depression patients' peripheral blood. *Cent Eur J Biol.* **8**, 1077–1082 (2013).
99. Zhao, J. *et al.* Correlative study of peripheral ATP1A1 gene expression level to anxiety severity score on major depressive disorder patients. *J Basic Clin Physiol Pharmacol.* **27**, 563–67 (2016).
100. Ji, L., Chauhan, A., Brown, W. T. & Chauhan, V. Increased activities of Na<sup>+</sup>/K<sup>+</sup>–ATPase and Ca<sup>2+</sup>/Mg<sup>2+</sup>–ATPase in the frontal cortex and cerebellum of autistic individuals. *Life Sci.* **85**, 788–93 (2009).
101. Goldstein, I. *et al.* Association between sodium- and potassium-activated adenosine triphosphatase alpha isoforms and bipolar disorders. *Biol Psychiatry.* **65**, 985–91 (2009).
102. Nazeen, S., Palmer, N. P., Berger, B. & Kohane, I. S. Integrative analysis of genetic data sets reveals a shared innate immune component in autism spectrum disorder and its co-morbidities. *Genome Biol.* **17**, 228 (2016).
103. Woodbury-Smith, M. *et al.* Using extended pedigrees to identify novel autism spectrum disorder (ASD) candidate genes. *Hum Genet.* **134**, 191–201 (2015).
104. Nebel, R. A. *et al.* Reciprocal Relationship between Head Size, an Autism Endophenotype, and Gene Dosage at 19p13.12 Points to AKAP8 and AKAP8L. *PLoS One.* **10**, e0129270, doi:10.1371/journal.pone.0129270 (2015).
105. Bergen, S. E. *et al.* Genome-wide association study in a Swedish population yields support for greater CNV and MHC involvement in schizophrenia compared with bipolar disorder. *Mol Psychiatry.* **17**, 880–6 (2012).
106. Psychosis Endophenotypes International Consortium. *et al.* A genome-wide association analysis of a broad psychosis phenotype identifies three loci for further investigation. *Biol Psychiatry.* **75**, 386–97 (2014).
107. Karszen, A. M. *et al.* Stress-induced changes in primate prefrontal profiles of gene expression. *Mol Psychiatry.* **12**, 1089–102 (2007).
108. Smith, E. N. *et al.* Genome-wide association study of bipolar disorder in European American and African American individuals. *Mol Psychiatry.* **14**, 755–63 (2009).
109. Cotney, J. *et al.* The autism-associated chromatin modifier CHD8 regulates other autism risk genes during human neurodevelopment. *Nat Commun.* **6**, 6404 (2015).
110. Tallaksen, C. M. *et al.* Subtle cognitive impairment but no dementia in patients with spastin mutations. *Arch Neurol.* **60**, 1113–8 (2003).
111. Herberth, M. *et al.* Peripheral profiling analysis for bipolar disorder reveals markers associated with reduced cell survival. *Proteomics.* **11**, 94–105 (2011).
112. Sun, L., Cheng, Z., Zhang, F. & Xu, Y. Gene expression profiling in peripheral blood mononuclear cells of early-onset schizophrenia. *Genom Data.* **5**, 169–70 (2015).
113. Quintela, I. *et al.* Female patient with autistic disorder, intellectual disability, and co-morbid anxiety disorder: Expanding the phenotype associated with the recurrent 3q13.2-q13.31 microdeletion. *Am J Med Genet A.* **167A**, 3121–9 (2015).
114. Wang, J. C. *et al.* A genome-wide association study of alcohol-dependence symptom counts in extended pedigrees identifies C15orf53. *Mol Psychiatry.* **18**, 1218–24 (2013).
115. O'Roak, B. J. *et al.* Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature.* **485**, 246–50 (2012).
116. Gai, X. *et al.* Rare structural variation of synapse and neurotransmission genes in autism. *Mol Psychiatry.* **17**, 402–11 (2012).
117. Hill, W. D. *et al.* Examining non-syndromic autosomal recessive intellectual disability (NS-ARID) genes for an enriched association with intelligence differences. *Intelligence.* **54**, 80–9 (2016).
118. Najmabadi, H. *et al.* Deep sequencing reveals 50 novel genes for recessive cognitive disorders. *Nature.* **478**, 57–63 (2011).
119. Matsunami, N. *et al.* Identification of rare DNA sequence variants in high-risk autism families and their prevalence in a large case/control population. *Mol Autism.* **5**, 5 (2014).
120. Liao, H. M. *et al.* Clinical and molecular characterization of a transmitted reciprocal translocation t(1;12)(p32.1;q21.3) in a family co-segregating with mental retardation, language delay, and microcephaly. *BMC Med Genet.* **12**, 70 (2011).
121. Kastenberger, I. & Schwarzer, C. GPER1 (GPR30) knockout mice display reduced anxiety and altered stress response in a sex and paradigm dependent manner. *Horm Behav.* **66**, 628–36 (2014).
122. Fındıklı, E. *et al.* Serum levels of G protein-coupled estrogen receptor 1 (GPER1) in drug-naive patients with generalized anxiety disorder. *Psychiatry Res.* **244**, 312–6 (2016).

123. Munkholm, K., Peijs, L., Vinberg, M. & Kessing, L. V. A composite peripheral blood gene expression measure as a potential diagnostic biomarker in bipolar disorder. *Transl Psychiatry*. 5, e614, doi:10.1038/tp.2015.110 (2015).
124. Crespo-Facorro, B., Prieto, C. & Sainz, J. Schizophrenia gene expression profile reverted to normal levels by antipsychotics. *Int J Neuropsychopharmacol*. 18, pii: pyu066;10.1093/ijnp/pyu066 (2014).
125. Latham, C. F. *et al.* Molecular dissection of the Munc18c/syntaxin4 interaction: implications for regulation of membrane trafficking. *Traffic*. 7, 1408–19 (2006).

## Acknowledgements

We are grateful to patients and controls who kindly participated in this research. The microarrays service was carried out at High Technology Unit (UAT) at Vall d'Hebron Research Institute (VHIR), Barcelona (Spain) and the Statistics and Bioinformatics Unit (UEB) at the VHIR provided statistical support. Over the course of this investigation, I.G. M. has been a recipient of a predoctoral fellowship from the Vall d'Hebron Research Institute (PRED-VHIR-2012), Barcelona, Spain, and currently she is a recipient of a contract from the 7th Framework Programme for Research, Technological Development and Demonstration, European Commission (AGGRESSOTYPE\_FP7HEALTH2013/602805). C.S. M. is a recipient of a Sara Borrell contract from the Instituto de Salud Carlos III, Ministerio de Economía, Industria y Competitividad, Spain (CD15/00199) and a mobility grant from the Instituto de Salud Carlos III, Ministerio de Economía, Industria y Competitividad, Spain (MV16/00039). M.S. A. is a recipient of a contract from the Biomedical Research Networking Centre in Mental Health (CIBERSAM), Madrid, Spain. P.R. is a recipient of a predoctoral fellowship from the Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR), Generalitat de Catalunya, Spain (2016FI\_B 00899). M.P. is recipient of a predoctoral fellowship from the Vall d'Hebron Research Institute, Barcelona, Spain (PRED-VHIR-2013) and a research grant from the Deutscher Akademischer Austauschdienst (DAAD), Germany (Research Grants - Short-Term Grants, 2017). E.C. S. is a recipient of a predoctoral fellowship from the Collaborative Research Training Programme for Medical Doctors (PhD4MD), Centre for Genomic Regulation (CRG) and Vall d'Hebron Research Institute (VHIR), Barcelona, Spain (II14/00018). M.R. is a recipient of a Miguel de Servet contract from the Instituto de Salud Carlos III, Ministerio de Economía, Industria y Competitividad, Spain (CP09/00119 and CPII15/00023). This investigation was supported by Instituto de Salud Carlos III (PI12/01139, PI14/01700, PI15/01789, PI16/01505), and cofinanced by the European Regional Development Fund (ERDF), Agència de Gestió d'Ajuts Universitaris i de Recerca-AGAUR, Generalitat de Catalunya (2014SGR1357, 2014SGR0932), Ministerio de Economía, Industria y Competitividad, Spain (SAF2012-33484, SAF2015-68341-R), the European College of Neuropsychopharmacology (ECNP network: 'ADHD across the lifespan'), Departament de Salut, Generalitat de Catalunya, Spain, and a NARSAD Young Investigator Grant from the Brain & Behavior Research Foundation. This project has also received funding from the European Union's Horizon 2020 Research and Innovation Programme under the grant agreements No 667302 and 643051. This paper reflects only the authors' views, and the European Union is not liable for any use that may be made of the information contained therein.

## Author Contributions

I.G.M., C.S.M., M.P., E.C. and P.R. participated in the DNA and RNA isolation and preparation of samples. I.G.M., C.S.M. and M.S. undertook the statistical analyses. V.R. and E.C. participated in the clinical assessment and in the recruitment of patients. J.A.R.Q. and M.C. participated in the study design, clinical assessment and coordination of the clinical research. M.R. conceived of the project concept, wrote the protocol, and coordinated the study design and the statistical analyses. M.R., J.A.R.Q. and B.C. supervised the project and the manuscript preparation. All authors contributed to and have approved the final manuscript.

## Additional Information

**Supplementary information** accompanies this paper at doi:10.1038/s41598-017-05514-7

**Competing Interests:** Prof. M.C. has received travel grants and research support from Eli Lilly and Co., Janssen-Cilag, Shire, and Laboratorios Rubió. He has been on the advisory board and served as a consultant for Eli Lilly and Co., Janssen-Cilag, Shire, and Laboratorios Rubió. Dr. J.A. R. Q. has served on the speakers' bureau and acted as consultant for Eli Lilly and Co., Janssen-Cilag, Novartis, Lundbeck, Shire, Ferrer, and Laboratorios Rubió. He has received travel awards from Eli Lilly and Co., Janssen-Cilag, and Shire for participating in psychiatric meetings. The ADHD Program chaired by Dr. J.A. R. Q. has received unrestricted educational and research support from Eli Lilly and Co., Janssen-Cilag, Shire, Ferrer, Rovi, and Laboratorios Rubió in the past two years.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017