Rare DNA variants in the brain derived neurotrophic factor (BDNF) gene increase risk for attention deficit hyperactivity disorder (ADHD): a next generation sequencing study

Ziarih Hawi[1φ] PhD, Tarrant D.R Cummins[1φ] PhD, Janette Tong[1] PhD, Mauricio Arcos-Burgos[2] PhD, Qiongyi Zhao[3] PhD, Natasha Matthews[3] PhD, Daniel P. Newman[1] BPsySc, Beth Johnson[1] PhD, Alasdair Vance[4] PhD, Helen S. Heussler[5] FRACP DM, Florence Levy[6] PhD, Simon Easteal[2] PhD, Naomi Wray[3] PhD, Elaine Kenny[7] PhD, Derek Morris[8] PhD, Lindsey Kent[9] PhD, Michael Gill[7] PhD, and Mark A. Bellgrove PhD [1, 3]

φ These authors contributed equally and should both be considered as first author

1. School of Psychological Sciences and Monash Institute for Cognitive and Clinical Neurosciences (MICCN), Monash University, Melbourne, Australia
2. John Curtin School of Medical Research, Australian National University, Canberra, Australia
3. Queensland Brain Institute, The University of Queensland, Brisbane, Australia
4. Academic Child Psychiatry Unit, Department of Paediatrics, University of Melbourne; Royal Children's Hospital, Melbourne, Australia
5. Mater Research Institute, University of Queensland, Brisbane, Australia
6. School of Psychiatry, University of New South Wales, Sydney, Australia; Child and Family East, Prince of Wales Hospital, Australia
7. Neuropsychiatric Genetics Research Group, Dept. of Psychiatry and Institute of Molecular Medicine, Trinity College Dublin, Dublin, Ireland
8. Department of Biochemistry, National University of Ireland Galway, Galway, Ireland
9. School of Medicine, University of St Andrews, St Andrews, Scotland, UK

Address correspondence to:

Professor Mark A. Bellgrove
School of Psychological Sciences
18 Innovation Walk, Clayton Campus,
Wellington Road,
Monash University, VIC, 3800
Australia
Ph: +61 3 9902 4200
Email: mark.bellgrove@monash.edu

Attention deficit hyperactivity disorder (ADHD) is a prevalent and highly heritable disorder of childhood with negative lifetime outcomes. Although candidate gene and genome wide association studies have identified promising common variant signals, these explain only a fraction of the heritability of ADHD. The observation that rare structural variants confer substantial risk to psychiatric disorders suggests that rare variants might explain a portion of the missing heritability for ADHD. Here we performed the first large-scale next generation targeted sequencing study of ADHD in 152 child and adolescent cases and 188 controls across an *a priori* set of 117 genes. A multi-marker gene level analysis of rare (<1% frequency) single nucleotide variants (SNVs) revealed that the gene encoding brain derived neurotrophic factor (*BDNF*) was associated with ADHD at Bonferroni corrected levels. Sanger sequencing confirmed the existence of all novel rare *BDNF* variants. Our results implicate BDNF as a genetic risk factor for ADHD, potentially by virtue of its critical role in neurodevelopment and synaptic plasticity.

ADHD is a highly heritable psychiatric condition with a worldwide prevalence in childhood and adolescence of approximately 5.3%[1]. ADHD is characterised by a persistent pattern of inattention and/or hyperactivity-impulsivity that interferes with social, academic or occupational functioning (DSM V[2]). These symptoms are chronic, persist into adulthood in approximately half of affected individuals and are associated with impaired family and peer relationships, increased risk for drug abuse and criminality and significantly increased mortality rates[3].

Large twin studies indicate that the key aetiological factors of ADHD are genetic, with heritability estimated at 75-90%[4]. Although recent polygenic risk analyses suggest that a

substantial portion of this heritability can be explained by the cumulative effects of common genetic variants, approximately 70% of the heritability of ADHD remains unexplained[5]. As a growing body of evidence now supports the involvement of rare variants in the aetiology of common diseases such as ADHD, it is possible that much of the so called 'missing heritability' of ADHD will be accounted for by rare variants. In line with this hypothesis, rare structural DNA abnormalities known as copy number variations (CNVs) have been implicated in the aetiology of several psychiatric disorders including autism, schizophrenia and ADHD. For example, a number of CNV studies have demonstrated that children with ADHD have a significant excess of large CNVs[6,7]. Yet while the CNV work represents a promising start to investigating the role of rare variants in ADHD, the majority of implicated CNVs are not highly penetrant and have limited overlap between studies (see Hawi 2015 for a review[8]).

Aside from CNVs, the most prevalent rare variants in the genome are single nucleotide variants (SNVs) and insertions/deletions (InDels). Promising results for analyses focusing on SNVs and InDels are just beginning to emerge for psychiatric phenotypes. For example Kenny et al[9] recently employed next generation sequencing and observed an increased overall burden of SNVs and InDels in both schizophrenia and autism spectrum disorder. Hence these two classes of rare variation represent a sensible starting point for investigations of the role of rare variants in ADHD. To define the search space for potential rare variant associations with ADHD we curated the published molecular genetic literature of ADHD to yield a set of genes for targeted (exon plus untranslated region (UTR)) sequencing that had *a priori* evidence of association. We included genes showing evidence of association with ADHD from candidate gene, genome wide association (GWA) and CNV studies. We then conducted burden analyses for SNVs and InDels at the level of each gene. These analyses indicated for the first

time that rare variation in the brain derived neurotrophic factor gene (*BDNF*) is associated with ADHD.

Methods

Gene Selection

Recent research indicates that ADHD is associated with a significant overlap of biological pathways enriched for both common variants and rare CNVs[8,10,11], suggesting that perturbation in these pathways is critical for conferring risk to ADHD. Thus we restricted the search space for rare SNVs and InDels that may be associated with ADHD by focusing on those genes implicated in previous research. We conducted a comprehensive literature review in order to identify genes showing evidence of association with ADHD under candidate gene and genome wide association (GWA) designs. Specifically, genes were selected from candidate gene studies if the original report of association with ADHD was replicated via independent GWAS, meta-analysis or at least one independent candidate gene study (both sources significant at $p<.05$). Two additional candidate genes (*PER2, CSNK1E*) were included because an association with ADHD at $p<.05$ was supported by very strong functional evidence. We also included genes from childhood ADHD GWA studies[12-15] where variants were associated with ADHD at $p<1*10^{-4}$ and the copy number variant GWAS (CNV-GWAS) of Williams et al (2010)[6] when 2 or more ADHD cases (but not controls) showed CNV's that overlapped a given gene (we note that the latter genes were selected just prior to the Williams et al, 2012, multisite CNV study). The application of these inclusion criteria yielded 117 genes for targeted sequencing (31, 48, 38 genes; candidate gene, SNP-GWAS, CNV-GWAS, respectively). See **Supplementary Table 1**.

Participants

Our sample comprised 152 ADHD cases and 188 controls from amongst subjects recruited in the Republic of Ireland[11,16]. Recruitment and ascertainment of ADHD cases was approved by the Eastern Regional Health Authority research ethics committee and written informed consent was obtained from parents. The control subjects were ascertained with written informed consent from the Irish GeneBank. Detailed demographic, epidemiological and clinical descriptions of these samples have been presented previously in separate GWAS[16-18]. Briefly, the ADHD sample was composed of clinically diagnosed ADHD children and adolescents recruited from child guidance clinics and ADHD support groups. Exclusion criteria included: epilepsy, fragile X syndrome, foetal alcohol syndrome, pervasive developmental disorder, Tourette syndrome, psychosis and IQ < 70. To confirm DSM-IV and/or ICD-10 diagnoses of ADHD, one or both parents of each child were interviewed using the child and adolescent psychiatric assessment (CAPA)[19]. DSM-IV[20] and ICD-10[21] criteria for symptom pervasiveness were confirmed by obtaining information about ADHD symptoms at school via a semi-structured teacher telephone interview. The sample of ADHD cases had a mean age at collection of 10.10 years (*SD* = 3.49). The control sample was composed of blood donors to the Irish Blood Transfusion Service that had given informed consent to the Trinity Biobank[16]. As individuals regularly taking prescription medication (including for psychiatric and other disorders) are excluded from blood donation in Ireland and, as the lifetime prevalence of ADHD is relatively low, there is no obvious reason to expect that individuals with ADHD would be over represented in this control sample. The control sample had a mean age at collection of 44.50 years (*SD* = 12.49).

DNA Isolation, Targeted Sequencing and Quality Control

Identical DNA collection and isolation protocols were used for the case and control samples. DNA was isolated from blood using standard procedures. The physical condition of

the DNA was comparable for both samples and the purity measured as optical density at 260/280 ranged between 1.7-1.9 indicating good DNA quality. An Agilent SureSelect (Agilent Technologies, Santa Clara, CA, USA) custom capture target was designed that contained capture baits for the coding and UTR regions of genes of interest, including 117 genes specifically chosen for this study. In order to increase the likelihood that we could target each region with at least two SureSelect 120mer probes, any target region under 120 bases was artificially inflated to 121 bases. This final co-ordinate list was uploaded to Agilent Technology's eArray design website to design the SureSelect Target Enrichment array. Library preparation of each sample involved a two-step process. In the first step, the DNA was prepared as an Illumina sequencing library, and in the second step, the sequencing library was enriched for the desired target using the Agilent SureSelect enrichment protocol (as described by Kenny et al, 2014[9]) See **Supplementary Information**. Next generation sequencing was then conducted on an Illumina HiSeq2000 at the Queensland Brain Institute, Brisbane, Australia. Image processing and sequence extraction were performed using the standard Illumina Genome Analyzer software and the quality of raw reads was evaluated by FastQC (v0.10.1 devel) software. A custom PERL script was then used to de-multiplex the samples and generate the short reads for each sample in "fastq" format.

Approximately 20Gb of sequence data was generated for a total sample of 340 individuals. Sequence alignment and calling of SNVs and InDels involved mapping with BWA v0.6.2[22] and SNV and InDel identification with GATK v 2.2-8[23,24]. To ensure high quality variant calling by GATK, we filtered variants by following the GATK Best Practice Variant Detection protocol. Briefly, we aligned the paired-end reads to the human reference genome (B37) with BWA followed by marking of duplicates with Picard software (http://picard.sourceforge.net). We then applied the GATK genotyping pipeline that includes

"base quality score recalibration", "InDel realignment", and "multi-sample SNP and InDel calling". SNP and InDel discovery and genotyping were performed across all 340 samples simultaneously (multi-sample calling) by using GATK Unifiedgenotyper. The GATK Best Practice Variant Detection protocol excluded SNVs with: a quality by depth score (QD) < 2.0, a mapping quality score (MQ) < 40.0, a mapping quality rank sum score (MQranksum) < 12.5, a Fisher strand score (FS) > 60.0, a haplotype score> 13.0 or a read position rank sum test score (ReadPosRankSum) < -8.0. These filters ensured that: 1) there was a high variant call confidence based on unfiltered depth of non-reference samples (QD); 2) the mapping quality of the reads across all samples was high (MQ, MQranksum); 3) there was low strand bias for detection of variants (FS) - as strand bias is indicative of false positive calls; 4) the site was consistent with two, and only two segregating haplotypes (HaplotypeScore) - as high scores are indicative of regions with bad alignments, often leading to artifactual SNP calls; and 5) positions of any detected variants were not biased to the end of the reads (ReadPosRankSum) - as alternate allele reads that only occur near the end of the read are indicative of error. The GATK Best Practice protocol was also applied to InDels to remove variant calls with: QD < 2.0, ReadPosRankSum < -8, FS >200 and inbreeding coefficients < -0.8 (the former 3 filters were employed for aforementioned reasons and the inbreeding coefficient filter accounted for inbreeding among samples).

A further genotyping quality control (QC) was then conducted in SNP and Variation Suite v8.3 (SVS 8.3; Golden Helix, Inc., Bozeman, MT). For SNV data we removed variants that: 1) were monoallelic or had more than one alternative allele; 2) had an average read depth (across samples) <10x, 3) had a genotyping fail rate across all samples >10%. We also applied a filter to remove subjects with a genotyping fail rate across all variants >10%,

however the high quality of the data entailed that no subjects were removed under this criterion.

Relatedness and population stratification analyses

Analyses for relatedness and population stratification were also performed using SVS 8.3. An analysis for relatedness amongst individuals in the sequencing sample was conducted with the data from the corresponding subset of GWAS participants (n=340). The genotyping for the ADHD- GWAS analysis had previously been performed on an Illumina Human 660W-Quad BeadChip, while the control samples had been genotyped using the Affymetrix 6.0 platform. The SNPs that were common to both case and control GWAS data sets (139,100 SNPs) provided ample data for relatedness analyses. The combined data set was LD pruned (window size: 100, window increment: 5, LD r^2 threshold=.5, LD computation: CHM) to leave 101,163 SNPs for identity by state (IBS) estimation. IBS values of 1 and .309 were obtained for a control pair and a case/control pair respectively. These IBS values were consistent with: the existence of a duplicate sample in the Control group (IBS=1) and a second degree relative case/control pairing (IBS=.309). Consequently three individuals on which sequencing had been conducted were excluded, leaving a sample of 337 subjects.

In order to test for population structure, the combined case/control common SNV (MAF >5%) data arising from our next generation sequencing (N=337) was then submitted to a principal components analysis (with LD pruning conducted as above). In line with previous GWAS investigations that were conducted separately in the case and control samples[11,16], no ethnic sub-groups were observed. Thus a final sample of 337 participants (152 cases, 185 controls) remained for further analyses. The mean coverage in this final sample was greater than 100 times for both ADHD cases and controls (102.4x, 113.2x respectively; ≥ 40x mean coverage achieved across 98% of all samples) and we observed >99% matching between common

variant calls (MAF > .05) in our sequenced regions and corresponding GWAS data[11,16]. This level of coverage was associated with high quality data that was comparable across cases and controls; phred-scaled confidence values for genotype call were above 20 in 97.65% of cases and 97.16% of controls and above 99 in 85.04% of cases and 84.73% of controls.

We note that since our samples have contributed to previous (GWA) common variant[11] and CNV[7] studies we do not include such analyses in this paper.

Rare variant annotation and analysis

The final rare variant data set was constructed and analysed in SVS 8.3. Sequencing variants were classified as rare if they had a minor allele frequency (MAF) of <.01 in the combined case-control sample[9,25,26]. Our sample size of 337 subjects has power in excess of 80% to detect rare variants with a minor allele frequency ≥ .0024. Filtering for rare variants resulted in a set of 2702 rare SNV's which were then functionally annotated with the Combined Annotation Dependent Depletion (CADD) tool to provide both PHRED-scaled CADD scores (C-score) and information on variant consequences[27]. Since C scores >10 are indicative of a high likelihood of pathogenicity[27] (predicted to be in the top 10% of deleterious substitutions), we filtered for rare variants using this criterion in order to construct a set of putatively functional variants. This filtering process resulted in a set of 1288 rare putatively functional SNVs (hereafter referred to as *rare functional SNVs*). Gene level analyses were then conducted across our rare functional SNV set using the adaptive permutation version of the Kernel-Based Adaptive Cluster method (KBAC)[28]. For each gene we conducted 100,000 iterations and used a Bonferroni corrected threshold (.05/number of genes) to determine significance.

A complementary analysis tested for the impact of rare Loss-of-Function (LoF) variants (both SNVs and InDels); that is those variants that are predicted to severely disrupt protein-coding sequences. Following recent studies[9,29] we defined the following variant types as LoF if they were located in a proportion of protein-coding transcripts: nonsense SNVs that introduce stop codons; SNVs and InDels that disrupt splice sites; and InDels that disrupt a transcript's open reading frame. With these criteria, we observed a total of 10 unique LoF variants distributed over 10 genes (4 stop-gain SNVs, 5 splice site disruption SNVs, 1 frameshift InDel). For each of these genes we conducted Fisher's exact tests for association with case-control status and corrected for multiple comparisons with a Bonferroni correction, See **Supplementary Table 4**.

Results

Gene level analyses were conducted across our rare functional SNV set using the adaptive permutation version of the Kernel-Based Adaptive Cluster method (KBAC)[28]. As **Table 1** shows, this analysis revealed a significant association between ADHD and the gene encoding brain derived neurotrophic factor (*BDNF*), with 20 ADHD cases versus 6 controls showing variation across 12 rare SNVs, (*KBAC* stat = 8.594, $p = 3.00 \times 10^{-4}$; **Fig. 1**). Each of these individuals contributed a single rare variant with the exception of one ADHD case that had a rare *BDNF* variant at two positions (positions: 27695869, 27741140), indicating that this result was not driven by a high burden of rare variants in one or more cases. A further 2 genes (*ITPR2, CLYBL*) showed nominally significant associations (significant at uncorrected level, α=.05) **Supplementary Table 3**. In order to demonstrate the stability of our *BDNF* result we ran an alternative gene level analysis, the Weighted Sum Test[30], on our rare functional SNV set. Notably, this test also returned a significant result for *BDNF* at corrected levels, $p = 2.85 \times 10^{-4}$.

We next sought to validate each of the 12 *BDNF* SNVs identified in the above analysis. **Table 1** lists the chromosomal positions of each of the rare functional *BDNF* SNVs identified in our study. Sanger sequencing on the PCR product of these variants was conducted for every novel rare *BDNF* SNV. In every case Sanger sequencing confirmed the existence of the SNV confirming the fidelity of our next generation sequencing data.

Bioinformatic analysis of the 12 *BDNF* variants was then performed. We used the Alamut splicing module (Interactive Biosoftware, Rouen, France) to predict alternative splicing for variants located at exon/intron boundaries. Further, variants in proximity to the transcription start site were assessed for Cis-regulatory transcriptional control using data from the Encyclopedia of DNA Elements (ENCODE) consortium[31] and the US National Institutes of Health (NIH) Roadmap Epigenomics Project[32]. Although one novel non-synonymous variant was identified in the pro-region of BDNF (Position: 27695742), PolyPhen2 and SIFT both predicted the C-G substitution to be benign. Multiple variants were predicted to affect both splicing (Position: 27695893) and transcriptional control (Positions: 27722647; 27741048; 27741140; 27743438; 27743449; 27743481; 27743556) of *BDNF* (**Supplementary Table 2**).

Further analyses involved testing for the impact of rare LoF variants (SNVs and InDELs). Of the 10 genes that held LoF variants, 9 had insufficient variation to achieve the asymptotic properties of the Fisher's test statistic (see Keizun et al., 2012[33]). That is, for the majority of genes, the rarity of the LoF variants in combination with the sample size entailed that a p value <.05 was not achievable (regardless of how the variants were distributed between cases and controls). The one gene that was sufficiently powered (ZNF544) for a loss of function

gene-level analysis, did not show a significant case/control difference (**Supplementary Table 4**).

Discussion

It is widely appreciated that genetic risk for psychiatric disorders is conferred by multiple gene variants acting across a range of allele frequencies. Although the bulk of previous molecular genetic work in ADHD has focused on the contribution of relatively common gene variants, here we examined the contribution of rare genetic variation, assayed using next generation sequencing. We performed the first large-scale targeted (exon plus UTR) capture across 117 genes in 152 child and adolescent cases with ADHD compared to 188 controls. Our results show that ADHD is associated with an enrichment of putatively functional, rare SNVs mapped to the gene encoding brain derived neurotrophic factor (BDNF).

BDNF is critically important for neural development, differentiation and plasticity in both the developing and adult brain[34,35]. The relevance of BDNF to ADHD is underscored by several observations. First, BDNF conditional knockout mice display a phenotype that includes impulsivity, a core feature of ADHD[36,37]. Second, BDNF is an essential neurotrophic factor for both dopaminergic and serotonergic neurons, disruption of which has been implicated in the pathophysiology of ADHD[38-40]. Third, animal studies demonstrate that psychostimulants such as methylphenidate that are used in the treatment of ADHD, modulate BDNF expression[41,42]. A common exonic Val66Met substitution within the pro-region of *BDNF* has been widely implicated as contributing to a diverse range of cognitive functions and psychiatric disorders, including ADHD[35,43]. The Val66Met polymorphism is widely believed to disrupt activity-dependent release of BDNF with consequences for physiological processes

modulated by BDNF such as neurotransmitter release. Although we identified one non-synonymous variant located within the pro-region of *BDNF*, its rarity entails low LD between this variant and the Val66Met SNP. Further, bioinformatics suggested a low probability of pathogenicity. Our next generation sequencing nevertheless identified other variants with high likelihoods of pathogenicity including splice variants that could result in aberrant transcript processing of *BDNF*, and multiple variants located close to the 5′ UTR which are likely to result in changes to cis-regulatory transcriptional control of *BDNF*. The high likelihood of pathogenicity associated with the identified *BDNF* variants, in combination with the observation that all the associated BDNF transcripts are expressed in brain regions relevant to ADHD[44] such as the frontal cortex and cerebellum, highlights the potential relevance of our findings to the aetiology of ADHD.

Our novel application of next generation sequencing to ADHD raises several methodological issues that warrant comment. First, we conducted targeted exon plus UTR capture across 117 genes rather than an exome-wide approach. This approach was justified given emerging evidence of overlap between common and rare variant signals in complex traits and the high costs of whole-exome compared to targeted sequencing approaches at the time of commencement of this study. Our decision to restrict our analysis to variants with a high likelihood of pathogenicity based on CADD scores and the subsequent confirmation of all BDNF variants by Sanger sequencing, suggests that our study has likely identified a meaningful rare variant signal. Nevertheless, we recognise the inherent bias in our targeted sequencing approach and the requirement for replication in independent whole-exome sequencing studies. Second, although we identified an SNV association at Bonferroni corrected levels, our study was underpowered for a gene-level analysis of LoF variants. Our finding that only two genes carried sufficient LoF variants is consistent with past studies[9],

including a recent survey of LoF variants which observed an average of only 100 LoF variants across an individual's entire genome[29]. Together these results show that establishing an aetiological role for LoF variants at the gene level for complex traits will require extremely large samples. Third, our samples of cases and controls were not tightly matched on demographic variables such as age, rendering this a limitation of our study. Although the use of age-matched controls should be an aspiration for future sequencing work, it is worth noting that the use of opportunistic control cohorts is not uncommon in psychiatric genetic studies of both common and rare variants.

In summary, this study provides preliminary evidence that rare DNA variation in *BDNF*, putatively via its varying effects on neuro-developmental or plastic processes, confers risk to ADHD. Although our finding requires replication, it may encourage functional genomic exploration of the molecular risk mechanism.

Supplementary Information

Supplementary information is available at *Molecular Psychiatry's* website

Acknowledgments

This work was supported by a grant from the National Health and Medical Research Council of Australia (NHMRC) (APP1065677).

Conflict of Interest

The authors declare no conflicts of interest.

References

1. Polanczyk GV, Willcutt EG, Salum GA, Kieling C, Rohde LA. ADHD prevalence estimates across three decades: an updated systematic review and meta-regression analysis. *Int J Epidemiol* 2014; **43**: 434-42.

2. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders.* 5th ed Washington, DC, 2013.

3. Dalsgaard S, Østergaard SD, Leckman JF, Mortensen PB, Pedersen MG. Mortality in children, adolescents, and adults with attention deficit hyperactivity disorder: a nationwide cohort study. *Lancet* 2015; **385**: 2190-6.

4. Stergiakouli E, Thapar A. Fitting the pieces together: current research on the genetic basis of attention-deficit/hyperactivity disorder (ADHD). *Neuropsychiatr Dis Treat* 2010; **6**: 551-60.

5. Cross-Disorder Group of the Psychiatric Genomics Consortium. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet* 2013; **45**: 984–94.

6. Williams NM, Zaharieva I, Martin A, Langley K, Mantripragada K, Fossdal R, et al. Rare chromosomal deletions and duplications in attention-deficit hyperactivity disorder: a genome-wide analysis. *Lancet* 2010; **376**: 1401–8.

7. Williams NM, Franke B, Mick E, Anney RJL, Freitag CM, Gill M, et al. Genome-wide analysis of copy number variants in attention deficit hyperactivity disorder: the role of rare variants and duplications at 5q13.3. *Am J Psychiatry* 2012; **169**: 195–204.

8. Hawi Z, Cummins TDR, Tong J, Johnson B, Lau R, Samarrai W, et al. The molecular genetic architecture of attention deficit hyperactivity disorder. *Mol Psychiatry* 2015; **20**: 289-97.

9. Kenny EM, Cormican P, Furlong S, Heron E, Kenny G, Fahey C, et al. Excess of rare novel loss-of-function variants in synaptic genes in schizophrenia and autism spectrum disorders. *Mol Psychiatry* 2014; **19**: 872-9.

10. Elia J, Glessner JT, Wang K, Takahashi N, Shtir CJ, Hadley D, et al. Genome-wide copy number variation study associates metabotropic glutamate receptor gene networks with attention deficit hyperactivity disorder. *Nat Genet* 2012; **44**: 78–84.

11. Stergiakouli E, Hamshere M, Holmans P, Langley K, Zaharieva I, deCODE Genetics, et al. Investigating the contribution of common genetic variants to the risk and pathogenesis of ADHD. *Am J Psychiatry* 2012; **169**: 186-94.

12. Lasky-Su J, Neale BM, Franke B, Anney RJL, Zhou K, Maller JB, et al. Genome-wide association scan of quantitative traits for attention deficit hyperactivity disorder identifies novel associations and confirms candidate gene associations. *Am J Med Genet B Neuropsychiatr Genet* 2008; **147B**: 1345–54.

13. Neale BM, Lasky-Su J, Anney R, Franke B, Zhou K, Maller JB, et al. Genome-wide association scan of attention deficit hyperactivity disorder. *Am J Med Genet B Neuropsychiatr Genet* 2008; **147B**: 1337–44.

14. Neale BM, Medland S, Ripke S, Anney RJL, Asherson P, Buitelaar J, et al. Case control genome-wide association study of attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry* 2010; **49**: 906–20.

15. Mick E, Todorov A, Smalley S, Hu X, Loo S, Todd RD, et al. Family-based genomewide association scan of attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry* 2010; **2010**: 898-905.

16. Irish Schizophrenia Genomics Consortium and the Wellcome Trust Case Control Consortium 2. Genome-wide association study implicates HLA-C*01:02 as a risk factor at the major histocompatibility complex locus in schizophrenia. *Biol Psychiatry* 2012; **72**: 620-8.

17. Hawi Z, Matthews N, Barry E, Kirley A, Wagner J, Wallace RH, et al. A high density linkage disequilibrium mapping in 14 noradrenergic genes: evidence of association between SLC6A2, ADRA1B and ADHD. *Psychopharmacology (Berl)* 2013; **225**: 895-902.

18. Hawi Z, Matthews N, Wagner J, Wallace RH, Butler TJ, Vance A, et al. DNA variation in the SNAP25 gene confers risk to ADHD and is associated with reduced expression in prefrontal cortex. *Plos One* 2013; **8**: 4:e60274.

19. Angold A, Prendergast M, Cox A, Harrington R, Simonoff E, Rutter M. The Child and Adolescent Psychiatric Assessment (CAPA). *Psychol Med* 1995; **25**: 739-53.

20. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders.* 4th ed American Psychiatric Association: Washington DC, 1994.

21. World Health Organization. *The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines.* World Health Organization: Geneva, 1992.

22. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; **25**: 1754-60.

23. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011; **43**: 491-8.

24. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010; **20**: 1297-303.

25. Lemire M. Defining rare variants by their frequencies in controls may increase type I error. *Nat Genet* 2011; **43**: 391–2.

26. Pearson RD. Bias due to selection of rare variants using frequency in controls. *Nat Genet* 2011; **43**: 394-5.

27. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014; **46**: 310-5.

28. Liu DJ, Leal SM. A novel adaptive method for the analysis of next-generation sequencing data to detect complex trait associations with rare variants due to gene main effects and interactions. *PLOS Genet* 2010; **6**: e1001156.

29. MacArthur DG, Balasubramanian S, Frankish A, Huang N, Morris J, K. W, et al. A systematic survey of loss-of-function variants in human protein-coding genes. *Science* 2012; **335**: 823-8.

30. Madsen B, Browning S. A groupwise association test for rare mutations using a weighted sum statistic. *PLOS Genet* 2009; **5**: e1000384

31. The Encode Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012; **489**: 57-74.

32. Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, et al. The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol* 2010; **28**: 1045-8.

33. Kiezun A, Garimella K, Do R, Stitziel NO, Neale BM, McLaren PJ, et al. Exome sequencing and the genetic basis of complex traits. *Nat Genet* 2012; **44**: 623-30.

34. Notaras M, Hill R, van den Buuse M. The BDNF gene Val66Met polymorphism as a modifier of psychiatric disorder susceptibility: progress and controversy. *Mol Psychiatry* 2015; **20**: 916-30.

35. Notaras M, Hill R, van den Buuse M. A role for the BDNF gene Val66Met polymorphism in schizophrenia? A comprehensive review. *Neurosci Biobehav Rev* 2015; **51**: 15-30.

36. Chourbaji S, Hellweg R, Brandis D, Zorner B, Zacher C, Lang UE, et al. Mice with reduced brain-derived neurotrophic factor expression show decreased choline acetyltransferase activity, but

regular brain monoamine levels and unaltered emotional behavior. *Brain Res Mol Brain Res* 2004; **121**: 28-36.

37. Rios M, Fan G, Fekete C, Kelly J, Bates B, Kuehn R, et al. Conditional deletion of brain-derived neurotrophic factor in the postnatal brain leads to obesity and hyperactivity. *Mol Endocrinol* 2001; **15**: 1748–57.

38. Eaton MJ, Staley JK, Globus MY, Whittemore SR. Developmental regulation of early serotonergic neuronal differentiation: the role of brain-derived neurotrophic factor and membrane depolarization. *Dev Biol* 1995; **170**: 169-82.

39. Hyman C, Hofer M, Barde YA, Juhasz M, Yancopoulos GD, Squinto SP, et al. BDNF is a neurotrophic factor for dopaminergic neurons of the substantia nigra. *Nature* 1991; **350**: 230-2.

40. Lyons WE, Mamounas LA, Ricaurte GA, Coppola V, Reid SW, Bora SH, et al. Brain-derived neurotrophic factor-deficient mice develop aggressiveness and hyperphagia in conjunction with brain serotonergic abnormalities. *PNAS* 1999; **96**: 15239-44.

41. Andersen SL, Sonntag KC. Juvenile methylphenidate reduces prefrontal cortex plasticity via D3 receptor and BDNF in adulthood. *Front Synaptic Neurosci* 2014; **6**: 1-8.

42. Meredith GE, Callen S, Scheuer DA. Brain-derived neurotrophic factor expression is increased in the rat amygdala, piriform cortex and hypothalamus following repeated amphetamine administration. *Brain Res* 2002; **949**: 218–27.

43. Kent L, Green E, Hawi Z, Kirley A, Dudbridge F, Lowe N, et al. Association of the paternally transmitted copy of common Valine allele of the Val66Met polymorphism of the brain-derived neurotrophic factor (BDNF) gene with susceptibility to ADHD. *Mol Psychiatry* 2005; **10**: 939-43.

44. Pruunsild P, Kazantseva A, Aid T, Palm K, Timmusk T. Dissecting the human BDNF locus: Bidirectional transcription, complex splicing, and multiple promoters. *Genomics* 2007; **90**: 397–406.

Figure Legends

Figure 1. Plot of –log10 p values for each gene showing that the BDNF gene is significant at Bonferroni corrected significance levels. The red horizontal line represents the threshold for Bonferroni corrected significance, the grey horizontal line represents the threshold for nominal significance ($p<.05$).

Cook EH Jr, Stein MA, Krasowski MD, Cox NJ, Olkon DM, Kieffer JE et al. Association of attention-deficit disorder and the dopamine transporter gene. Am J Hum Genet. 1995 ;**56**:993-8.

Gizer I, Ficks C, Waldman I. Candidate gene studies of ADHD: a meta-analytic review. Hum Genet 2009;1**26**:51-90.

Bobb AJ, Addington AM, Sidransky E, Gornick MC, Lerch JP, Greenstein DK et al. Support for association between ADHD and two candidate genes: NET1 and DRD1. Am J Med Genet B Neuropsychiatr Genet. 2005;1**34B**:67-72.

Brookes K, Xu X, Chen W, Zhou K, Neale B, Lowe N et al. The analysis of 51 genes in DSM-IV combined type attention deficit hyperactivity disorder: association signals in DRD4, DAT1 and 16 other genes. Mol Psychiatry. 2006;**11**(10):934-53.

Comings DE, Wu S, Chiu C, Ring RH, Gade R, Ahn C et al Polygenic inheritance of Tourette syndrome, stuttering, attention deficit hyperactivity, conduct, and oppositional defiant disorder: the additive and subtractive effect of the three dopaminergic genes--DRD2, D beta H, and DAT1. Am J Med Genet. 1996 ;**67**:264-88.

Lasky-Su J, Neale BM, Franke B, Anney RJL, Zhou K, Maller JB et al. Genome-wide association scan of quantitative traits for attention deficit hyperactivity disorder identifies novel associations and confirms candidate gene associations. Am J Med Genet B Neuropsychiatr Genet 2008;**147B**:1345-54.

LaHoste GJ, Swanson JM, Wigal SB, Glabe C, Wigal T, King N et al. Dopamine D4 receptor gene polymorphism is associated with attention deficit hyperactivity disorder. Mol Psychiatry 1996;**1**:121-4.

Daly G, Hawi Z, Fitzgerald M, Gill M. Mapping susceptibility loci in attention deficit hyperactivity disorder: preferential transmission of parental alleles at DAT1, DBH and DRD5 to affected children. Mol Psychiatry 1999;**4**:192-6.

Lowe N, Kirley A, Hawi Z, Sham P, Wickham H, Kratochvil CH et al. Joint Analysis of the DRD5 Marker Concludes Association with Attention-Deficit/Hyperactivity Disorder Confined to the Predominantly Inattentive and Combined Subtypes. Am J Hum Genet. 2004; **74**: 348-356.

Ribasés M, Ramos-Quiroga JA, Hervás A, Bosch R, Bielsa A, Gastaminza X et al. Exploration of 19 serotoninergic candidate genes in adults and children with attention-deficit/hyperactivity disorder identifies association for 5HT2A, DDC and MAOB. Mol Psychiatry. 2009;**14**:71-85.

Brophy K, Hawi Z, Kirley A, Fitzgerald M, Gill M. Synaptosomal-associated protein 25 (SNAP-25) and attention deficit hyperactivity disorder (ADHD): evidence of linkage and association in the Irish population. Mol Psychiatry 2002;**7:**913-7.

Eisenberg J, Mei-Tal G, Steinberg A, Tartakovsky E, Zohar A, Gritsenko I et al Haplotype relative risk study of catechol-O-methyltransferase (COMT) and attention deficit hyperactivity disorder (ADHD): association of the high-enzyme activity Val allele with ADHD impulsive-hyperactive phenotype. Am J Med Genet. 1999 ;**88**:497-502.

Oades RD, Lasky-Su J, Christiansen H, Faraone SV, Sonuga-Barke EJ, Banaschewski T et al. The influence of serotonin- and other genes on impulsive behavioral aggression and cognitive impulsivity in children with attention-deficit/hyperactivity disorder (ADHD): Findings from a family-based association test (FBAT) analysis. Behav Brain Funct. 2008 **20**;4:48.

Manor I, Tyano S, Mel E, Eisenberg J, Bachner-Melman R, Kotler M et al. Family-based and association studies of monoamine oxidase A and attention deficit hyperactivity disorder (ADHD): preferential transmission of the long promoter-region repeat and its association with impaired performance on a continuous performance test. Mol Psychiatry 2002;**7:**626-32.

Hawi Z, Dring M, Kirley A, Foley D, Kent L, Craddock N et al. Serotonergic system and attention deficit hyperactivity disorder (ADHD): a potential susceptibility locus at the 5-HT(1B) receptor gene in 273 nuclear families from a multi-centre sample. Mol Psychiatry 2002;**7:**718-25.

Levitan RD, Masellis M, Basile VS, Lam RW, Jain U, Kaplan AS et al. Polymorphism of the serotonin-2A receptor gene (HTR2A) associated with childhood attention deficit hyperactivity disorder (ADHD) in adult women with seasonal affective disorder. J Affect Disord. 2002;**71**:229-33.

Sheehan K, Lowe N, Kirley A, Mullins C, Fitzgerald M, Gill M et al. Tryptophan hydroxylase 2 (TPH2) gene variants associated with ADHD. Mol Psychiatry. 2005;**10**:944-9.

Kent L, Green E, Hawi Z, Kirley A, Dudbridge F, Lowe N, Raybould R. Association of the paternally transmitted copy of common Valine allele of the Val66Met polymorphism of the brain-derived neurotrophic factor (BDNF) gene with susceptibility to ADHD. Mol Psychiatry. 2005 ;**10**:939-43.

Todd RD, Lobos EA, Sun LW, Neuman RJ. Mutational analysis of the nicotinic acetylcholine receptor alpha 4 subunit gene in attention deficit/hyperactivity disorder: evidence for association of an intronic polymorphism with attention problems. Todd RD, Lobos EA, Sun LW, Neuman RJ.Mol Psychiatry. 2003;**8**:103-8.

Baird AL, Coogan AN, Siddiqui A, Donev RM, Thome J. Adult attention-deficit hyperactivity disorder is associated with alterations in circadian rhythms at the behavioural, endocrine and molecular levels. Mol Psychiatry. 2012;**17**:988-95.

Neale BM, Medland SE, Ripke S, Asherson P, Franke B, Lesch K-P et al. Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. J Am Acad Child Adolesc Psychiatry 2010;**49**:884-97.

de Silva MG, Elliott K, Dahl HH, Fitzpatrick E, Wilcox S, Delatycki M et al. Disruption of a novel member of a sodium/hydrogen exchanger family and DOCK3 is associated with an ADHD like phenotype. J Med Genet. 2003; **40**:733-740.

Brookes KJ, Chen W, Xu X, Taylor E, Asherson P. Association of fatty acid desaturase genes with attention-deficit/hyperactivity disorder. Biol Psychiatry. 2006 ;**60**:1053-61.

Mick E, Todorov A, Smalley S, Hu X, Loo S, Todd RD et al. Family-based genome-wide association scan of attention-deficit/hyperactivity disorder. J Am Acad Child Adolesc Psychiatry. 2010 ;**49**:898-905.

Lasky-Su J, Anney RJ, Neale BM, Franke B, Zhou K, Maller JB et al. Genome-wide association scan of the time to onset of attention deficit hyperactivity disorder. Am J Med Genet B Neuropsychiatr Genet. 2008 ;**147B**:1355-8.

Bryant CD, Graham ME, Distler MG, Munoz MB, Li D, Vezina P et al. A role for casein kinase 1 epsilon in the locomotor stimulant response to methamphetamine. Psychopharmacology (Berl). 2009 ;**203**:703-11.

Neale BM, Lasky-Su J, Anney R, Franke B, Zhou K, Maller JB et al. Genome-wide association scan of attention deficit hyperactivity disorder. Am J Med Genet B Neuropsychiatr Genet. 2008 ;**147B**:1337-44.

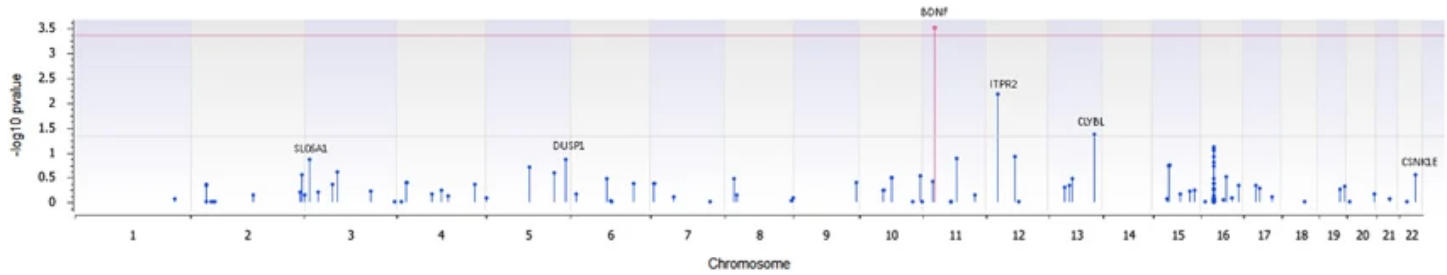
Neale BM, Medland S, Ripke S, Anney RJ, Asherson P, Buitelaar J et al. Case-control genome-wide association study of attention-deficit/hyperactivity disorder.J Am Acad Child Adolesc Psychiatry. 2010;**49**:906-20.

Williams NM, Zaharieva I, Martin A, Langley K, Mantripragada K, Fossdal R et al. Rare chromosomal deletions and duplications in attention-deficit hyperactivity disorder: a genome-wide analysis. Lancet. 2010 ;**376**:1401-8.

Table 1. Count of rare functional *BDNF* variants in cases vs controls

Position MAF	27695742 .00148	27695869 .00742	27695893 .00148	27720908 .00593	27721136 .00148	27722647 .00148	27741048 .002967	27741140 .00445	27743438 .00148	27743449 .00148	27743481 .00148	27743556 .00890	**Total**
Case (n=152)	1	3	0	4	1	1	2	2	1	1	0	5	**21**
Control (n=185)	0	2	1	0	0	0	0	1	0	0	1	1	**6**

Position: single nucleotide variant (SNV) positions on chromosome 11 (under build GRCh37.p13). MAF: minor allele frequency. Note: Each individual contributed a single rare variant with the exception of one ADHD case that had a BDNF rare variant at two positions (positions: 27695869, 27741140).


BDNF
ITPR2
CLYBL
SLC6A1
DUSP1
CSNK1E
-log10 pvalue
3.5
3
2.5
2
1.5
1
0.5
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
Chromosome

Supplementary Table 1: Discovery sample gene list (117 genes) annotated with source and inclusion criteria

Study design	Gene ID	Chr	Description	Associated variant α	Location	Reference
Candidate Gene	*SLC6A3*	5	solute carrier family 6 (dopamine transporter)	VNTR	3′ UTR	Cook et al(1995)[§], Gizer et al(2009)*
	DRD1	5	dopamine receptor D1	rs10061709	5′ UTR	Bobb et al(2005)[§], Brookes et al(2006b)*
	DRD2	11	dopamine receptor D2	rs7125415	intron 1	Comings et al(1996)[§], Lasky-Su et al(2008a)†
	DRD4	11	dopamine receptor D4	VNTR	exon 3	La Hoste et al(1996) [§], Gizer et al(2009)*
	DRD5	4	dopamine receptor D5	Dinucleotide repeat	5′ flanking	Daly et al(1999)[§], Lowe et al(2004)*
	DDC	7	dopa decarboxylase (aromatic L-amino acid decarboxylase)	rs1466163	intron 2	Brookes et al(2006b)[§], M Ribasés et al(2009)
	SNAP25	20	synaptosomal-associated protein, 25kDa	rs3746544	3′ UTR	Brophy et al(2003)[§], Gizer et al(2009)*
	SLC6A2	16	solute carrier family 6 (noradrenalin transporter)	rs17307096	intron 1	Bobb et al(2005)[§], Lasky-Su et al(2008a)†
	ADRA1A	8	adrenergic, alpha-1A- receptor	rs1470221	3′ UTR	Brookes et al(2006b)[§], Lasky-Su et al(2008a)†
	ADRB2	5	adrenergic, beta-2-, receptor	rs12654778	5′ flanking	Brookes et al (2006b)[§], Lasky-Su et al(2008a)†
	DBH	9	dopamine beta-hydroxylase	rs4531	exon 5	Daly et al(1999)[§], Lasky-Su et al(2008a)†
	COMT	22	catechol-O-methyltransferase	rs9332377	intron 5	Eisenberg et al(1999)[§], Lasky-Su et al(2008a)†
	PNMT	17	phenylethanolamine N-methyltransferase	rs2934966	5′ flanking	Brookes et al(2006b)[§], Oades et al(2008)
	SLC6A4	17	solute carrier family 6 (serotonin transporter)	indel	promoter	Manor et al(2002)[§], Gizer et al(2009)*
	HTR1B	6	5-hydroxytryptamine (serotonin) receptor 1B	rs6296	exon 1	Hawi et al(2002)[§], Gizer et al (2009) *
	HTR2A	13	5-hydroxytryptamine (serotonin) receptor 2A	rs1002513	intron 3	Levitan et al(2002)[§], Lasky-Su et al(2008a) †
	HTR1E	6	5-hydroxytryptamine (serotonin) receptor 1E	rs17222848	3′ flanking	Brookes et al(2006b)[§], Lasky-Su et al(2008)†
	TPH2	12	tryptophan hydroxylase 2	rs1843809	Intron 5	Sheehan et al(2005)[§], Brookes et al(2006b)£
	BDNF	11	brain-derived neurotrophic factor	rs2049048	5′ flanking	Kent et al(2005)[§], Lasky-Su et

						al(2008a)†
	CHRNA4	20	cholinergic receptor, nicotinic, alpha 4	rs7267923	5′ flanking	Todd et bal(2003)[§], Brookes et al(2006b)£
	HES1	3	hairy and enhancer of split 1, (Drosophila)	rs6767127	3′ flanking	Brookes et al(2006b)[§], Lasky-Su et al(2008a)†
	PER2[+]	2	period homolog 2 (Drosophila)	rs13382977	intron 11	Brookes et al(2006b)[§], Baird et al(2012)
	SLC6A1	3	solute carrier family 6 (GABA transporter)	rs6778281	intron 1	Lasky-Su et al(2008a)[§], Neale et (2010a)†
	SLC9A9	3	solute carrier family 9 (sodium/hydrogen exchanger)	inversion breakpoint	3p14 - q21	de Silva et al(1998)[§], Mick et 2010)†
	FADS2	11	fatty acid desaturase 2	rs174611	Intron 7	Brookes et al(2006a)[§], Brookes et (2006b)
	ARRB1	11	arrestin, beta 1	rs512797	Intron 3	Lasky-Su et al(2008b)[§], Mick et (2010)†
	FADS1	11	fatty acid desaturase 1	rs174556	Intron 3	Brookes et al[§] (2006a)[§], Mick et (2010)†
	NFIL3	9	nuclear factor, interleukin 3 regulated	rs968821	5′ flanking	Lasky –Su et al (2008b)[§], Mick et (2010) †
	CSNK1E[+]	22	casein kinase 1, epsilon	rs17753394	3′ flanking	Mick et al(2010)§, Bryant et al(2009)
TDT-GWAS	*NUCB1*	19	nucleobindin 1	rs9676447	intron 9	Neale et al(2008)
	RAB27B	18	RAB27B, member RAS oncogene family	rs2311120	5′ flanking	Neale et al(2008)
	CNR1	6	cannabinoid receptor 1 (brain)	rs964647	3′ flanking	Neale et al(2008)
	KCNIP4	4	Kv channel interacting protein 4	rs876477	intron 2	Neale et al(2008)
	DCLK1	13	doublecortin-like kinase 1	rs1539549	intron 4	Neale et al(2008)
	SPOCK3	4	sparc/osteonectin, cwcv and kazal-like domains proteoglycan (testican) 3	rs7657608	intron 1	Neale et al(2008)
	PTPN14	1	protein tyrosine phosphatase, non-receptor type 14	rs6657749	intron 5	Neale et al(2008)
	PEX7	6	peroxisomal biogenesis factor 7	rs6919857	5′ flanking	Neale et al(2008)
	ITPR2	12	inositol 1,4,5-triphosphate receptor, type 2	rs3782309	intron 9	Neale et al(2008)
	TRUB1	10	TruB pseudouridine (psi) synthase homolog 1	rs12772737	3′ flanking	Neale et al(2008)
	RORA	15	RAR-related orphan receptor A	rs922781	intron 2	Neale et al(2008)
	DUSP1	5	dual specificity phosphatase 1	rs7702178	intron 2	Mick E et al(2010)

	DPH6/ATPBD4	15	ATP binding domain 4	rs4923705	5′ flanking[#]	Mick E et al(2010)
	LECT1	13	leukocyte cell derived chemotaxin 1	rs6561686	intron 4	Mick E et al(2010)
	BMPR1B	4	bone morphogenetic protein receptor, type IB	rs1859156	intron 2	Mick E et al(2010)
	RBMS3	3	RNA binding motif, single stranded interacting protein 3	rs17023218	intron 1	Mick et al(2010)
	CEP112/CCDC46	17	coiled-coil domain containing 46	rs8074751	intron 3	Mick et al(2010)
	ELOVL6	4	ELOVL family member 6, elongation of long chain fatty acids	rs10011926	intron 2	Mick et al(2010)
	ANO5/TMEM16E	11	anoctamin 5	rs10833716	5′ flanking	Mick et al(2010)
	MCTP1	5	multiple C2 domains, transmembrane 1	rs11953346	intron 20	Mick et al(2010)
	EMP2	16	epithelial membrane protein 2	rs11074889	intron 4	Mick et al(2010)
	UGT1A9	2	UDP glucuronosyltransferase 1 family, polypeptide A9	rs2602381	intron 1	Mick et al(2010)
Case-Control GWAS	*PRKG1*	10	protein kinase, cGMP-dependent, type I	rs10823954	intron 7	Neale et al(2010b)
	FLNC	7	filamin C, gamma	rs13227216	intron 32	Neale et al(2010b)
	TCERG1L	10	transcription elongation regulator 1-like	rs1583670	intron 8	Neale et al(2010b)
	PPM1H	12	protein phosphatase, Mg2+/Mn2+ dependent, 1H	rs12317552	3′ UTR	Neale et al(2010b)
	NXPH1	7	neurexophilin 1	rs17151821	intron 1	Neale et al(2010b)
	HK1	10	hexokinase 1	rs2394538	intron 1	Neale et al(2010b)
	NRXN1	2	neurexin 1	rs17495366	intron 23	Neale et al(2010b)
	HKDC1	10	hexokinase domain containing 1	rs906219	exon 18	Neale et al(2010b)
QTL-GWAS	*CDH13*	16	cadherin 13, H-cadherin (heart)	rs6565113	intron 3	Lasky-Su et al(2008a)
	GFOD1	6	glucose-fructose oxidoreductase domain containing 1	rs552655	intron 5	Lasky-Su et al(2008a)
	ZNF423	16	zinc finger protein 423	rs17281813	intron 7	Lasky-Su et al(2008a)
	HAS3	16	hyaluronan synthase 3	rs8047014	5′ flanking	Lasky-Su et al(2008a)
	FOXP1	3	forkhead box P1	rs17651978	intron/3′UTR	Lasky-Su et al(2008a)
	EREG	4	epiregulin	rs1350666	5′ flanking	Lasky-Su et al(2008a)
	CLYBL	13	citrate lyase beta like	rs7992643	intron 1	Lasky-Su et al(2008a)
	MEIS2	15	Meis homeobox 2	rs8041675	intron 6	Lasky-Su et al(2008a)
	NAPRT1	8	nicotinate phosphoribosyltransferase domain containing 1	rs2290416	exon 10	Lasky-Su et al(2008a)

	OXER1	2	oxoeicosanoid (OXE) receptor 1	rs930421	3′ UTR	Lasky-Su et al(2008a)
	GRIK1	21	glutamate receptor, ionotropic, kainate 1	rs363512	intron 3	Lasky-Su et al(2008a)
	IL16	15	interleukin 16 (lymphocyte chemoattractant factor)	rs7172689	intron 3	Lasky-Su et al(2008a)
	DMRT2	9	doublesex and mab-3 related transcription factor 2	rs17641078	Missense UTR	Lasky-Su et al(2008a)
	LPL	8	lipoprotein lipase	rs7816032	5′ flanking	Lasky-Su et al(2008a)
	SLCO3A1	15	solute carrier organic anion transporter family, member 3A1	rs7495052	intron 2	Lasky-Su et al(2008a)
	NCKAP5/FLJ34870	2	NCK-associated protein 5	rs7577925	intron 5	Lasky-Su et al(2008a)
	FHIT	3	fragile histidine triad gene	rs6791644	intron 4	Lasky-Su et al(2008a)
	ZNF544	19	zinc finger protein 544	rs260461	intron 6	Lasky-Su et al(2008a)
CNV-GWAS	*TTC27*	2	tetratricopeptide repeat domain 27	CNV		Williams et al(2010)
	BIRC6	2	baculoviral IAP repeat-containing 6	CNV		Williams et al(2010)
	LTBP1	2	Latent transforming growth factor beta binding protein 1	CNV		Williams et al(2010)
	CNTN6	3	contactin 6	CNV		Williams et al(2010)
	OTUD7A	15	OTU domain containing 7A	CNV		Williams et al(2010)
	BOLA2	16	bolA family member 2	CNV		Williams et al(2010)
	BOLA2B	16	bolA family member 2B	CNV		Williams et al(2010)
	MVP	16	major vault protein	CNV		Williams et al(2010)
	YPEL3	16	yippee-like 3 (Drosophila)	CNV		Williams et al(2010)
	CORO1A	16	coronin, actin binding protein, 1A	CNV		Williams et al(2010)
	MAPK3	16	mitogen-activated protein kinase 3	CNV		Williams et al(2010)
	GDPD3	16	glycerophosphodiester phosphodiesterase domain containing 3	CNV		Williams et al(2010)
	QPRT	16	quinolinate phosphoribosyltransferase	CNV		Williams et al(2010)
	MAZ	16	MYC-associated zinc finger protein (purine-binding transcription factor)	CNV		Williams et al(2010)
	CDIPT	16	CDP-diacylglycerol--inositol 3-phosphatidyltransferase	CNV		Williams et al(2010)
	GIYD1/SLX1A	16	SLX1 structure-specific endonuclease subunit homolog A (S. cerevisiae)	CNV		Williams et al(2010)

TBX6	16	T-box 6	CNV	Williams et al(2010)
PPP4C	16	protein phosphatase 4, catalytic subunit	CNV	Williams et al(2010)
ALDOA	16	aldolase A, fructose-bisphosphate	CNV	Williams et al(2010)
FAM57B	16	family with sequence similarity 57, member B	CNV	Williams et al(2010)
DOC2A	16	double C2-like domains, alpha	CNV	Williams et al(2010)
HIRIP3	16	HIRA interacting protein 3	CNV	Williams et al(2010)
TAOK2	16	TAO kinase 2	CNV	Williams et al(2010)
TMEM219	16	transmembrane protein 219	CNV	Williams et al(2010)
PRRT2	16	proline-rich transmembrane protein 2	CNV	Williams et al(2010)
CD2BP2	16	CD2 (cytoplasmic tail) binding protein 2	CNV	Williams et al(2010)
TBC1D10B	16	TBC1 domain family, member 10B	CNV	Williams et al(2010)
INO80E/CCDC95	16	INO80 complex subunit E	CNV	Williams et al(2010)
SEZ6L2	16	seizure related 6 homolog (mouse)-like 2	CNV	Williams et al(2010)
ASPHD1	16	aspartate beta-hydroxylase domain containing 1	CNV	Williams et al(2010)
KCTD13	16	potassium channel tetramerisation domain containing 13	CNV	Williams et al(2010)
CHRNA7	15	cholinergic receptor, nicotinic, alpha 7	CNV	Williams et al(2010)
SEPHS2	16	selenophosphate synthetase 2	CNV	Williams et al(2010)
ZNF771	16	zinc finger protein 771	CNV	Williams et al(2010)
ZNF48/ZNF553	16	zinc finger protein 48	CNV	Williams et al(2010)
MYLPF	16	myosin light chain, phosphorylatable, fast skeletal muscle	CNV	Williams et al(2010)
GIYD2/SLX1B	16	SLX1 structure-specific endonuclease subunit homolog B (S. cerevisiae)	CNV	Williams et al(2010)
SPN	16	sialophorin	CNV	Williams et al(2010)
SULT1A3	16	sulfotransferase family, cytosolic, 1A, phenol-preferring, member 3	CNV	Williams et al(2010)
SULT1A4	16	sulfotransferase family, cytosolic, 1A, phenol-preferring, member 4	CNV	Williams et al(2010)

α = Associated SNP taken from the study of largest sample size, Chr = Chromosome, § = first reported by, * = meta-analysis article, †= GWAS finding, £ = association in large sample # = 3′ flanking for antisense RNA 1, + Two additional candidate genes that were included because an association with ADHD at $p<.05$ was supported by very strong functional evidence - Selective pharmacological inhibition of *CSNK1E* using PF-670462 blocks the locomotor response to methamphetamine in mice and rats, healthy *PER2* circadian rhythmicity expression is lost in individuals with ADHD.

Supplementary Table 2. *in silico* functional predictions for the identified rare functional BDNF variants.

Position/ ID	Ref	Alt	Consequence	CADD C-score*	Predicted functional annotations	Potential consequences
27695742	C	G	NON_SYNON	13.39	This missense variant is predicted to be benign by both PolyPhen2 and SIFT.	Low likelihood of resulting in altered protein folding/ function
27695869 (rs147943376)	T	C	REGULATORY	13.70	This variant overlaps with strong enhancer regions of HMEC (breast) and NHEK (skin) cell lines (part of the ENCODE project) and DNase I hypersensitivity clusters.	Possible role in cis-regulatory transcriptional control
27695893†	T	A	SPLICE_SITE	19.04	This variant lies 4 bases from the acceptor 3' splicing element. The consensus sequences in mammals at the 3' splice site are comprised of a pyrimidine-rich tract followed by an AG acceptor site. In the case of this SNV, the rare A allele decreases the pyrimidine content upstream to the AG acceptor site which is likely to influence splicing efficiency. Although this hypothesis remained to be tested, there is transcriptional evidence of an alternatively spliced transcript containing an extra exon of 75 amino acids (RefSeq: NM_001143810.1).	**Likely to result in aberrant transcript processing of BDNF**
27720908† (rs20124401)	T	C	REGULATORY	10.68	Using the splicing module in Alamut, this intronic variant was not predicted to alter either;	Low likelihood of resulting in aberrant transcript processing

					normal splicing or, de novo acceptor/ donor splice sites.	
27721136[†]	C	G	REGULATORY	12.54	Using the splicing module in Alamut, this intronic variant was not predicted to alter either; normal splicing or, de novo acceptor/ donor splice sites.	Low likelihood of resulting in aberrant transcript processing
27722647[§]	C	G	REGULATORY	22.3	This variant overlaps with an alternative promoter, multiple TFs (RBBP5, CTBP2, EZH2, STAT3 and GTF2F1)) ChIP-seq regions and DNase I hypersensitivity clusters. Likely to regulate expression of BDNF.	**Likely to result in changes in cis-regulatory transcriptional control of BDNF**
27741048[§]	T	A	REGULATORY	22.4	This variant overlaps a CpG island, multiple TFs (EZH2, REST and SIN3AK20) ChIP-seq regions, and DNase I hypersensitivity clusters. Likely to regulate expression of BDNF.	**Likely to result in changes in cis-regulatory transcriptional control of BDNF**
27741140 (rs183873737)	C	T	REGULATORY	22.2	This variant overlaps a CpG island, transcription factor EZH2 ChIP-seq regions, and active promoter region of H1-hESC cell lines of the ENCODE project.	**Likely to result in changes in cis-regulatory transcriptional control of BDNF**
27743438[§]	T	A	REGULATORY	22.2	This variant overlaps TFs (EZH2 and SUZ12) ChIP-seq regions and DNase I hypersensitivity clusters. Likely to regulate expression of BDNF.	**Likely to result in changes in cis-regulatory transcriptional control of BDNF**
27743449[§]	C	G	REGULATORY	21.7	This variant overlaps TF EZH2 ChIP-seq regions and histone marks (H3K27Me3) were	**Likely to result in changes in cis-regulatory transcriptional control of BDNF**

					detected in the anterior caudate region (NIH Roadmap Epigenomics Project cell types). Likely to regulate expression of BDNF.	
27743481[§]	C	G	REGULATORY	21.1	This variant overlaps TF EZH2 ChIP-seq regions and histone marks (H3K27Me3) were detected in the anterior caudate region (NIH Roadmap Epigenomics Project cell types). Likely to regulate expression of BDNF.	**Likely to result in changes in cis-regulatory transcriptional control of BDNF**
27743556 (rs28383487)	G	T	REGULATORY	21.8	This variant overlaps a CpG island, histone marks (H3K27me3) in the anterior caudate region and transcription factor RBBP5 ChIP-seq regions.	**Likely to result in changes in cis-regulatory transcriptional control of BDNF**

Position: genomic position under build GRCh37.p13; ID: annotated SNP (where available); Ref: reference allele; Alt: alternative allele; Non-synon: Non-synonymous. * All BDNF variants were ranked by the Combined Annotation-Dependent Depletion (CADD) tool based on their likely pathogenicity. A C-score above 10 indicates a high likelihood that the variant will be linked to disease. †To assess variants located at exon/intron boundaries, the Alamut tool was used to predict alternative RNA splicing. § Cis-regulatory transcriptional control of variants in proximity to the transcription start site was assessed using available data of the Encyclopedia of DNA Elements (ENCODE) consortium and the US National Institutes of Health (NIH) Roadmap Epigenomics Project.

Supplementary Table 3. Kernel Based Adaptive Cluster (KBAC) results for each gene

Gene	Chr.	Start Pos.	Stop Pos.	Vars.	Multi.	Case(n)	Control(n)	Perms.	KBAC	*P*-value
PTPN14	1	214521039	214726024	16	16	8	17	52	0.246	0.846
BIRC6	2	32580735	32844966	29	27	21	21	86	4.045	0.442
TTC27	2	32852087	33047118	11	12	5	14	52	-0.963	0.962
LTBP1	2	33171221	33625576	18	18	14	13	86	2.695	0.453
OXER1	2	42988639	42992401	1	2	1	4	52	-0.317	1.000
NRXN1	2	50144643	51260674	29	28	13	32	52	-0.713	0.981
NCKAP5	2	133428361	134400118	22	20	11	23	52	1.785	0.731
UGT1A9	2	234579544	234682951	7	8	4	8	69	0.631	0.623
PER2	2	239151679	239199678	11	10	5	7	137	1.731	0.285
CNTN6	3	1133342	1446278	6	6	3	5	69	0.394	0.710
SLC6A1	3	11033420	11081935	7	8	10	7	307	2.832	0.137
RBMS3	3	29321351	30052886	19	18	15	18	69	2.365	0.623
FHIT	3	59734036	61238133	3	4	3	2	103	0.705	0.447
FOXP1	3	71002865	71634140	11	13	8	5	188	2.460	0.250
SLC9A9	3	142983063	143568373	6	4	4	5	69	0.317	0.609
HES1	3	193852931	193857401	9	9	2	11	52	-1.210	0.981
DRD5	4	9782258	9786633	3	4	0	5	52	-0.717	1.000
KCNIP4	4	20729234	21951424	12	13	10	12	103	1.850	0.408
EREG	4	75229860	75255477	3	4	1	2	69	0.054	0.696
BMPR1B	4	95678128	96080601	11	12	6	7	69	1.114	0.580
ELOVL6	4	110969187	111121355	4	5	2	3	69	0.132	0.739
SPOCK3	4	167653535	168156741	4	4	2	1	103	0.850	0.437
SLC6A3	5	1391905	1446543	8	8	3	6	52	0.162	0.808
MCTP1	5	94037280	94621279	10	10	8	6	205	2.448	0.195
ADRB2	5	148205156	148255628	4	5	4	5	171	1.702	0.257
DUSP1	5	172194093	172199203	5	8	3	4	307	2.527	0.140
GFOD1	6	13362587	13488869	2	3	1	1	69	0.301	0.696
HTR1B	6	78170948	78174120	2	2	1	0	120	0.549	0.333
HTR1E	6	87646024	87727397	7	8	3	8	52	-0.260	0.923
CNR1	6	88848583	88877063	1	2	0	1	52	-0.248	1.000

PEX7	6	137142702	137236072	3	4	4	3	103	0.776	0.417
NXPH1	7	8472585	8793593	4	5	3	2	103	0.928	0.417
DDC	7	50525134	50634154	4	3	2	4	52	0.232	0.788
FLNC	7	128469436	128500328	31	29	18	30	52	0.680	0.962
LPL	8	19795582	19825770	7	8	6	5	120	1.531	0.342
ADRA1A	8	26604667	26728003	10	9	5	7	52	0.536	0.731
NAPRT1	8	144655955	144661881	8	9	6	12	52	-0.383	0.942
DMRT2	9	1048858	1058554	2	3	1	1	52	0.301	0.808
DBH	9	136500485	136525466	11	10	6	5	103	1.609	0.408
PRKG1	10	52749911	54059110	8	9	6	7	69	1.309	0.565
HKDC1	10	70979059	71028315	11	12	8	11	137	2.311	0.328
HK1	10	71028740	71162638	17	15	11	17	120	2.811	0.325
TRUB1	10	116696952	116738439	4	5	0	4	52	-0.990	1.000
TCERG1L	10	132889655	133110984	2	3	3	1	137	1.137	0.292
DRD4	11	636305	641706	5	6	2	7	52	-0.865	0.981
ANO5	11	22213199	22305913	6	7	4	5	103	0.926	0.388
BDNF	11	27675440	27744605	13	14	21	6	100001	8.594	**0.0003****
FADS1	11	61566097	61585529	1	2	0	1	52	-0.248	1.000
FADS2	11	61582675	61635826	4	3	0	2	52	-0.495	1.000
ARRB1	11	74970166	75063875	6	6	4	2	324	1.948	0.133
DRD2	11	113279317	113347413	7	6	2	4	52	-0.116	0.731
ITPR2	12	26487285	26987131	25	25	21	12	100001	8.318	**0.007***
PPM1H	12	63036762	63329665	3	4	3	1	358	1.399	0.123
TPH2	12	72331626	72427221	6	7	1	8	52	-1.135	1.000
DCLK1	13	36341789	36706514	15	16	15	15	86	2.471	0.500
HTR2A	13	47404677	47472211	3	4	2	3	86	0.099	0.465
LECT1	13	53276399	53314947	7	6	3	2	120	1.152	0.342
CLYBL	13	100257919	100550388	11	9	10	5	1395	3.709	**0.044***
OTUD7A	15	31772884	32163875	6	7	2	4	52	0.108	0.865
DPH6	15	35656670	35839597	10	8	5	4	222	1.978	0.189
MEIS2	15	37182222	37394500	9	10	12	9	222	3.289	0.185
RORA	15	60779483	61522502	17	13	8	13	69	0.944	0.696
IL16	15	81473941	81606104	6	8	3	7	69	0.568	0.594

SLCO3A1	15	92395938	92716665	12	13	7	10	69	1.582	0.580
EMP2	16	10621279	10675539	3	3	0	2	52	-0.495	1.000
SPN	16	29673271	29682828	1	2	1	4	52	-0.317	1.000
QPRT	16	29689334	29710324	1	2	0	2	52	-0.271	1.000
MAZ	16	29816417	29823504	14	14	4	11	52	-0.080	0.827
PRRT2	16	29822409	29828202	4	4	6	4	256	1.825	0.160
MVP	16	29830715	29860360	10	11	3	13	52	-2.170	1.000
CDIPT	16	29868678	29875801	1	2	1	0	103	0.549	0.427
SEZ6L2	16	29881480	29911585	24	25	15	15	69	2.841	0.551
ASPHD1	16	29911147	29918377	3	4	3	0	494	1.647	0.093
KCTD13	16	29916657	29938545	3	4	3	0	664	1.647	0.077
TMEM219	16	29972351	29985373	2	3	2	0	239	1.098	0.188
TAOK2	16	29984188	30004583	18	15	10	10	120	2.758	0.333
HIRIP3	16	30002642	30008417	9	8	10	5	375	2.600	0.120
INO80E	16	30006485	30018112	7	6	9	4	562	2.299	0.085
DOC2A	16	30015835	30024424	2	3	1	3	52	-0.450	0.904
FAM57B	16	30034748	30065338	3	4	0	3	52	-0.743	1.000
ALDOA	16	30063411	30082741	7	7	4	7	52	-0.102	0.808
PPP4C	16	30086297	30097698	1	2	0	1	52	-0.248	1.000
TBX6	16	30096114	30104258	9	10	6	13	52	-0.986	0.923
YPEL3	16	30102635	30109248	6	6	8	7	103	1.176	0.417
GDPD3	16	30115131	30125878	6	7	2	6	52	0.133	0.827
MAPK3	16	30124426	30135630	6	6	2	4	52	-0.116	0.885
CORO1A	16	30193731	30201397	6	8	6	6	69	0.825	0.565
CD2BP2	16	30361087	30367682	6	6	1	5	52	-0.465	0.865
TBC1D10B	16	30367422	30382522	4	4	2	6	52	0.028	0.731
MYLPF	16	30385123	30390310	1	2	0	1	52	-0.248	1.000
ZNF48	16	30388633	30412429	5	6	3	4	86	0.400	0.523
ZNF771	16	30417735	30430916	1	2	0	1	52	-0.248	1.000
SEPHS2	16	30453952	30458224	4	5	1	3	52	-0.194	0.962
ZNF423	16	49523515	49892830	25	23	16	24	52	1.186	0.904
SLC6A2	16	55688542	55741104	9	10	5	5	137	1.731	0.307
HAS3	16	69138467	69153622	8	8	3	5	52	0.186	0.808

CDH13	16	82659399	83831215	20	18	13	13	86	2.804	0.453
SLC6A4	17	28520337	28563986	3	4	2	1	86	0.850	0.453
PNMT	17	37823706	37827728	7	9	8	10	86	1.605	0.523
CEP112	17	63630658	64189212	8	9	4	11	52	0.405	0.769
RAB27B	18	52384097	52563747	2	3	0	2	52	-0.495	1.000
NUCB1	19	49402307	49427540	8	8	4	5	69	0.702	0.551
ZNF544	19	58738960	58789923	3	4	5	5	86	0.559	0.477
SNAP25	20	10198477	10289068	5	6	1	9	52	-1.107	0.981
CHRNA4	20	61973662	62010487	7	6	4	6	69	0.293	0.681
GRIK1	21	30908254	31313282	13	12	6	13	52	-0.094	0.846
COMT	22	19928263	19958498	1	2	0	1	52	-0.248	1.000
CSNK1E	22	38685697	38715089	2	3	3	4	137	0.781	0.285

9 genes (DRD1, NFIL3, CHRNA7, GIYD1/SLX1A, BOLA2B, GIYD2/SLX1B, BOLA2, SULT1A3, SULT1A4) did not show any rare variants across the samples and hence are not included in the table. Chr: chromosome; Start pos, stop pos: genomic positions of gene start and stop (build GRCh37.p13); Vars: number of unique variants per gene; Multi: number of unique multi-marker genotypes per gene; Case(n): count of rare variants per gene in cases; Control(n): count of rare variants per gene in controls; perms: number of iterations of the adaptive permutation procedure; KBAC: kernel based adaptive cluster statistic. * significant at nominal level (uncorrected, α=.05), ** significant at Bonferroni corrected level.

Supplementary Table 4. Count of rare loss of function variants in cases vs. controls

Gene	Chr.	Position	Variant type	Count cases	Count controls	Ref. allele	Alt. Allele	P value
BIRC6	2	32733048	Splice Site	0	1	C	T	1.00
RBMS3	3	29977687	Splice Site	0	2	G	A	0.50
DRD5	4	9784658	Stop Gain	0	1	C	A	1.00
PEX7	6	137219351	Stop Gain	1	1	T	A	1.00
DBH	9	136508537	Splice Site	2	1	C	T	0.59
BDNF	11	27695731	Frame Trunc.	0	1	CAT	C	1.00
SEZ6L2	16	29883600	Stop Gain	1	0	G	A	0.45
ZNF423	16	49559275	Splice Site	1	0	T	G	0.45
ZNF544	19	58773014	Stop Gain	3	3	C	T	1.00
GRIK1	21	31045482	Splice Site	0	1	G	A	1.00

Chr: Chromosome; Ref. allele: reference allele; Alt. allele: alternative allele; Frame Trunc: Frameshift truncation.