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 ($p<.05$).