Ultra-rare *de novo* damaging coding variants are enriched in attention-deficit/hyperactivity disorder and identify risk genes

Authors: Emily Olfson[1,2*], Luis C. Farhat[1,3], Wenzhong Liu[1], Lawrence A. Vitulano[1], Gwyneth Zai[4,5], Monicke O. Lima[3], Justin Parent[6,7,8], Guilherme V. Polanczyk[3], Carolina Cappi[9], James L. Kennedy[4,5], Thomas V. Fernandez[1,10*]

[1] Child Study Center, Yale University

[2] Wu Tsai Institute, Yale University

[3] Faculdade de Medicina FMUSP, Universidade de São Paulo

[4] Neurogenetics Section, Molecular Brain Sciences Department, Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health

[5] Institute of Medical Science and Department of Psychiatry, University of Toronto

[6] Center for Children and Families, Florida International University

[7] Bradley/Hasbro Children's Research Center, E.P. Bradley Hospital

[8] Alpert Medical School of Brown University

[9] Icahn School of Medicine at Mount Sinai

[10] Department of Psychiatry, Yale University

*Corresponding authors: Emily Olfson and Thomas Fernandez

Email: emily.olfson@yale.edu, thomas.fernandez@yale.edu

Major classification: biological sciences

Minor classification: genetics

Keywords: attention-deficit/hyperactivity disorder; DNA sequencing; parent-child trios

Abstract:

Attention-deficit/hyperactivity disorder (ADHD) is a common and impairing neurodevelopmental disorder in which genetic factors play an important role. DNA sequencing of parent-child trios provides a powerful approach for identifying *de novo* (spontaneous) variants, which has led to the discovery of hundreds of clinically informative risk genes for other neurodevelopmental disorders but has yet to be extensively leveraged in studying ADHD. Here, we conducted whole-exome DNA sequencing in 152 parent-child trios with ADHD and demonstrate for the first time a significant enrichment of rare and ultra-rare *de novo* protein-truncating variants and missense variants predicted to be damaging in ADHD cases compared to unaffected controls. Combining these results with a large independent case-control DNA sequencing cohort (3,206 ADHD cases and 5,002 controls), we identify *lysine demethylase 5B* (*KDM5B)* as a high-confidence risk gene for ADHD as well as two likely risk genes. We estimate that 862 genes contribute to ADHD risk. Finally, using our list of genes harboring ultra-rare *de novo* damaging variants, we show that these genes overlap with previously reported risk genes for other neuropsychiatric conditions in both DNA sequencing and genome-wide association studies. We also show that these genes are enriched for several canonical biological pathways, suggesting early neurodevelopmental underpinnings of ADHD. Overall, this work provides critical new insight into the biology of ADHD and demonstrates the discovery potential of DNA sequencing in larger parent-child trio cohorts.

Significance statement:

Given the important role of genetic factors in the development of attention-deficit/hyperactivity disorder (ADHD), research aimed at identifying risk genes can provide critical insight into underlying biological processes. We conducted whole-exome DNA sequencing in parent-child trios with ADHD, showing that these children have a significantly greater rate of rare and ultra-rare *de novo* gene-damaging mutations compared to unaffected controls, expanding our understanding of the genetic landscape of ADHD. We then use this information to identify *KDM5B* as a high-confidence risk gene for ADHD and highlight several enriched biological pathways. This work advances our etiologic understanding of ADHD and illustrates a previously unexplored path for risk gene discovery in this common neurodevelopmental disorder.

Introduction:

Attention-deficit/hyperactivity disorder (ADHD) is a common neurodevelopmental disorder in childhood (1) that places a significant burden on individuals, their families, and the community (2). ADHD is highly heritable (~70-80%) (3), so identifying genes associated with the disorder will increase our understanding of underlying biological processes. Recent case-control genome-wide studies have identified ADHD risk loci by assessing common single-nucleotide polymorphisms (SNPs) through genome-wide association studies (GWAS) (4, 5). However, to date, SNP-heritability has only accounted for a small portion (~15-30%) of the overall heritability estimates from twin studies, suggesting that other genetic factors, including rare genetic variants, may play an important role in ADHD risk (6). Indeed, previous studies have demonstrated that rare copy number variants (7) and very rare protein-truncating variants in evolutionarily constrained genes (8) are enriched in ADHD. Therefore, assessing rare variation may help identify potential ADHD risk genes. Despite previous research considering these different categories of genetic variation in ADHD, few specific high-confidence risk genes have yet been identified.

Studies of rare *de novo* genetic variants using parent-child trios have proven to be powerful for risk gene discovery in other neurodevelopmental disorders such as autism spectrum disorder (ASD) (9), developmental delay/intellectual disability (10), and Tourette's disorder (11), leading to the discovery of risk genes. Since the background rate of *de novo* variants in the population is low, finding an elevated rate of damaging *de novo* variants suggests that we can leverage these variants to identify risk genes and underlying biological pathways. However, this approach has yet to be extensively leveraged for ADHD.

Only a few studies have used parent-child trios to investigate the genetics of ADHD. Studies examining *de novo* copy number variants (CNVs) suggest a greater rate of these variants in ADHD cases compared to rates in controls (12, 13), with the largest study suggesting a rate similar to that previously reported in ASD and Tourette's disorder (13). However, given the large number of genes disrupted by CNVs, it is challenging to identify specific risk genes from these variants. Whole-exome DNA sequencing studies enable the identification of *de novo* sequence variants affecting single genes. There have been a few small whole-exome DNA sequencing studies of parent-child trios focused on ADHD (14-16), each identifying rare damaging *de novo* sequence variants in 11-30 parent-child trios, supporting the discovery potential of applying this approach in larger ADHD cohorts.

Here, we conducted whole-exome DNA sequencing in 152 parent-child trios (456 individuals in total), comprising a child with ADHD and both biological parents, to identify rare and ultra-rare *de novo* genetic variants and compare rates with previously sequenced controls without ADHD. For the first time, we demonstrate that rare and ultra-rare *de novo* protein-truncating variants (PTVs, including single nucleotide variants introducing premature stop codons, frameshift indels, and canonical splice site variants), as well as missense variants predicted to be damaging (Mis-D), are enriched in ADHD cases compared to controls. Combining our results with a large independent case-control DNA sequencing study (3,206 ADHD cases and 5,002 typically developing controls) (8), we identify *lysine demethylase 5B* (*KDM5B)* as a high-confidence risk gene for ADHD (FDR<0.1). Finally, we identify overlap among genes harboring *de novo* damaging variants in ADHD and previously reported risk genes for other psychiatric conditions, and we conduct exploratory analyses to identify biological pathway enrichment. These new findings provide a critical step forward toward improving our etiologic understanding of ADHD, which may, in the future, inform the treatment of this common and impairing condition.

Results:

Rare and ultra-rare de novo damaging variants are enriched in ADHD probands:

We performed whole-exome DNA sequencing in 152 parent-child trios with ADHD collected from four sites (**Dataset S1**). We pooled this sequencing data and performed joint variant calling with whole-exome sequencing from 788 parent-child trios without ADHD, already sequenced as part of the Simons Simplex Collection. After applying our quality control methods, we compared rates of *de novo* variants in 147 ADHD parent-child trios and 780 control parent-child trios. Based on studies of other childhood-onset neuropsychiatric conditions (9, 11, 17, 18), we expected to find a greater rate of rare *de novo* damaging variants in ADHD probands versus controls. Damaging variants include protein-truncating variants (PTVs, including premature stop codons, frameshift, and splice site variants) and missense variants predicted to be damaging (Mis-D) by a “missense badness, PolyPhen-2, constraint” (MPC) score > 2 (19).

Results from this burden analysis demonstrate a greater rate of both rare and ultra-rare *de novo* damaging variants (PTVs + Mis-D) in ADHD cases versus unaffected controls (**Figure 1**, **Table S1, Dataset S2**). For rare *de novo* damaging variants (non-neuro gnomAD allele frequency < 0.001), the rate ratio was 1.55 (95% CI 1.02-2.30). We found a greater difference

between cases and controls when narrowing our analysis to ultra-rare *de novo* damaging variants (non-neuro gnomAD allele frequency < 0.00005), with a rate ratio of 1.86 (95% CI 1.21-2.83, p=0.009) (**Table S1, Figure 1**). Within the subset of ultra-rare *de novo* damaging variants, we found a greater rate of PTVs (rate ratio 1.85, 95% CI 1.10-3.03, p=0.03) and a trend towards an increased rate of Mis-D variants in cases versus controls (rate ratio 1.91, 95% CI 0.78-4.36, p=0.12). As anticipated, we did not find differences in rates of *de novo* variants between cases and controls when including all (damaging and non-damaging) rare or all ultra-rare variants (**Table S1**).

Recurrent ultra-rare damaging variants identify ADHD risk genes:

Among 147 ADHD parent-child trios passing quality control, we identified 25 ultra-rare *de novo* damaging variants in 24 individuals (**Table 1, Dataset S2**). One gene, *KDM5B*, had two *de novo* PTVs in unrelated individuals in our ADHD trio cohort. To identify ADHD risk genes (genes harboring damaging variants more often than expected by chance), we combined our *de novo* parent-child trio findings with counts of ultra-rare PTVs and Mis-D (MPC > 2) variants identified in a large independent ADHD case-control dataset (3,206 ADHD cases and 5,002 typically developing controls) (8). Using this combined dataset, we applied the Transmission And De novo Association test (extTADA) (20) and identified *KDM5B* as a high-confidence risk gene (FDR=0.05), *POMT1* as a probable risk gene (FDR=0.21), and *YLPM1* as a probable risk gene (FDR=0.28) for ADHD (**Figure 2, Dataset S3**). This extTADA analysis estimates that 862 genes (95% CI 243-2,502) contribute to ADHD risk.

Genes with de novo damaging variants in ADHD overlap with risk genes for other psychiatric conditions:

Using the list of 25 genes with ultra-rare *de novo* damaging variants (PTV and Mis-D) in 147 ADHD probands (**Table 1, Dataset S2**), we identified overlap with risk genes for other conditions (**Table S2**), using the Gene4Denovo database (21). *KDM5B* is also a risk gene for autism spectrum disorder (FDR=0), general developmental disorders (FDR=0), congenital heart disease (FDR=.005), complex motor stereotypies (FDR=0.06), and across all disorders in the Gene4Denovo database (FDR=0). *FBXO11* and *STAG1* were also both associated with developmental disorders (FDR= 0 and 0.00002, respectively) and across disorders (FDR=0 for both), and *PAK1* was a risk gene across disorders (FDR=.009). Additionally, we identified

overlap between our list of 25 genes with ultra-rare *de novo* damaging variants in ADHD probands and gene-mapped loci from common variant GWAS studies in neuropsychiatric disorders in the GWAS Catalog (**Table S3**).

Exploratory gene ontology and pathway enrichment:

Using this same list of 25 genes harboring ultra-rare *de novo* damaging variants in ADHD trios, we also conducted exploratory analyses to identify enriched gene ontology and biological pathways. Several gene ontology and pathway-based sets were enriched for these 25 genes identified in ADHD (**Table S4**). The top pathway-based sets were CXCR4-mediated signaling events (q=0.004), Sema3A PAK dependent Axon repulsion (q=0.004), and ectoderm differentiation (q=0.008).

Discussion:

In this largest parent-child trio whole-exome DNA sequencing study of ADHD to date, we found a significantly greater rate of rare and ultra-rare *de novo* damaging variants in children with ADHD compared to unaffected controls (**Figure 1**). Combining our trio sequencing data with results from a large independent case-control DNA sequencing dataset, we identified *KDM5B* as a high-confidence risk gene and *POMT1* and *YLMP1* as two probable risk genes for ADHD (**Figure 2**).

Our sequencing data identified a 1.55-fold enrichment of rare *de novo* damaging variants in ADHD cases compared to unaffected controls (**Figure 1, Table S1**). Narrowing to ultra-rare *de novo* damaging variants increases this enrichment to 1.86-fold and strengthens statistical significance. It is important to note that these estimated enrichments have wide confidence intervals, so caution is warranted in interpreting these results, and replication in larger ADHD parent-child trio cohorts is needed. Nevertheless, our observed enrichment of rare *de novo* PTV and Mis-D variants is of a similar magnitude to enrichments reported in other neurodevelopmental disorders, including ASD and Tourette's disorder (9, 11). This enrichment of rare and ultra-rare *de novo* damaging variants in ADHD cases compared to controls is also consistent with findings from the largest case-control DNA sequencing study that observed an enrichment of rare protein-truncating variants in constrained genes in ADHD cases, and this rate was similar in ASD cases (8). Finding an enrichment of rare *de novo* damaging variants in ADHD adds information about the genomic architecture of ADHD and supports the value of

DNA sequencing studies in larger ADHD parent-child trio cohorts to identify risk genes in a manner which has led to the identification of over 100 high-confidence risk genes in ASD. The discovery potential of this approach is further reinforced by our estimate that 862 genes contribute to ADHD risk. Although the confidence interval of this estimate is wide and will need to be refined in larger cohorts, this finding further highlights a path toward systematic risk gene discovery in ADHD.

Our study identified ultra-rare *de novo* PTV variants in *KDM5B* in two unrelated individuals with ADHD (**Table 1, Dataset S2**). These individuals did not have diagnoses of ASD or intellectual disability. *KDM5B* is a histone-modifying enzyme that demethylates H3K4 and plays an important role in normal embryonic development (22), likely through epigenetic regulation of gene expression. This gene has been previously identified as a high-confidence risk gene for ASD (9) and developmental disorders more broadly (10). Interestingly, PTV mutations in *KDM5B* have also been reported in unaffected control subjects (8), and recessive mutations have been reported to cause a syndrome with developmental delay (23, 24). A review of the literature suggests that *KDM5B* likely causes an autosomal recessive developmental disorder, while dominant disease variants may exist (22). Our findings suggest, for the first time, that ADHD is included in the spectrum of phenotypic changes that may occur in the context of rare damaging variants in *KDM5B*.

We found several additional genes with *de novo* damaging variants in ADHD that are worth highlighting. First, we identified a *de novo* PTV in *POMT1*, which we identify as a probable risk gene for ADHD, based on FDR < 0.3 (**Table 1, Figure 2, Dataset S3**). POMT1 is a key enzyme in glycosylation of alpha-dystroglycan, and biallelic mutations are associated with dystroglycanopathies characterized by proximal muscular dystrophy (25) and often accompanied by intellectual disability (26). Second, we also identify *YLPM1* as a probable risk gene for ADHD. YLPM1 is involved in RNA binding and has been predicted to be involved in telomere maintenance, but to our knowledge, psychiatric manifestations related to *YLPM1* mutations have not been described previously. Third, although no other gene beyond *KDM5B* was found to have more than one ultra-rare *de novo* gene-damaging mutation in unrelated individuals using our definition of PTV and Mis-D, two additional genes, *CTNND2* and *EML6*, had a *de novo* PTV variant in one individual with ADHD and a *de novo* missense variant that was not predicted to be damaging (MPC<2) in an unrelated individual with ADHD (**Dataset S2**). Although not predicted to be damaging by MPC, the ultra-rare *de novo* missense variant in *CTNND2* was predicted to be damaging by PolyPhen2 (**Dataset S2**), and another PTV in *CTNND2* was identified in an ADHD case in the case-control dataset, but not in controls

(**Dataset S3**). *CTNND2* encodes an adhesive junction protein, and mutations have been previously associated with intellectual disability in Cri-du-Chat syndrome, ASD, and epilepsy (27-29). Research suggests that *CTNND2* is important for the formation of dendritic spines and synapses (28). Finally, we identified individuals with ADHD who had *de novo* damaging variants in the genes *FBXO11* and *STAG1* (**Table 1**). These two genes have been previously identified as high-confidence risk genes for neurodevelopmental disorders in general (10). We did not see damaging variants in these genes in controls (**Datasets S2 and S3**). *FBXO11* encodes an F-box protein, and *de novo* variants have been associated with syndromic intellectual disability and behavioral difficulties, including ADHD (30, 31). *STAG1* encodes a component of cohesion involved with the separation of sister chromatids and has been associated with syndromic intellectual disability (32). In our study, ultra-rare damaging *de novo* variants in these genes were identified in children with ADHD who did not have intellectual disability or other known genetic syndromes. This highlights the potential range of clinical manifestations that may occur due to *de novo* damaging variants in these genes and suggests potential clinical implications for identifying *de novo* damaging variants.

Genes harboring rare *de novo* gene-damaging variants in the ADHD cases not only overlapped with high-confidence risk genes identified in previous DNA sequencing studies of other neuropsychiatric conditions (**Table S2**) but also overlapped with genes mapped from genome-wide significant common variants identified in previous GWA studies (**Table S3**). Although there was no overlap with the 76 prioritized risk genes identified in the recent large ADHD GWAS (4), there was overlap between genes mapped from externalizing-related disorders more broadly (33). These findings add to the growing evidence supporting the convergence of common and rare variants in ADHD (4) and psychiatric disorders in general (9, 34).

Finally, we conducted exploratory ontology and pathway analyses of genes harboring *de novo* damaging variants in our ADHD cases. In interpreting these results, it is important to note that many of these genes may not be true ADHD risk genes and replication of these exploratory findings are needed as more high-confidence risk genes are identified. Nevertheless, we observed a significant enrichment of several biological processes. Of note, one of the top pathways is ectoderm differentiation (**Table S4**), suggesting early neurodevelopmental underpinnings of ADHD. In the largest recent GWAS study of ADHD, gene-linked loci were enriched for expression in early brain development (4), also suggesting the possible role of early embryonic changes in the development of ADHD.

This study has limitations that should be considered. For comparing mutation rates, the ideal controls would have been sequenced simultaneously with cases and assessed for ADHD. This study prioritized sequencing ADHD parent-child trios and used controls that had been previously sequenced using similar methods and scored in the normal range of the ADHD subscale of the Child Behavioral Checklist (CBCL) or the Adult Behavioral Checklist (ABCL). We attempted to minimize batch effects by focusing on the intersection of the capture platforms as done in other DNA sequencing studies (11, 18, 35). Another limitation is that our study focused on the coding region of the genome, and it is possible that important rare variants also occur in the noncoding region. Currently, understanding the biological and clinical relevance of non-coding variation remains challenging, but future studies of ADHD may utilize whole-genome sequencing technologies.

Despite these limitations, our results are important by demonstrating, for the first time, an enrichment of rare and ultra-rare *de novo* damaging variants in ADHD cases compared to unaffected controls and identifying *KDM5B* as a high-confidence risk gene for ADHD. These findings reinforce the value of DNA sequencing of parent-child trios in larger cohorts to identify additional high-confidence risk genes for ADHD. Identifying risk genes that can be studied in model systems may offer additional insight into the underlying biology of ADHD and has the potential to inform clinical care for individuals and families.

Materials and Methods:

Participants:

This study was approved by the local institutional review boards of all participating institutions and informed consent/assent was obtained from all participants. A total of 152 parent-child trios (456 individuals in total), comprising a child meeting DSM-IV or DSM-5 criteria for the diagnosis of ADHD and both biological parents, were included in this study. Trios were identified from four sites: the University of São Paulo School of Medicine (n=30), the Centre for Addiction and Mental Health in Toronto (n=37), Florida International University (n=13), and the Genizon biobank from Génome Québec (n=72). All subjects were assessed by structured clinical interviews. Exclusion criteria included a diagnosis of ASD, intellectual disability, psychosis, mood disorders (including bipolar disorder), and clinically significant medical or neurological disease. We prioritized the study of simplex (no known family history of ADHD) parent-child trios to increase the likelihood of detecting *de novo* variants. Control subjects were

788 unaffected parent-child trios, selected from the Simons Simplex Collection from the National Institutes of Health Data Archive (https://nda.nih.gov/edit_collection.html?id=2042) (36). Control subjects did not have ASD and were selected to be in the normal range for the attentional problems subscale from the CBCL or the ABCL (t score < 64.5), which has been shown to predict ADHD diagnosis (37).

Whole-exome DNA sequencing:

Exome capture and whole-exome DNA sequencing of DNA from 80 children with ADHD and their parents were conducted at the Yale Center for Genomic Analysis (YCGA) using the IDT xGen V1 capture and the Illumina NovaSeq6000 sequencing instrument. An additional 72 ADHD parent-child trios were sequenced by Genome Quebec using the Agilent SureSelect All Exon V7 capture and the Illumina NovaSeq6000 sequencing instrument. 788 control parent-child trios were previously sequenced as part of the Simons Simplex Collection, using the NimbleGen SeqCap EzExomeV2 capture and the Illumina HiSeq 2000 sequencing instrument. We performed joint variant calling with sequencing data from all cases and controls (940 trios, 2,820 individuals in total).

Sequencing alignment and variant identification:

Alignment and variant calling of the DNA sequencing reads followed the Genome Analysis Toolkit (GATK) best practice guidelines (38) as previously described by our group (35). To minimize potential downstream effects of differential coverage between the different capture platforms, a target bed file was created using the intersection of target regions of the three capture platforms (IDT xGen V1, Agilent SureSelect All Exon V7, and SeqCap EzExome V2). Case and control samples were called jointly using GATK GenotypeGVCF tools and variant score recalibration was applied to all called variants. Passing variants were then annotated using the RefSeq hg19 gene definitions and external databases using ANNOVAR (39).

Quality control of de novo variants:

Parent-child trios were excluded if unexpected family relationships were identified using relatedness statistics (40). Trios were also omitted if the children were observed to have an outlier number of *de novo* variants (>20). PLINK/SEQ istats was used to generate quality control

statistics for both cases and controls, and principal component analyses were used to remove outliers from the analysis (see **Figure S1** and **Dataset S1** for details). After these quality control steps, we analyzed 147 parent-child trios with ADHD and 780 control parent-child trios for *de novo* variants.

As previously described (35), we then used stringent thresholds to assess *de novo* mutations. This included that the child was heterozygous for the variant with an alternate allele frequency between 0.3 and 0.7 in the child and < 0.05 in the parents, sequencing depth of ≥ 20 in all family members at the variant position, alternate allele depth ≥ 5, mapping quality ≥ 30. Calls were limited to one variant per person per gene, retaining variants with the most severe consequence (9). We filtered to include rare *de novo* variants with an allele frequency < 0.001 (0.1%) in the "non-neuro" subset of the Genome Aggregation Database (gnomAD v2.2.1). Within this rare set of *de novo* variants, we defined an ultra-rare subset, defined as having an allele frequency of < 0.00005 in the non-neuro subset of gnomAD (41). The gnomAD v2.2.1 non-neuro dataset contains exome sequencing data from 104,068 individuals who were not ascertained for having a neurologic or psychiatric condition in case-control studies.

Mutation rate analysis:

We calculated the rate of *de novo* variants per base pair in cases and controls. The GATK DepthofCoverage tool was used to determine the denominator of the "callable" base pairs per family. We required callable bases to have a sequencing depth of ≥ 20x in all family members and mapping quality ≥ 30. We limited our analyses to variants in the callable exome to further minimize potential calling bias between cases and controls. Mutation rates were divided by 2 to calculate haploid rates, and confidence intervals were calculated (pois.conf.int, pois.exact function from epitools v0.5.10.1 in R). We used a one-tailed rate ratio test to compare *de novo* mutation rates between cases and controls (rateratio.test v1.1 in R). Based on studies of other childhood-onset neuropsychiatric conditions (9, 11, 17, 18), we hypothesized that rare and ultra-rare *de novo* PTV variants and missense variants predicted to be damaging (Mis-D) would be enriched in cases compared to controls. Mis-D variants were identified using the integrated "missense badness, PolyPhen-2, constraint" (MPC) score > 2 (19) as done in other recent studies (8, 18, 42). The combined group of *de novo* PTV and Mis-D variants were considered "damaging" variants.

Transmission and De Novo Association Test Analysis:

We used a Bayesian extension of the original Transmission And De novo Association test (extTADA) (20) to integrate *de novo* and case-control variants in a hierarchical model to increase the power of identification of risk genes for ADHD. We obtained mutation counts for PTVs and Mis-D variants (MPC > 2) from an independent case-control study including 3,206 individuals with ADHD and 5,002 typically developing controls (8). These individuals did not have diagnoses of autism, intellectual disability, schizophrenia, bipolar disorder, affective disorders, or anorexia (8). We ran extTADA to calculate the Bayes factor and q-values (false discovery rate, FDR) for each gene (details in Supplemental methods) (**Dataset S3**). We applied commonly used statistical thresholds to define ”probable” (FDR < 0.3) and ”high confidence” (FDR < 0.1) risk genes (17).

Gene set overlap:

We examined if our list of genes with ultra-rare *de novo* damaging variants (PTV or Mis-D) in the ADHD probands overlapped with genes implicated in other DNA sequencing studies and genome-wide association studies. The Gene4Denovo database (21) (http://www.genemed.tech/gene4denovo/home) integrates *de novo* mutations from 68,404 individuals across 37 different phenotypes, including several neuropsychiatric conditions, but not including ADHD. We assessed the overlap between the Gene4Denovo candidate gene list (release version updated 07/08/2022) with FDR < 0.1 and our list of genes with ultra-rare damaging *de novo* variants. The GWAS Catalog (43, 44) was used to examine if this same list of genes harboring *de novo* damaging variants overlapped with loci mapped to genes in previous genome-wide association studies of neuropsychiatric phenotypes. The GWAS Catalog identifies past studies through weekly PubMed searches and extracts data for SNPs with $p < 1 \times 10^{-5}$ in the overall (initial GWAS + replication) population. All curated trait descriptions in the GWAS Catalog are mapped to terms from the Experimental Factor Ontology (EFO), which provide a systematic description of traits to support the annotation, analysis, and visualization of data. We limited our overlap analysis to traits in the GWAS Catalog that were categorized under the umbrella terms ‘nervous system disease’ or ‘psychiatric disorder’ (additional details found at https://www.ebi.ac.uk/gwas//docs).

Exploratory pathway analysis:

We used ConsensusPathDB (45) (http://cpdb.molgen.mpg.de/, Latest Release 35, 05.06.2021) to assess if the list of genes harboring rare *de novo* damaging variants in ADHD probands were over-represented in gene-ontology based sets and biological pathways. This network interaction tool integrates information from 31 public resources. The following default settings were used for the gene set over-representation analysis with gene identifier type of Ensembl; Pathway based sets: pathways as defined by pathway databases, select all resources, minimum overlap with input list=2, p-value cutoff=0.01; Gene ontology categories: gene ontology levels 2, 3, 4, and 5 categories, select all categories, p-value cutoff=0.01.

Acknowledgments: We thank all of the families who contributed to this research study. We would also like to thank Kyle Satterstrom, Anders Børglum, and Mark Daly for sharing their case-control PTV and Mis-D counts for this analysis and all of the team members who contributed to data collection, including Chelsea Dale and Juliana Acosta. This work was supported by a Klingenstein Third Generation Foundation ADHD Fellowship grant (E.O.), a Yale Child Study Center Faculty Development Award (T.V.F.), and the Allison Family Foundation (T.V.F.). Brazilian samples were recruited and collected with support from São Paulo Research Foundation (FAPESP, grant 2016/22455-8). E.O. was supported by the National Institute of Mental Health grants R25MH077823, T32MH018268, and K08MH128665. L.C.F. was supported by grant #2021/08540-0, São Paulo Research Foundation (FAPESP). J.P. was supported by the Bradley Hospital COBRE Center for Sleep and Circadian Rhythms in Child and Adolescent Mental Health (P20GM139743, PI Carskadon). Seventy-two of the sequenced ADHD parent-child trios were from the Génome Québec Genizon Biobank. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Simons Simplex Collection (SSC) control parent-child trio genetic data used in the preparation of this manuscript were obtained from the NIH- supported National Database for Autism Research (NDAR). NDAR is a collaborative informatics system created by the National Institutes of Health to provide a national resource to support and accelerate research in autism. Dataset identifier: https://nda.nih.gov/edit_collection.html?id=2042. This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH or of the Submitters submitting original data to NDAR. We are grateful to all of the families at the participating SSC sites, as well as the principal investigators (A. Beaudet, R. Bernier, J. Constantino, E. Cook, E. Fombonne, D. Geschwind, R. Goin-Kochel, E. Hanson, D. Grice, A. Klin, D. Ledbetter, C. Lord, C. Martin, D. Martin, R. Maxim, J. Miles, O. Ousley, K. Pelphrey, B. Peterson, J. Piggot, C. Saulnier, M. State, W. Stone, J. Sutcliffe, C. Walsh, Z. Warren, E.

Wijsman). We appreciate obtaining access to phenotypic data on SFARI Base. Approved researchers can obtain the SSC population dataset described in this study (https://www.sfari.org/resource/simons-simplex-collection/) by applying at https://base.sfari.org.

Disclosures: The authors declare no conflict of interest. In the past 3 years, GVP has been consultant, advisory board member, and/or speaker for Aché, Abbott, Apsen, Medice, Novo Nordisk, Pfizer and Takeda. GVP also receives royalties from Editora Manole.

References:

1. G. V. Polanczyk, G. A. Salum, L. S. Sugaya, A. Caye, L. A. Rohde, Annual research review: A meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. *Journal of child psychology and psychiatry* **56**, 345-365 (2015).
2. J. Posner, G. V. Polanczyk, E. Sonuga-Barke, Attention-deficit hyperactivity disorder. *Lancet* **395**, 450-462 (2020).
3. S. V. Faraone, H. Larsson, Genetics of attention deficit hyperactivity disorder. *Molecular psychiatry* **24**, 562-575 (2019).
4. D. Demontis *et al.*, Genome-wide analyses of ADHD identify 27 risk loci, refine the genetic architecture, and implicate several cognitive domains. *Nature genetics*, 1-11 (2023).
5. D. Demontis *et al.*, Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nature genetics* **51**, 63-75 (2019).
6. E. J. Sonuga-Barke *et al.*, Annual Research Review: Perspectives on progress in ADHD science–from characterization to cause. *Journal of Child Psychology and Psychiatry* (2022).
7. B. Harich *et al.*, From rare copy number variants to biological processes in ADHD. *American Journal of Psychiatry* **177**, 855-866 (2020).
8. F. K. Satterstrom *et al.*, Autism spectrum disorder and attention deficit hyperactivity disorder have a similar burden of rare protein-truncating variants. *Nat Neurosci* **22**, 1961-1965 (2019).
9. F. K. Satterstrom *et al.*, Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. *Cell* **180**, 568-584 e523 (2020).
10. J. Kaplanis *et al.*, Evidence for 28 genetic disorders discovered by combining healthcare and research data. *Nature* **586**, 757-762 (2020).
11. S. Wang *et al.*, De novo sequence and copy number variants are strongly associated with tourette disorder and implicate cell polarity in pathogenesis. *Cell reports* **24**, 3441-3454 e3412 (2018).
12. A. C. Lionel *et al.*, Rare copy number variation discovery and cross-disorder comparisons identify risk genes for ADHD. *Science translational medicine* **3**, 95ra75-95ra75 (2011).
13. J. Martin *et al.*, A brief report: de novo copy number variants in children with attention deficit hyperactivity disorder. *Transl Psychiatry* **10**, 135 (2020).
14. B. R. Al-Mubarak *et al.*, Whole exome sequencing in ADHD trios from single and multi-incident families implicates new candidate genes and highlights polygenic transmission. *European Journal of Human Genetics* **28**, 1098-1110 (2020).

15. L. de Araújo Lima *et al.*, An integrative approach to investigate the respective roles of single-nucleotide variants and copy-number variants in Attention-Deficit/Hyperactivity Disorder. *Scientific reports* **6**, 22851 (2016).
16. D. S. Kim *et al.*, Sequencing of sporadic Attention-Deficit Hyperactivity Disorder (ADHD) identifies novel and potentially pathogenic de novo variants and excludes overlap with genes associated with autism spectrum disorder. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics* **174**, 381-389 (2017).
17. M. Halvorsen *et al.*, Exome sequencing in obsessive-compulsive disorder reveals a burden of rare damaging coding variants. *Nat Neurosci* **24**, 1071-1076 (2021).
18. E. Olfson *et al.*, Whole-exome DNA sequencing in childhood anxiety disorders identifies rare de novo damaging coding variants. *Depress Anxiety* **39**, 474-484 (2022).
19. K. E. Samocha *et al.*, Regional missense constraint improves variant deleteriousness prediction
bioRxiv 10.1101/148353 (2017).
20. H. T. Nguyen *et al.*, Integrated Bayesian analysis of rare exonic variants to identify risk genes for schizophrenia and neurodevelopmental disorders. *Genome medicine* **9**, 1-22 (2017).
21. G. Zhao *et al.*, Gene4Denovo: an integrated database and analytic platform for de novo mutations in humans. *Nucleic acids research* **48**, D913-D926 (2020).
22. J. Harrington, G. Wheway, S. Willaime-Morawek, J. Gibson, Z. S. Walters, Pathogenic KDM5B variants in the context of developmental disorders. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 194848 (2022).
23. V. Faundes *et al.*, Histone lysine methylases and demethylases in the landscape of human developmental disorders. *The American Journal of Human Genetics* **102**, 175-187 (2018).
24. H. C. Martin *et al.*, Quantifying the contribution of recessive coding variation to developmental disorders. *Science (New York, N.Y.)* **362**, 1161-1164 (2018).
25. T. Geis *et al.*, Clinical long-time course, novel mutations and genotype-phenotype correlation in a cohort of 27 families with POMT1-related disorders. *Orphanet Journal of Rare Diseases* **14**, 1-17 (2019).
26. D. Song *et al.*, Genetic variations and clinical spectrum of dystroglycanopathy in a large cohort of Chinese patients. *Clinical genetics* **99**, 384-395 (2021).
27. M. Medina, R. C. Marinescu, J. Overhauser, K. S. Kosik, Hemizygosity of δ-catenin (CTNND2) is associated with severe mental retardation in cri-du-chat syndrome. *Genomics* **63**, 157-164 (2000).
28. T. N. Turner *et al.*, Loss of δ-catenin function in severe autism. *Nature* **520**, 51-56 (2015).
29. A.-F. van Rootselaar *et al.*, δ-Catenin (CTNND2) missense mutation in familial cortical myoclonic tremor and epilepsy. *Neurology* **89**, 2341-2350 (2017).
30. A. Gregor *et al.*, De novo variants in the F-box protein FBXO11 in 20 individuals with a variable neurodevelopmental disorder. *The American Journal of Human Genetics* **103**, 305-316 (2018).
31. S. Jansen *et al.*, De novo variants in FBXO11 cause a syndromic form of intellectual disability with behavioral problems and dysmorphisms. *European Journal of Human Genetics* **27**, 738-746 (2019).
32. D. Lehalle *et al.*, STAG1 mutations cause a novel cohesinopathy characterised by unspecific syndromic intellectual disability. *Journal of medical genetics* **54**, 479-488 (2017).
33. R. Karlsson Linnér *et al.*, Multivariate analysis of 1.5 million people identifies genetic associations with traits related to self-regulation and addiction. *Nature neuroscience* **24**, 1367-1376 (2021).

34. V. Trubetskoy *et al.*, Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* **604**, 502-508 (2022).
35. C. Cappi *et al.*, De novo damaging DNA coding mutations are associated with obsessive-compulsive disorder and overlap with Tourette's disorder and autism. *Biological psychiatry* **87**, 1035-1044 (2020).
36. G. D. Fischbach, C. Lord, The Simons Simplex Collection: a resource for identification of autism genetic risk factors. *Neuron* **68**, 192-195 (2010).
37. W. J. Chen, S. V. Faraone, J. Biederman, M. T. Tsuang, Diagnostic accuracy of the Child Behavior Checklist scales for attention-deficit hyperactivity disorder: a receiver-operating characteristic analysis. *Journal of consulting and clinical psychology* **62**, 1017 (1994).
38. A. McKenna *et al.*, The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research* **20**, 1297-1303 (2010).
39. K. Wang, M. Li, H. Hakonarson, ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research* **38**, e164 (2010).
40. A. Manichaikul *et al.*, Robust relationship inference in genome-wide association studies. *Bioinformatics (Oxford, England)* **26**, 2867-2873 (2010).
41. K. J. Karczewski *et al.*, The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434-443 (2020).
42. P. Feliciano *et al.*, Exome sequencing of 457 autism families recruited online provides evidence for autism risk genes. *NPJ Genom Med* **4**, 19 (2019).
43. A. Buniello *et al.*, The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic acids research* **47**, D1005-D1012 (2019).
44. E. Sollis *et al.*, The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic acids research* **51**, D977-D985 (2023).
45. R. Herwig, C. Hardt, M. Lienhard, A. Kamburov, Analyzing and interpreting genome data at the network level with ConsensusPathDB. *Nature protocols* **11**, 1889-1907 (2016).

Table 1: Ultra-rare *de novo* damaging variants identified in ADHD probands

ID	Gene[a]	Protein change	Genomic coordinate	Ref. allele	Alt. allele	Variant class[b]	Non-neuro gnomAD AF
ADHD6.p1	*RBBP6*	p.Y170C	Chr16:24567213	A	G	missense (MPC 2.1)	0
ADHD8.p1	*POMT1*	p.V563fs	Chr9:134398355	G	GTCCAACAC	frameshift insertion	0
ADHD14.p1	*YLPM1*	p.R1646X	Chr14:75276497	C	T	stopgain	0
ADHD25.p1	*SECISBP2L*	p.T934fs	Chr15:49284809	TG	T	frameshift deletion	0
ADHD33.p1	*NARS2*	p.R54X	Chr11:78282471	G	A	stopgain	3.36E-05
ADHD37.p1	*CTNNA2*	p.R348W	Chr2:80808942	C	T	missense (MPC 2.6)	0
ADHD44.p1	*BEST4*	p.L207fs	Chr1:45251759	CAGAG	C	frameshift deletion	0
ADHD50.p1	***KDM5B***	p.R1093X	Chr1:202704703	G	A	stopgain	1.12E-05
ADHD57.p1	*STAG1*	p.R1088X	Chr3:136068009	G	A	stopgain	0
ADHD58.p1	***KDM5B***	p.D806fs	Chr1:202711840	TC	T	frameshift deletion	0
ADHD61.p1	*EML6*	p.R313X	Chr2:55071273	C	T	stopgain	0
ADHD69.p1	*CFL1*	p.Y82C	Chr11:65623472	T	C	missense (MPC 3.2)	0
ADHD71.p1	*OCEL1*	p.G34X	Chr19:17337532	G	T	stopgain	0
	TTC26	p.G24D	Chr7:138819468	G	A	splicing	0
ADHD84.p1	*PLD5*	p.D28fs	Chr1:242511463	TG	T	frameshift deletion	0
ADHD86.p1	*FBXO11*	p.P918S	Chr2:48035289	G	A	missense (MPC 2.6)	0
ADHD95.p1	*CHST15*	p.L214fs	Chr10:125804342	G	GGT	frameshift insertion	0
ADHD98.p1	*EHBP1L1*	p.M1429fs	Chr11:65359271	CATGG	C	frameshift deletion	0
ADHD99.p1	*TUBB*	p.G233E	Chr6:30691669	G	A	missense (MPC 3.3)	0
ADHD107.p1	*CTNND2*	p.V229fs	Chr5:11236867	AC	A	frameshift deletion	0
ADHD117.p1	*GOLGB1*	p.Y288X	Chr3:121435768	A	C	stopgain	0
ADHD130.p1	*L1TD1*	p.S580X	Chr1:62676185	C	A	stopgain	0
ADHD134.p1	*GNB2L1*	p.N244D	Chr5:180665146	T	C	missense (MPC 2.2)	0
ADHD141.p1	*PAK1*	c.C44T	Chr11:77103522	G	A	missense (MPC 2.8)	0
ADHD144.p1	*USP54*	p.R58X	Chr10:75331247	G	A	stopgain	3.30E-05

Ref. allele, reference allele; Alt. allele, alternative allele; gnomAD, the Genome Aggregation Database; AF, allele frequency; MPC, “missense badness, PolyPhen-2, constraint” scores.
[a]Genetic variants were annotated with ANNOVAR using RefGene hg19 definitions.
[b]Damaging *de novo* variants were defined as protein-truncating variants (frameshift insertions, frameshift deletions, stop codon change, canonical splice site variants) or missense variants predicted to be damaging by a MPC “missense badness, PolyPhen-2, constraint” score >2. For missense variants, the MPC score of the variant is listed in parenthesis.

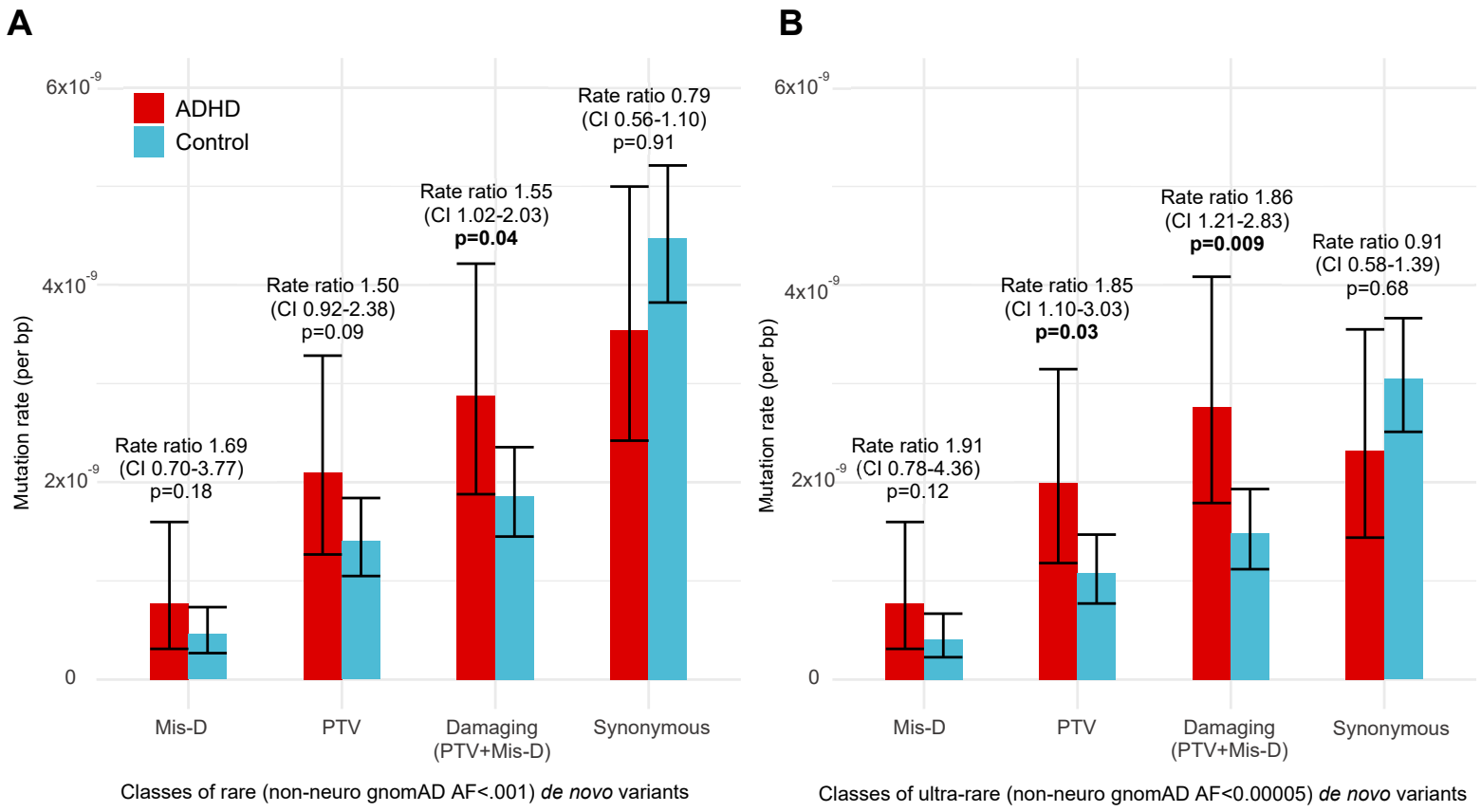

Figure 1: Rates of (A) rare and (B) ultra-rare *de novo* damaging mutations are enriched in ADHD probands (n=147) compared to controls (n=780). Rare variants have an allele frequency <0.001 (0.1%) in the non-neuro subset of the Genome Aggregation Database (gnomAD) and ultra-rare de novo variants have an allele frequency of <0.00005 (0.005%) in the non-neuro subset of gnomAD. The mutation rate per base pair (bp) includes only the "callable" loci in each family that meet sequencing depth and quality scores. Mutation rates are compared with a one-tailed rate ratio test with a p<0.05 considered significant. Bold p-values are significant. Error bars show 95% confidence intervals. PTVs, protein truncating variants, including frameshift, splice site, and stop-gain variants. Mis-D, missense variants predicted to be damaging with "missense badness, PolypPhen-2 constraint" (MPC) score>2.

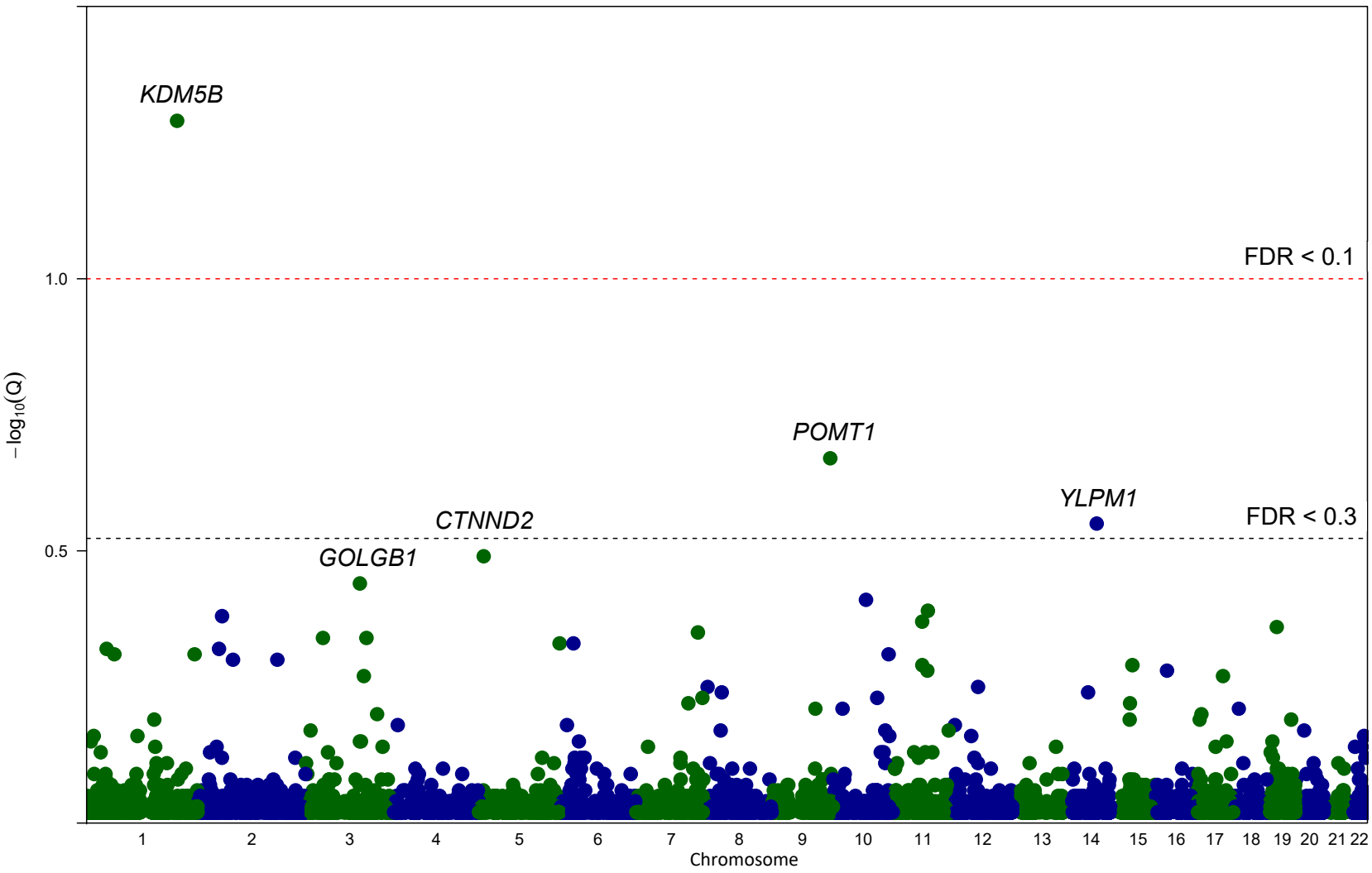

Figure 2: Gene-based test results for ADHD, combining the ultra-rare *de novo* damaging variants and independent case-control data. Results from extension of the Transmission And De novo Association test (extTADA) examining ultra-rare *de novo* protein-truncating variants and missense variants predicted to be damaging (MPC score>2) from the 147 ADHD parent-child trios and an independent group of 3,206 ADHD cases and 5,002 unaffected controls. Genes are

organized by chromosome and the top 5 gene symbols are listed that have the lowest q values. Only one gene *KDM5B* is classified as a high-confidence risk gene (FDR, false discovery rate < 0.1) and two genes, *POMT1* and *YLPM1*, are classified as probable risk genes (FDR<0.3)

Supporting Information for

Ultra-rare *de novo* damaging coding variants are enriched in attention-deficit/hyperactivity disorder and identify risk genes

Emily Olfson[1,2*], Luis C. Farhat[1,3], Wenzhong Liu[1], Lawrence A. Vitulano[1], Gwyneth Zai[4,5], Monicke O. Lima[3], Justin Parent[6,7,8], Guilherme V. Polanczyk[3], Carolina Cappi[9], James L. Kennedy[4,5], Thomas V. Fernandez[1,10*]

[1] Child Study Center, Yale University
[2] Wu Tsai Institute, Yale University
[3] Faculdade de Medicina FMUSP, Universidade de São Paulo
[4] Neurogenetics Section, Molecular Brain Sciences Department, Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health
[5] Institute of Medical Science and Department of Psychiatry, University of Toronto
[6] Center for Children and Families, Florida International University
[7] Bradley/Hasbro Children's Research Center, E.P. Bradley Hospital
[8] Alpert Medical School of Brown University
[9] Icahn School of Medicine at Mount Sinai
[10] Department of Psychiatry, Yale University

*Corresponding authors: Emily Olfson and Thomas Fernandez
Email: emily.olfson@yale.edu, thomas.fernandez@yale.edu

This PDF file includes:

Supporting text
Figure S1
Tables S1 to S4
Legends for Datasets S1 to S3
SI References

Other supporting materials for this manuscript include the following:

Datasets S1 to S3

Supporting Information Text

Details of Transmission and De Novo Association Test Analysis:

We ran extTADA following the code outlined at https://github.com/hoangtn/extTADA/blob/master/examples/extTADA_OneStep.ipynb (1). Unlike the original TADA, extTADA uses a Markov chain Monte Carlo (MCMC) approach to calculate all parameters that are used as input in the traditional TADA (2) through sampling from the posterior in one step with resulting credible intervals. Parameter estimation led to the following estimates of (1) proportion of risk genes (π) (lower-upper credible intervals): 4.40% (1.24% – 12.79%); (2) average relative risk (γ) (lower-upper credible intervals): misD DN = 13.96 (1.22 – 67.51), PTV DN = 20.42 (3.27 – 66.51), misD CC = 1.60 (1.0 – 4.63), PTV CC = 1.74 (1.12 – 5.52); and (3) variability in relative risk estimates per gene (β) (lower-upper credible intervals): misD DN = 0.83, PTV DN = 0.82, misD CC = 5.65, PTV CC = 3.54. These parameters were used by the extTADA function to calculate the Bayes factor and q-values (FDR) for each gene.

For the calculation of absolute number of ADHD risk genes, we multiplied the total number of genes included in the extTADA analysis (19,560) by the proportion of risk genes estimated by the extTADA pipeline. All genes from the list generated by denovolyzeR(3) except for American College of Medical Genetics genes (*ACTA2, ACTC1, APC, APOB, ATP7B, BMPR1A, BRCA1, BRCA2, CACNA1S, COL3A1, DSC2, DSG2, DSP, FBN1, GLA, KCNH2, KCNQ1, LDLR, LMNA, MEN1, MLH1, MSH2, MSH6, MUTYH, MYBPC3, MYH11, MYH7, MYL2, MYL3, NF2, OTC, PCSK9, PKP2, PMS2, PRKAG2, PTEN, RB1, RET, RYR1, RYR2, SCN5A, SDHAF2, SDHB, SDHC, SDHD, SMAD3, SMAD4, STK11, TGFBR1, TGFBR2, TMEM43, TNNI3, TNNT2, TP53, TPM1, TSC1, TSC2, VHL*) were included in the exTADA analysis.

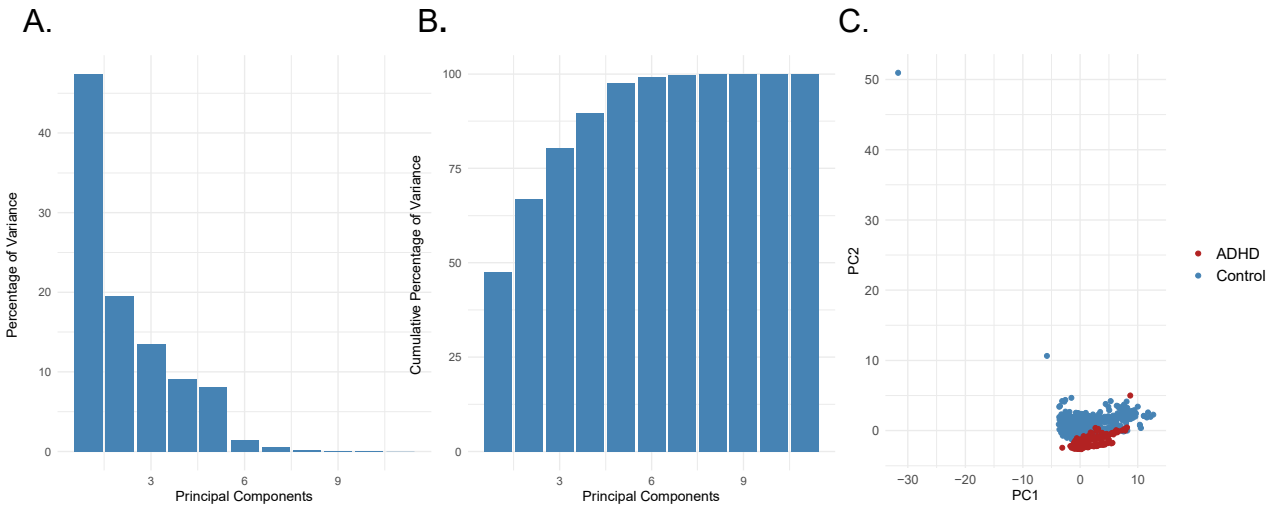

Fig. S1. Plots from the principal components analysis (PCA). (A) Shows the percentage of variance captured by the 11 principal components from the exome metrics data from cases and controls. (B) Shows the cumulative percentage of variance captured by these components and demonstrates that over 75% of the cumulative variance was captured by the first 3 principal components. (C) Shows the first two principal components based on the PCA of the exome sequencing quality metrics. ADHD cases are plotted in red and controls in blue. This figure includes PCA outliers (>5 standard deviations in PCAs 1-3), which were removed during the quality control.

Table S1. Distribution of classes of rare and ultra-rare *de novo* variants in ADHD cases and controls

	Variant counts		Mutation rate (x10-8) per basepair (95% CI)[b]		Estimated coding variants per individual (95% CI)[c]		Rate Ratio (95% CI)[d]	*p*-value
	ADHD case (n=147)	Control (n=780)	ADHD case (n=147)	Control (n=780)	ADHD case (n=147)	Control (n=780)		
Rare *de novo* variant class (non-neuro gnomAD AF<0.001)[a]								
Synonymous SNVs	32	166	0.35 (0.24-0.50)	0.45 (0.38-0.52)	0.23 (0.16-0.32)	0.29 (0.25-0.34)	0.79 (0.56-1.10)	0.91
All missense (Mis)[e]	95	430	1.05 (0.85-1.28)	1.16 (1.05-1.27)	0.68 (0.55-0.83)	0.75 (0.68-0.83)	0.91 (0.75-1.10)	0.82
Missense with MPC 0-1 (Mis-B)[f]	68	308	0.75 (0.58-0.95)	0.83 (0.74-0.93)	0.49 (0.38-0.62)	0.54 (0.48-0.60)	0.91 (0.72-1.13)	0.79
Missense with MPC 1-2 (Mis-P)[g]	15	76	0.17 (0.09-0.27)	0.20 (0.16-0.26)	0.11 (0.06-0.18)	0.13 (0.10-0.17)	0.81 (0.48-1.31)	0.81
Missense with MPC >2 (Mis-D)[h]	7	17	0.08 (0.03-0.16)	0.05 (0.03-0.07)	0.05 (0.02-0.10)	0.03 (0.02-0.05)	1.69 (0.70-3.77)	0.18
All PTV[i]	19	52	0.21 (0.13-0.33)	0.14 (0.10-0.18)	0.14 (0.08-0.21)	0.09 (0.07-0.12)	1.50 (0.92-2.38)	0.09
PTV frameshift indels	9	32	0.10 (0.05-0.19)	0.09 (0.06-0.12)	0.06 (0.03-0.12)	0.06 (0.04-0.08)	1.15 (0.56-2.23)	0.41
PTV stopgain	9	20	0.10 (0.05-0.19)	0.05 (0.03-0.08)	0.06 (0.03-0.12)	0.04 (0.02-0.05)	1.85 (0.85-3.77)	0.10
PTV splicing	1	0	0.01 (0.00-0.06)	0.00 (0.00-0.01)	0.01 (0.00-0.04)	0.00 (0.00-0.01)	Inf (0.22-Inf)	0.20
Nonframeshift indels	3	7	0.03 (0.01-0.10)	0.02 (0.01-0.04)	0.02 (0.00-0.06)	0.01 (0.00-0.03)	1.76 (0.39-6.33)	0.31
Damaging (PTV + Mis-D)	26	69	0.29 (0.19-0.42)	0.19 (0.14-0.24)	0.19 (0.12-0.27)	0.12 (0.09-0.15)	1.55 (1.02-2.30)	**0.04**
All[j]	150	662	1.66 (1.40-1.95)	1.79 (1.65-1.93)	1.08 (0.91-1.26)	1.16 (1.07-1.25)	0.93 (0.80-1.08)	0.80
Ultra-rare *de novo* variant class (non-neuro gnomAD AF<0.00005)[a]								
synonymous SNVs	21	113	0.23 (0.14-0.36)	0.30 (0.25-0.37)	0.15 (0.09-0.23)	0.20 (0.16-0.24)	0.76 (0.49-1.14)	0.90
All missense (Mis)[k]	73	344	0.81 (0.63-1.02)	0.93 (0.83-1.03)	0.52 (0.41-0.66)	0.60 (0.54-0.67)	0.87 (0.70-1.08)	0.87
Missense with MPC 0-1 (Mis-B)[f]	50	243	0.55 (0.41-0.73)	0.66 (0.58-0.74)	0.36 (0.27-0.47)	0.43 (0.37-0.48)	0.84 (0.64-1.10)	0.88
Missense with MPC 1-2 (Mis-P)[g]	13	62	0.14 (0.08-0.25)	0.17 (0.13-0.21)	0.09 (0.05-0.16)	0.11 (0.08-0.14)	0.86 (0.48-1.45)	0.73
Missense with MPC >2 (Mis-D)[h]	7	15	0.08 (0.03-0.16)	0.04 (0.02-0.07)	0.05 (0.02-0.1)	0.03 (0.01-0.04)	1.91 (0.78-4.36)	0.12
All PTV[i]	18	40	0.20 (0.12-0.31)	0.11 (0.08-0.15)	0.13 (0.08-0.2)	0.07 (0.05-0.1)	1.85 (1.10-3.03)	**0.03**
PTV frameshift indels	8	25	0.09 (0.04-0.17)	0.07 (0.04-0.1)	0.06 (0.02-0.11)	0.04 (0.03-0.06)	1.31 (0.6-2.68)	0.31
PTV stopgain	9	15	0.1 (0.05-0.19)	0.04 (0.02-0.07)	0.06 (0.03-0.12)	0.03 (0.01-0.04)	2.46 (1.1-5.28)	0.03
PTV splicing	1	0	0.01 (0.00-0.06)	0.00 (0.00-0.01)	0.01 (0.00-0.04)	0.00 (0.00-0.01)	Inf (0.22-Inf)	0.20
Nonframeshift indels	1	5	0.01 (0.00-0.06)	0.01 (0.00-0.03)	0.01 (0.00-0.04)	0.01 (0.00-0.02)	0.82 (0.04-5.71)	0.73
Damaging (PTV+Mis-D)	25	55	0.28 (0.18-0.41)	0.15 (0.11-0.19)	0.18 (0.12-0.27)	0.10 (0.07-0.13)	1.86 (1.21-2.83)	**0.009**
All[l]	114	508	1.26 (1.04-1.52)	1.37 (1.25-1.49)	0.82 (0.68-0.98)	0.89 (0.81-0.97)	0.92 (0.77-1.10)	0.80

gnomAD, the Genome Aggregation Database; AF, allele frequency; CI, confidence interval; SNVs, single nucleotide variants; PTV, protein-truncating variants; indel, insertion-deletion variants; Mis-B, benign missense variants; Mis-P, possibly damaging missense variants; Mis-D, damaging missense variants.
[a]Variants were annotated with ANNOVAR using RefSeq hg19 definitions.

[b]*De novo* mutation rates were calculated as the number of variants divided by the number of haploid "callable" base pairs.
[c]The estimated number of *de novo* mutations per individual was calculated by multiplying by the size of the RefSeq hg19 coding exome (33,828,798 bp).
[d]Rates were compared using a 1-sided rate ratio test with p-value <0.05 considered significant. Bold indicates significant p-values.
[e]All rare missense variants contain 5 rare missense variants in ADHD cases and 29 rare in control trios that were not annotated by the "missense badness, PolyPhen-2, constraint" scores (MPC).
[f]Missense variants with an MPC score of 0-1.
[g]Missense variants with an MPC score of 1-2.
[h]Missense variants with an MPC score of >2.
[i]Includes frameshift indels, premature stop codons, and canonical splice site variants
[j]All rare annotated coding variants, which includes 3 variant annotated as unknown in ADHD cases and 7 in controls.
[k]All ultra-rare missense variants contain 3 missense variants in ADHD cases and 24 in control trios that were not annotated by the "missense badness, PolyPhen-2, constraint" scores (MPC).
[l]All ultra-rare annotated coding variants, which includes 1 variant annotated as unknown in ADHD cases and 6 in controls.

Table S2. Overlap between genes harboring ultra-rare *de novo* damaging variants (PTV + Mis-D) in ADHD probands and genes identified in other DNA sequencing studies of parent-child trios using the Gene4Denovo database

Group name	Gene symbol	Mutation rate	LoF	Del.	Func.	Tol.	Syn.	Nonfram.	P-value	FDR	Sample size
ASD	*KDM5B*	0.00007	11	12	23	1	0	0	0	**0**	16991
Combined Disorders	*FBXO11*	0.00003	11	11	22	4	0	0	0	**0**	62549
Combined Disorders	*KDM5B*	0.00007	38	23	61	3	3	1	0	**0**	62549
Combined Disorders	*STAG1*	0.00005	6	10	16	1	1	1	0	**0**	62549
UDD	*FBXO11*	0.00003	8	10	18	3	0	0	0	**0**	31085
UDD	*KDM5B*	0.00007	20	10	30	2	3	1	0	**0**	31085
UDD	*STAG1*	0.00005	4	6	10	1	1	0	0	**0.00002**	31085
CHD	*KDM5B*	0.00007	3	0	3	0	0	0	0.00001	**0.00528**	3408
Combined Disorders	*PAK1*	0.00002	1	5	6	1	0	0	0.00018	**0.00918**	62549
CMS	*KDM5B*	0.00007	2	0	2	0	0	0	0	**0.0601**	118
ID	*FBXO11*	0.00003	1	0	1	0	0	0	0.00215	0.129	841
UDD	*PAK1*	0.00002	1	3	4	1	0	0	0.0018	0.145	31085
ID	*TUBB*	0.00002	0	1	1	0	0	0	0.00546	0.21	841
NDDs	*PAK1*	0.00002	0	1	1	0	0	0	0.00505	0.259	637
ID	*STAG1*	0.00005	0	1	1	0	0	0	0.00883	0.31	841
NDDs	*STAG1*	0.00005	0	1	1	0	0	0	0.00729	0.321	637
TD	*KDM5B*	0.00007	1	0	1	0	0	0	0.00393	0.349	909
CH	*FBXO11*	0.00003	1	0	1	0	0	0	0.00057	0.39	232
ASD	*FBXO11*	0.00003	1	1	2	1	0	0	0.0144	0.411	16991
OCD	*STAG1*	0.00005	1	0	1	0	0	0	0.00329	0.506	906
PNH	*GOLGB1*	0.0001	1	0	1	0	0	0	0.00096	0.52	202
CDH	*KDM5B*	0.00007	0	1	1	0	0	0	0.0104	0.621	827
SCZ	*POMT1*	0.00004	0	1	1	0	0	0	0.032	0.683	3402
SCZ	*USP54*	0.00006	1	0	1	0	0	0	0.0332	0.688	3402
SCZ	*KDM5B*	0.00007	1	0	1	0	0	0	0.0341	0.692	3402
SCZ	*CTNNA2*	0.00005	0	1	1	0	0	0	0.0357	0.7	3402
SCZ	*STAG1*	0.00005	0	1	1	0	0	0	0.0361	0.701	3402
ASD	*STAG1*	0.00005	1	1	2	0	0	1	0.0679	0.703	16991
SCZ	*YLPM1*	0.00009	1	0	1	0	1	0	0.0409	0.719	3402
SCZ	*GOLGB1*	0.0001	0	1	1	1	0	0	0.0449	0.739	3402
ASD	*TTC26*	0.00002	0	1	1	0	0	0	0.0863	0.741	16991

ASD	*PAK1*	0.00002	0	1	1	0	0	0	0.0979	0.76	16991
Combined Disorders	*EML6*	0.00006	4	3	7	7	2	1	0.147	0.81	62549
Combined Disorders	*CTNNA2*	0.00005	0	6	6	6	0	0	0.204	0.849	62549
Combined Disorders	*SECISBP2L*	0.00004	2	2	4	1	1	0	0.228	0.861	62549
Combined Disorders	*BEST4*	0.00002	0	1	1	0	0	0	0.281	0.881	62549
UDD	*BEST4*	0.00002	0	1	1	0	0	0	0.197	0.893	31085
ASD	*USP54*	0.00006	1	1	2	2	0	0	0.337	0.901	16991
UDD	*EML6*	0.00006	4	2	6	4	2	0	0.245	0.911	31085
Combined Disorders	*PLD5*	0.00002	0	1	1	2	0	0	0.415	0.913	62549
Combined Disorders	*CTNND2*	0.00006	2	5	7	1	2	0	0.445	0.918	62549
Combined Disorders	*NARS2*	0.00002	0	1	1	0	1	0	0.465	0.921	62549
UDD	*CTNNA2*	0.00005	0	5	5	3	0	0	0.288	0.922	31085
ASD	*SECISBP2L*	0.00004	0	1	1	1	0	0	0.465	0.923	16991
Combined Disorders	*TTC26*	0.00002	0	1	1	2	0	0	0.482	0.924	62549
UDD	*PLD5*	0.00002	0	1	1	1	0	0	0.32	0.928	31085
Combined Disorders	*TUBB*	0.00002	0	1	1	2	0	0	0.572	0.934	62549
UDD	*SECISBP2L*	0.00004	2	1	3	0	1	0	0.36	0.935	31085
UDD	*NARS2*	0.00002	0	1	1	0	1	0	0.371	0.937	31085
ASD	*YLPM1*	0.00009	2	0	2	4	1	0	0.661	0.942	16991
Combined Disorders	*CHST15*	0.00003	1	0	1	2	0	0	0.728	0.946	62549
Combined Disorders	*POMT1*	0.00004	0	1	1	1	0	0	0.839	0.952	62549
UDD	*CTNND2*	0.00006	2	4	6	1	1	0	0.52	0.953	31085
ASD	*EML6*	0.00006	0	1	1	1	0	0	0.86	0.954	16991
ASD	*CTNND2*	0.00006	0	1	1	0	1	0	0.898	0.955	16991
Combined Disorders	*USP54*	0.00006	2	1	3	3	0	0	0.892	0.955	62549
ASD	*RBBP6*	0.00007	0	1	1	2	2	0	0.915	0.956	16991
Combined Disorders	*EHBP1L1*	0.00005	1	0	1	2	0	0	0.916	0.956	62549
Combined Disorders	*YLPM1*	0.00009	3	2	5	10	2	0	0.925	0.956	62549
Combined Disorders	*GOLGB1*	0.0001	3	2	5	12	1	0	0.958	0.957	62549
Combined Disorders	*RBBP6*	0.00007	0	2	2	6	5	2	0.959	0.957	62549

ASD	*GOLGB1*	0.0001	1	0	1	1	1	0	0.969	0.958	16991
UDD	*CHST15*	0.00003	1	0	1	1	0	0	0.673	0.963	31085
UDD	*EHBP1L1*	0.00005	1	0	1	1	0	0	0.887	0.971	31085
UDD	*RBBP6*	0.00007	0	1	1	3	3	2	0.955	0.972	31085
UDD	*GOLGB1*	0.0001	1	1	2	6	0	0	0.983	0.973	31085
UDD	*YLPM1*	0.00009	0	2	2	4	0	0	0.975	0.973	31085

LoF, loss-of-function; Del., Deleterious nonsynonymous; Func., Functional; Tol., Tolerable nonsynonymous; Syn., synonymous; Nonfram., Non-frameshift; Combined Disorders, integration of all disorders; ASD, autism spectrum disorder; CDH, congenital diaphragmatic hernia; CHD, congenital heart disease; CH, congenital hydrocephalus; CMS, complex motor stereotypies; ID, intellectual disability; NDDs, neurodevelopmental disorder; OCD, obsessive-compulsive disorder; PNH, periventricular nodular heterotropia; SCZ, schizophrenia; TD, Tourette disorder; UDD, undiagnosed developmental disorder.

The Gene4Denovo database integrates *de novo* mutations from 68,404 individuals across 37 different phenotypes, including several neuropsychiatric conditions, but not including ADHD. We assessed the overlap between the Gene4Denovo candidate gene list (release version updated 07/08/2022) and our list of ultra-rare de novo damaging variants. Bold denotes FDR (false discovery rate) <0.1, consistent with high-confidence risk genes.

Table S3. Overlap between the genes harboring ultra-rare de novo damaging variants (PTV or Mis-D) in ADHD probands and loci mapped to genes in genome-wide association studies (GWAS) of neuropsychiatric conditions using the GWAS Catalog

Mapped gene in GWAS catalog	Variant and risk allele	Genomic coordinate	P-value	Reported neuropsychiatric trait in GWA study	PMID
PLD5	rs11802387	1:242256712	7 x 10-6	Dementia in non-APOE e4 carriers	35694926
	rs1553441	1:242546741	6 x 10-6	Hypersomnia during a major depressive episode in bipolar disorder	26207136
L1TD1	rs2886644	1:62210612	**6 x 10-14**	Neuroblastoma (pediatric)	3259975
FBXO11, MSH6	rs7562367-G	2:47805474	**1 x 10-9**	Externalizing behaviour (multivariate analysis)	34446935
FBXO11	rs43811823	2:47820196	6 x 10-6	Schizophrenia	23894747
	rs77969729	2:47879940	**6 x 10-11**	Alzheimer's disease polygenic risk score (upper quantile vs lower quantile)	35589863
	rs2881935-T	2:47892134	**1 x10-8**	Insomnia	35835914
CTNNA2	rs13407231-C	2:79206501	1 x 10-6	Schizophrenia	35396580
	rs1897784-C	2:79269432	**5 x 10-9**	Externalizing behaviour (multivariate analysis)	34446935
	rs13409348-G	2:79312862	3 x 10-6	Bipolar disorder	19416921
	rs399885	2:79460126	5 x 10-7	Response to antipsychotic treatment	20195266
	rs7570469	2:79482228	6 x 10-7		
	rs10196867-C	2:79751234	**5 x 10-9**	Alcohol dependence or heroin dependence or methamphetamine dependence	31462767
	rs17018359	2:79989050	2 x 10-6	Schizophrenia (MTAG)	32107650
	rs10180106-A	2:79994772	5 x 10-7 2 x 10-6	Response to lurasidone in schizophrenia	29730043
	rs6738962	2:80054047	**1 x 10-8**	Alzheimer's disease (cognitive decline)	23535033
ANKRD11P1, CTNNA2	rs55803020-C	2:80465766	8 x 10-14	Externalizing behaviour (multivariate analysis)	34446935
	rs10179482-G	2:80777501	**3 x 10-8**		
GOLGB1	rs115630863-A	3:121671594	3 x 10-6	Response to antidepressants in major depressive disorder	36228427
STAG1	rs6770476	3:136355078	**6 x 10-9**	Attention deficit hyperactivity disorder or autism spectrum disorder or intelligence (pleiotropy)	35764056
	rs10935182	3:136418580	**7 x 10-10**	Schizophrenia	31740837
	rs10935184-T	3:136434626	**1 x 10-14**		35396580
	rs10935184-C		**1 x 10-9**	Neuroticism	32231276
	rs66691851-C	3:136435986	**4 x 10-15**	Schizophrenia	28991256
			2 x 10-11		26198764
			2 x 10-10		30285260
			2 x 10-13		
	rs66691851		**4 x 10-18**		31740837
	rs7427564	3:136555593	**2 x 10-11**		31268507
	rs7427564-G		**2 x 10-10**		30285260
	rs7432375	3:136569563	**4 x 10-12**		29483656

			3 x 10-8	Autism spectrum disorder or schizophrenia	28540026
	rs7432375-G		**7 x 10-11**	Schizophrenia	25056061
	rs10935185	3:136571687	**5 x 10-9**	Schizophrenia (MTAG)	32606422
	rs940174	3:136590868	**3 x 10-9**	Schizophrenia vs autism spectrum disorder (ordinary least squares (OLS))	33686288
	rs7618871	3:136681578	**5 x 10-10**	Anorexia nervosa, attention-deficit/hyperactivity disorder, autism spectrum disorder, bipolar disorder, major depression, obsessive-compulsive disorder, schizophrenia, or Tourette syndrome (pleiotropy)	31835028
STAG1-DT, SLC35G2	rs6789329	3:136758020	**1 x 10-8**	Attention deficit hyperactivity disorder or autism spectrum disorder or intelligence (pleiotropy)	35764056
	rs6795372-G	3:136765048	**3 x 10-10**	Depressed affect	29942085
CTNND2	rs56044142-G	5:10900282	**1 x 10-9**	Insomnia	30804565
	rs2907292	5:11084600	**4 x 10-8**	Amyotrophic lateral sclerosis (sporadic)	24529757
	rs2530215	5:11298111	**3 x 10-8**	Bipolar disorder and schizophrenia	20889312
	rs6887317-A	5:11370935	7 x 10-6	Lewy body disease	25188341
CTNND2, RNU6-679P	rs10060040-A	5:11922959	9 x 10-6	Opioid addiction	36207451
	rs11744876	5:11930025	4 x 10-6	Late-onset Alzheimer's disease	27770636
	rs9686466	5:12202315	8 x 10-6	Alzheimer's disease in non-APOE e4 carriers	35694926
TUBB	rs114441450	6:30719037	9 x 10-7	Autism spectrum disorder or schizophrenia	28540026
CHST15	rs28719480-C	10:124040129	**2 x 10-8**	Alzheimer's disease or gastroesophageal reflux disease	35851147
OAT, CHST15	rs3884528	10:124235482	9 x 10-6	Opioid addiction	36207451
NARS2	rs4474465	11:78493334	1 x 10-6	Alzheimer's disease (survival time)	25649651
YLPM1	rs10144845-C	14:74771067	**1 x 10-10**	Depressed affect	29942085
			4 x 10-13	Neuroticism	32231276
	rs10148293-G	14:74830128	1 x 10-7	Depression (broad)	29662059
SECISBP2L	rs11854184-C	15:49000997	5 x 10-7	Schizophrenia	35396580

The GWAS Catalog identifies studies through weekly PubMed searches and extracts data for single nucleotide polymorphisms (SNPs) with $p < 1 \times 10^{-5}$ in the overall (initial GWAS + replication) population. Bold denotes p values $< 5 \times 10^{-8}$. Results are shown for traits that fall under the umbrella terms ‘nervous system disease’ or ‘psychiatric disorder’.

Table S4: Genes harboring ultra-rare de novo variants in ADHD cases are enriched for pathway and gene ontology based sets

Enriched pathway-based sets					
pathway name	**set size**	**candidates contained**	**p-value**	**q-value**	**pathway source**
CXCR4-mediated signaling events	86	3 (3.6%)	0.000152	0.00414	PID
Sema3A PAK dependent Axon repulsion	16	2 (12.5%)	0.00018	0.00414	Reactome
Ectoderm Differentiation	142	3 (2.1%)	0.00071	0.00856	Wikipathways
EPHB-mediated forward signaling	35	2 (5.7%)	0.000879	0.00856	Reactome
rac1 cell motility signaling pathway	36	2 (5.6%)	0.00093	0.00856	BioCarta
RAC1 signaling pathway	54	2 (3.7%)	0.00208	0.016	PID
Semaphorin interactions	64	2 (3.1%)	0.00292	0.0192	Reactome
CDC42 signaling events	71	2 (2.8%)	0.00357	0.0198	PID
EPH-Ephrin signaling	74	2 (2.7%)	0.00388	0.0198	Reactome
Fc gamma R-mediated phagocytosis - Homo sapiens (human)	97	2 (2.1%)	0.00644	0.0296	KEGG
Enriched gene ontology-based sets					
gene ontology term	**category, level**	**set size**	**candidates contained**	**p-value**	**q-value**
GO:0044877 protein-containing complex binding	MF 2	1223	7 (0.6%)	0.000384	0.00461
GO:0000226 microtubule cytoskeleton organization	BP 3	615	5 (0.8%)	0.000653	0.0559
GO:0005856 cytoskeleton	CC 4	2323	9 (0.4%)	0.000816	0.022
GO:0007010 cytoskeleton organization	BP 4	1405	7 (0.5%)	0.000879	0.0993
GO:0040019 positive regulation of embryonic development	BP 5	42	2 (4.8%)	0.00112	0.0564
GO:0050808 synapse organization	BP 3	410	4 (1.0%)	0.00126	0.0559
GO:0050807 regulation of synapse organization	BP 5	210	3 (1.4%)	0.00188	0.0564
GO:0050803 regulation of synapse structure or activity	BP 3	220	3 (1.4%)	0.00214	0.0574
GO:0060997 dendritic spine morphogenesis	BP 5	59	2 (3.4%)	0.0022	0.0564
GO:0007017 microtubule-based process	BP 2	808	5 (0.6%)	0.00221	0.112
GO:0016358 dendrite development	BP 3	235	3 (1.3%)	0.00258	0.0574
GO:0015629 actin cytoskeleton	CC 5	513	4 (0.8%)	0.00281	0.0112
GO:0015630 microtubule cytoskeleton	CC 5	1292	6 (0.5%)	0.00321	0.0112
GO:0044430 cytoskeletal part	CC 3	1777	7 (0.4%)	0.00342	0.0478
GO:0005912 adherens junction	CC 3	557	4 (0.7%)	0.00382	0.0478
GO:0097061 dendritic spine organization	BP 4	81	2 (2.5%)	0.0041	0.169
GO:0070161 anchoring junction	CC 2	573	4 (0.7%)	0.00423	0.114
GO:0031334 positive regulation of protein complex assembly	BP 5	287	3 (1.0%)	0.00453	0.0775
GO:0008013 beta-catenin binding	MF 3	86	2 (2.3%)	0.00461	0.0737
GO:0106027 neuron projection organization	BP 5	90	2 (2.2%)	0.00503	0.0775

GO:0032587	ruffle membrane	CC 4	96	2 (2.1%)	0.0057	0.0585
GO:0060996	dendritic spine development	BP 3	97	2 (2.1%)	0.00582	0.104
GO:0005200	structural constituent of cytoskeleton	MF 2	103	2 (1.9%)	0.00654	0.0392
GO:0045296	cadherin binding	MF 4	335	3 (0.9%)	0.00695	0.0581
GO:0048858	cell projection morphogenesis	BP 4	666	4 (0.6%)	0.00719	0.169
GO:0016363	nuclear matrix	CC 4	109	2 (1.8%)	0.0073	0.0585
GO:0032990	cell part morphogenesis	BP 4	682	4 (0.6%)	0.00781	0.169
GO:0005884	actin filament	CC 5	114	2 (1.8%)	0.00795	0.0186
GO:0120036	plasma membrane bounded cell projection organization	BP 4	1556	6 (0.4%)	0.00799	0.169
GO:0030036	actin cytoskeleton organization	BP 3	691	4 (0.6%)	0.00809	0.108
GO:0030030	cell projection organization	BP 3	1594	6 (0.4%)	0.00897	0.108
GO:0000904	cell morphogenesis involved in differentiation	BP 5	724	4 (0.6%)	0.0096	0.107
GO:0034399	nuclear periphery	CC 4	0	2 (1.6%)	0.00994	0.0585

ConsensusPathDB was used to query pathways and gene-ontology. All sets with a p value <.01 are listed.

Dataset S1 (separate file). Metrics and quality control information from whole-exome DNA sequencing data. Key included in dataset.

Dataset S2 (separate file). Details of rare *de novo* variants identified in ADHD probands and controls. Key included in dataset.

Dataset S3 (separate file). Gene-based test results combining the ultra-rare *de novo* damaging variants and independent case-control data. We used the Bayesian extension of the Transmission And De novo Association test (extTADA) to examine ultra-rare de novo protein-truncating variants (PTV) and missense variants predicted to be damaging (MPC score>2, Mis-D) from the 147 ADHD parent-child trios and an independent group of 3,206 ADHD cases and 5,002 unaffected controls. We ran extTADA to calculate the Bayes factor and q-values (false discovery rate, FDR) for each gene. One gene *KDM5B* is classified as a high-confidence risk gene (FDR, false discovery rate < 0.1) and two genes, *POMT1* and *YLPM1*, are classified as probable risk genes (FDR<0.3).

SI References

1. H. T. Nguyen *et al.*, Integrated Bayesian analysis of rare exonic variants to identify risk genes for schizophrenia and neurodevelopmental disorders. *Genome medicine* **9**, 1-22 (2017).
2. X. He *et al.*, Integrated model of de novo and inherited genetic variants yields greater power to identify risk genes. *PLoS genetics* **9**, e1003671 (2013).
3. J. S. Ware, K. E. Samocha, J. Homsy, M. J. Daly, Interpreting de novo Variation in Human Disease Using denovolyzeR. *Curr Protoc Hum Genet* **87**, 7.25.21-27.25.15 (2015).