

## The copy number variant architecture of psychopathology and cognitive development in the ABCD® study

Zhiqiang Sha<sup>1,2,3</sup>, Kevin Y. Sun<sup>1,2,3</sup>, Benjamin Jung<sup>4</sup>, Ran Barzilay<sup>1,2,3</sup>, Tyler M. Moore<sup>1</sup>, Laura Almasy<sup>3,5,6</sup>, Jennifer K. Forsyth<sup>7</sup>, Smriti Prem<sup>8,9</sup>, Michael J. Gandal<sup>1,3,6</sup>, Jakob Seidlitz<sup>1,2,3</sup>, Joseph T. Glessner<sup>10\*</sup>, Aaron F. Alexander-Bloch<sup>1,2,3\*</sup>

<sup>1</sup> Department of Psychiatry, University of Pennsylvania, Philadelphia, PA, USA

<sup>2</sup> Department of Child and Adolescent Psychiatry and Behavioral Science, The Children's Hospital of Philadelphia, Philadelphia, PA, USA

<sup>3</sup> Lifespan Brain Institute, The Children's Hospital of Philadelphia and Penn Medicine, Philadelphia, PA, USA

<sup>4</sup> Section on Neurobehavioral and Clinical Research, Social and Behavioral Research Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA

<sup>5</sup> Department of Biomedical and Health Informatics, Children's Hospital of Philadelphia, Philadelphia, PA, USA

<sup>6</sup> Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

<sup>7</sup> University of Washington, Department of Psychology, WA, USA

<sup>8</sup> Department of Neuroscience and Cell Biology, Rutgers Robert Wood Johnson Medical School, Piscataway, NJ, USA

<sup>9</sup> Graduate Program in Neuroscience, Rutgers University, Piscataway, NJ, USA

<sup>10</sup> Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA, USA

\*contributed equally as senior authors

**Manuscript information:** 4739 words (Introduction, Methods, Results, Discussion and Conclusion) and 5 figures.

(Supplementary Information: 10 figures, 24 tables)

### Corresponding authors:

Zhiqiang Sha Ph.D.

Lifespan Brain Institute at University of Pennsylvania and Children's Hospital of Philadelphia  
Department of Child and Adolescent Psychiatry and Behavioral Science, The Children's Hospital of Philadelphia

Department of Psychiatry, University of Pennsylvania

E-mail: [Zhiqiang.Sha@Pennmedicine.upenn.edu](mailto:Zhiqiang.Sha@Pennmedicine.upenn.edu)

Aaron Alexander-Bloch, M.D. Ph.D.

Lifespan Brain Institute at University of Pennsylvania and Children's Hospital of Philadelphia  
Department of Child and Adolescent Psychiatry and Behavioral Science, The Children's Hospital of Philadelphia

Department of Psychiatry, University of Pennsylvania

E-mail: [Aaron.Alexander-Bloch@Pennmedicine.upenn.edu](mailto:Aaron.Alexander-Bloch@Pennmedicine.upenn.edu)

## Abstract

### Importance

Childhood is a crucial developmental phase for mental health and cognitive function, both of which are commonly affected in patients with psychiatric disorders. This neurodevelopmental trajectory is shaped by a complex interplay of genetic and environmental factors. While common genetic variants account for a large proportion of inherited genetic risk, rare genetic variations, particularly copy number variants (CNVs), play a significant role in the genetic architecture of neurodevelopmental disorders. Despite their importance, the relevance of CNVs to child psychopathology and cognitive function in the general population remains underexplored.

### Objective

Investigating CNV associations with dimensions of child psychopathology and cognitive functions.

### Design, Setting, and Participants

ABCD<sup>®</sup> study focuses on a cohort of over 11,875 youth aged 9 to 10, recruited from 21 sites in the US, aiming to investigate the role of various factors, including brain, environment, and genetic factors, in the etiology of mental and physical health from middle childhood through early adulthood. Data analysis occurred from April 2023 to April 2024.

### Main Outcomes and Measures

In this study, we utilized PennCNV and QuantiSNP algorithms to identify duplications and deletions larger than 50Kb across a cohort of 11,088 individuals from the Adolescent Brain Cognitive Development<sup>®</sup> study. CNVs meeting quality control standards were subjected to a genome-wide association scan to identify regions associated with quantitative measures of broad psychiatric symptom domains and cognitive outcomes. Additionally, a CNV risk score, reflecting the aggregated burden of genetic intolerance to inactivation and dosage sensitivity, was calculated to assess its impact on variability in overall and dimensional child psychiatric and cognitive phenotypes.

### Results

In a final sample of 8,564 individuals (mean age=9.9 years, 4,532 males) passing quality control, we identified 4,111 individuals carrying 5,760 autosomal CNVs. Our results revealed significant associations between specific CNVs and our phenotypes of interest, psychopathology and cognitive function. For instance, a duplication at 10q26.3 was associated with overall psychopathology, and somatic complaints in particular. Additionally, deletions at 1q12.1, along with duplications at 14q11.2 and 10q26.3, were linked to overall cognitive function, with particular contributions from fluid intelligence (14q11.2), working memory (10q26.3), and reading ability (14q11.2). Moreover, individuals carrying CNVs previously associated with neurodevelopmental disorders exhibited greater impairment in social functioning and cognitive performance across multiple domains, in particular working memory. Notably, a higher deletion CNV risk score was significantly correlated with increased overall psychopathology (especially in dimensions of social functioning, thought disorder, and attention) as well as cognitive impairment across various domains.

### Conclusions and Relevance

In summary, our findings shed light on the contributions of CNVs to interindividual variability in complex traits related to neurocognitive development and child psychopathology.

## **Key Points**

### **Question**

Are copy number variants (CNVs) contributing to individualized variability in psychopathology outcomes and cognitive functions in children?

### **Findings**

Both regional CNVs at 1q12.1, 10q26.3 and 14q11.2 and global CNV burden accounted for the variance in dimensions of psychopathology, particularly for attention and social problems, as well as cognitive performance, especially for working memory and reading ability.

### **Meaning**

Our findings suggest that, in addition to common genetic variants, gene dosage changes confer genetic susceptibility to child psychopathology and cognitive development.

## Introduction

Neurodevelopmental disorders are characterized by a spectrum of mental health issues, developmental delays, and deficits in diverse cognitive functions<sup>1,2</sup>. Genetic factors are known to play a crucial role in these disorders<sup>3,4</sup> such as autism spectrum disorder<sup>5,6</sup>, attention-deficit/hyperactivity disorder<sup>7,8</sup> and psychosis spectrum disorders<sup>9</sup>. Population-based genetic studies have revealed robust associations between common variants, such as single-nucleotide polymorphisms (SNPs), and the pathogenesis of these conditions<sup>10,11</sup>. The advent of genome-wide association scans has shed light on the polygenic nature of neurodevelopmental disorders. However, the contribution of SNPs with subtle genetic effects does not fully account for the variance in these complex phenotypes due to genetic factors<sup>12</sup>. In addition to common genetic variants, other types of genetic variations including copy number variants (CNVs) – a form of genomic structural rearrangement involving duplications and deletions spanning at least 1000 base pairs throughout the human genome<sup>13</sup> – are increasingly recognized as crucial components in elucidating the detailed genetic architecture underlying the etiology of neurodevelopmental disorders<sup>14</sup>.

While the large majority of CNVs are benign, clinical and genetic epidemiological investigations have consistently demonstrated that individuals harboring pathogenic CNVs often exhibit early-onset symptoms of mental illness and are more predisposed to transitioning into clinically diagnosable psychiatric disorders<sup>15</sup>. For example, approximately one quarter of individuals with the 22q11.2 deletion develop schizophrenia<sup>16,17</sup>. Large-scale case-control studies have identified a spectrum of genomic duplications and deletions associated with increased genetic susceptibility to neurodevelopmental disorders, including schizophrenia, attention-deficit/hyperactivity disorder, and autism spectrum disorder<sup>5,6,18</sup>. However, these investigations have predominantly focused on identifying genetic correlates within the confines of a categorical diagnostic framework, while it is increasingly acknowledged that psychopathological manifestations and associated cognitive deficits exist along a continuum beyond traditional diagnostic boundaries<sup>19-21</sup>. To date, there remains a dearth of research exploring the CNV architecture underlying dimensions of psychopathology and cognitive functions during childhood.

The concept of the CNV risk score has emerged to estimate the tolerance to loss-of-function mutations and dosage sensitivity of genes by deletions or duplications<sup>22</sup>. CNV risk scores aim to measure the cumulative risk from deletions or duplications of large genomic segments. Different types of genetic variants exhibit cumulative effects, contributing to overall phenotypic variability<sup>12</sup>. Recently, we observed that higher CNV risk scores were significantly correlated with poor cognitive performance and greater psychopathology in youth<sup>23</sup>. These findings underscore the potential of CNV risk scores to detect genetic associations with neurodevelopmental outcomes in childhood and adolescence.

In the present study, we conducted a comprehensive exploration of the CNV architecture underlying dimensions of psychopathology and cognitive phenotypes within the Adolescent Brain Cognitive Development (ABCD<sup>®</sup>) Study. This prospective study comprises a community sample of N=11,876 children aged 9-10 years old, who were not ascertained for mental health problems<sup>24</sup>. Our investigation proceeded in several stages. First, CNVs for each individual were identified using two algorithms, PennCNV<sup>25</sup> and QuantiSNP<sup>26</sup>, leveraging genome-wide SNP array data. Second, CNVs that overlapped between the two algorithms were utilized in a genome-wide CNV association scan, aiming to identify specific CNV regions (CNVRs) associated with psychopathology and cognition independently. Third, we examined whether individuals carrying CNVs associated with neurodevelopmental disorders<sup>22</sup> exhibited poorer cognitive performance and greater psychopathology. Finally, we explored the association of CNV risk score with subdomains of psychopathology and cognition. Cumulatively, the present study sheds

significant light on the contribution of CNVs to interindividual variability in complex traits critical to neurodevelopmental outcomes.

## Methods

### Sample and CNV quality control

We downloaded the imputed genotyped data from the ABCD<sup>®</sup> portal (Data Release 3.0). These genetic data were acquired by the Chemagen bead-based/Chemagic STAR DNA Saliva4k Kit (CMG-1755-A) at the Rutgers University Cell and DNA Repository<sup>27</sup>. Genotypes were ascertained using the Affymetrix National Institute on Drug Abuse SmokeScreen Array<sup>28</sup>, and imputed using SHAPEIT3<sup>29</sup> and impute2<sup>30</sup>.

To call CNVs using ABCD<sup>®</sup> study, we compiled log R ratio and B allele frequency profiles for each SNP marker from the intensity files of genotyping calls obtained through the Affymetrix Smokescreen array<sup>27</sup>. Multiple sample log R ratio and B allele frequency matrix files, paired with sample and probe information files were split into single sample log R ratio and B allele frequency files. Population frequency of B allele and GC base content model definition files specific to the Affymetrix SmokeScreen Array were compiled. CNV calling was performed following established procedures detailed in our previous publications (Supplementary Figure 1)<sup>23,31</sup>. The ABCD<sup>®</sup> study encompasses genotype data from 138 plates. Initially, individuals from plate 461 were excluded due to potential sequencing issues identified in the genetic data report. Additionally, individuals were excluded if their data exhibited a standard deviation of log R ratio >0.35, a standard deviation of B allele frequency >0.1, or an absolute value of wave factor >0.05. All probe coordinates were established based on the GRCh37/hg19 reference genome.

We employed hidden Markov models to call CNVs based on SNP genotype data, using both PennCNV<sup>25</sup> and QuantiSNP<sup>26</sup> algorithms. PennCNV utilizes multiple sources of information, including the log R ratio and B allele frequency at each SNP marker, the distance between neighboring SNPs, and adjusts for "genomic waves" using a regression model based on GC content<sup>25</sup>. On the other hand, QuantiSNP utilizes an objective Bayes approach, leveraging the log R ratio and B allele frequency for each SNP marker<sup>26</sup>. Both algorithms have demonstrated sensitivity in detecting deletions and duplications in the human genome across genetic studies spanning the past two decades<sup>9,32,33</sup>. To ensure identifying reliable CNVs, we used CNVision<sup>34</sup> to merge CNVs detected by two algorithms. Furthermore, we excluded individuals carrying more than 500 merged CNVs, as this threshold may indicate batch effects, genotyping errors, or extreme chromosome abnormalities. CNVs exhibiting >70% overlap between the two algorithms were retained for follow-up genetic analysis.

Subsequently, we implemented a series of additional filtering steps aimed at minimizing false discoveries. Specifically, we excluded CNVs smaller than 50,000 base pairs in size and filtered out CNVs spanning <20 SNPs or with PennCNV or QuantiSNP confidence scores <30<sup>23</sup>. CNVs demonstrating over 50% reciprocal overlap with segmental duplications, centromeric regions, telomeric regions, or the major histocompatibility complex region were excluded due to the complex genomic structure in these regions. CNVs exhibiting an equivocal number of copies between the two algorithms were also removed. Moreover, in families with multiple members, only one random family member was included in the subsequent analysis. The final sample included N=8,564 individuals (mean age=9.9 years, 4,532 boys)

### Psychopathological and cognitive measures

The ABCD<sup>®</sup> study includes comprehensive assessments of psychopathological and cognitive measures<sup>35,36</sup>. We used derived measures of dimensional psychopathology from the 119-item Child Behavioral Checklist<sup>37</sup> completed for each individual. Parents rated the extent to which specific behaviors

(e.g., "Destroys others' things") were characteristic of their child over the past 6 months, using a 3-point scale: 0 ("not true"), 1 ("somewhat or sometimes true"), or 2 ("very true or often true"). These 119 items are aggregated to generate raw composite scores for 8 scales: anxious/depressed, withdrawn/depressed, somatic complaints, social problems, thought problems, attention problems, rule-breaking behavior, and aggressive behavior. The sum of scores on these scales provided an overall measure of psychopathology burden, encompassing various behavioral syndromes.

Cognitive performance was assessed by the NIH Toolbox Neurocognitive Battery. The NIH Toolbox encompasses 8 domains<sup>38,39</sup>, including: fluid intelligence (reflecting the ability to solve abstract reasoning problems); dimensional change card sort (measuring ability to plan, organize, and execute goal-directed behaviors); flanker inhibitory control and attention (reflecting the ability to handle multiple environmental stimuli); list sorting working memory (quantifying the ability to store, manipulate, and hold new information); oral reading recognition test (capturing reading ability and academic achievement); picture vocabulary test (characterizing language and verbal intellect); picture sequence memory (representing episodic memory including the acquisition, storage, and retrieval of information); and pattern comparison processing speed (indicating the ability to process new information within a certain amount of time). Scores across these domains were aggregated to generate a composite score reflecting overall cognitive performance.

For each cognitive and psychopathological measure, we regressed the effects of age, sex, and batch by linear regression models. Subsequently, rank-based inverse normalization was applied to ensure the residuals following a normal distribution.

### **Genome-wide CNVR association analysis**

We utilized ParseCNV2<sup>40</sup> to conduct genome-wide CNVR association analyses for each cognitive and psychopathological measure independently. This method involves mapping individual-level CNV calls to population-level probe-based CNV statistics (separately for deletions and duplications). Subsequently, the significance of each SNP marker was assessed using Fisher's exact test and incorporating covariates using linear regression for quantitative (continuous) trait phenotypes. The resulting association statistics were then used to combine SNPs in proximity (default distance of 1MB) with comparable p-values (default threshold of 1 power of 10) into genomic regions referred to as "CNVRs". This approach offers several advantages over methods that investigate genetic correlates for individual CNVs, increasing power and preventing redundant reporting of overlapping genomic regions associated with the phenotype. Previous studies have successfully applied this approach to identify genomic structural changes associated with brain disorders and physical health<sup>41-43</sup>.

### **Multiple testing correction of significance**

In contrast to genome-wide association scans involving SNPs, a p-value threshold of  $5 \times 10^{-4}$  was recommended to define genome-wide significance for CNV (spanning many nucleotides) association testing after surviving multiple testing corrections. We applied a Bonferroni-corrected threshold of  $p < 2.5 \times 10^{-4}$  (i.e.,  $5 \times 10^{-4}/2$ ) to identify CNVRs contributing to the overall psychopathological and cognitive measures. For the 8 dimensions of cognitive measures, after accounting for correlations between these traits using matrix spectral decomposition, the effective number of traits was estimated to be 6. Consequently, we applied a significance level of  $p < 8.3 \times 10^{-5}$  (i.e.,  $5 \times 10^{-4}/6$ ) to correct for multiple testing. Similarly, for the 8 dimensions of psychopathology, we set a p-value threshold of  $p < 1 \times 10^{-4}$  (i.e.,  $5 \times 10^{-4}/5$ ).

### **CNV annotation and CNV risk scores**

The CNV risk score is used to estimate the pathogenic functional consequences of CNV deletions and duplications in individuals -- encompassing genetic burden, intolerance, and dosage sensitivity. CNV burden quantifies the number of genes covered by CNVs or the size of CNVs carried by an individual. Genetic intolerance refers to the annotations of genes within CNVs concerning their ability to withstand loss-of-function mutations. Dosage sensitivity characterizes the sensitivity to the deletion and duplication of CNVs.

We annotated each CNV based on two public databases<sup>44,45</sup>, scoring them in terms of genetic burden, intolerance, and dosage sensitivity (Supplementary Figure 2). Genetic burden was quantified by the size of the CNV and the number of overlapping genes within the CNV. Intolerance metrics included the probability of loss intolerance (pLI)<sup>46</sup> and the loss of function observed/expected upper bound fraction (LOEUF). Specifically, pLI represents the likelihood of a gene being intolerant to loss-of-function mutations, ranging from 0 to 1. A higher pLI score indicates less tolerance to inactivation, implying potentially deleterious functional outcomes. LOEUF compares the observed and expected number of loss-of-function mutations for a gene in a reference population, with values ranging from 0 to 2. Genes with a LOEUF value less than 0.35 are typically considered intolerant. For CNVs, the cumulative inverse LOEUF (iLOEUF) of all encompassed genes is computed. Dosage sensitivity is measured by the probability of haploinsufficiency (pHI) and the probability of triplosensitivity (pTS). These metrics reflect the likelihood of a gene being sensitive to copy number gain or loss, respectively. Thus, the pHI of a deletion is computed as the sum of the pHI values of all genes overlapping with the deletion, while the pTS of a duplication is estimated by the sum of the pTS values of all genes within the duplication. At the individual level, each component of the CNV risk score is calculated as the sum of all deletions and duplications, respectively.

We used a linear regression model to investigate the association of each CNV risk score with individual cognitive and psychopathological measures across 8564 individuals (the size of the unrelated sample after quality control), while controlling for the effects of age, sex, and array batch. Prior to modeling, rank-based inverse normalization was conducted on each behavioral measure to ensure adherence to a normal distribution. To correct for multiple testing, false discovery rate (FDR) adjustment at a significance level of  $p < 0.05$  was applied separately for cognitive measures and psychopathological measures.

## Results

### The CNV landscape of the ABCD® study

Following conservative quality control (Supplementary Figure 1), a total of 5760 CNVs were identified across 4111 individuals (Figure 1) while 4453 individuals did not carry any CNVs. Among the detected CNVs, 1391 (24.15%) were deletions (43 homozygous), while 4369 (75.85%) were duplications (161 with copy number >3). While 28.8% of individuals with a CNV carried a single deletion, instances of up to three deletions were observed in nine individuals. Similarly, 61.5% of individuals had one duplication (totaling 2527 out of 4111), with up to four duplications found in 18 individuals (Supplementary Figure 3). Summary statistics at the CNV level were as follows (Supplementary Figure 3): number of SNPs encompassed by deletions, mean=78 (range 20-1811); number of SNPs encompassed by duplications, mean=151 (20-12,635); deletion length, mean=300.99Kb (50Kb-6.14Mb); duplication length, mean=357.76Kb (50Kb-11.10Mb); number of genes within deletions, mean=1 (0-46); number of genes within duplication, mean=2 (0-51).



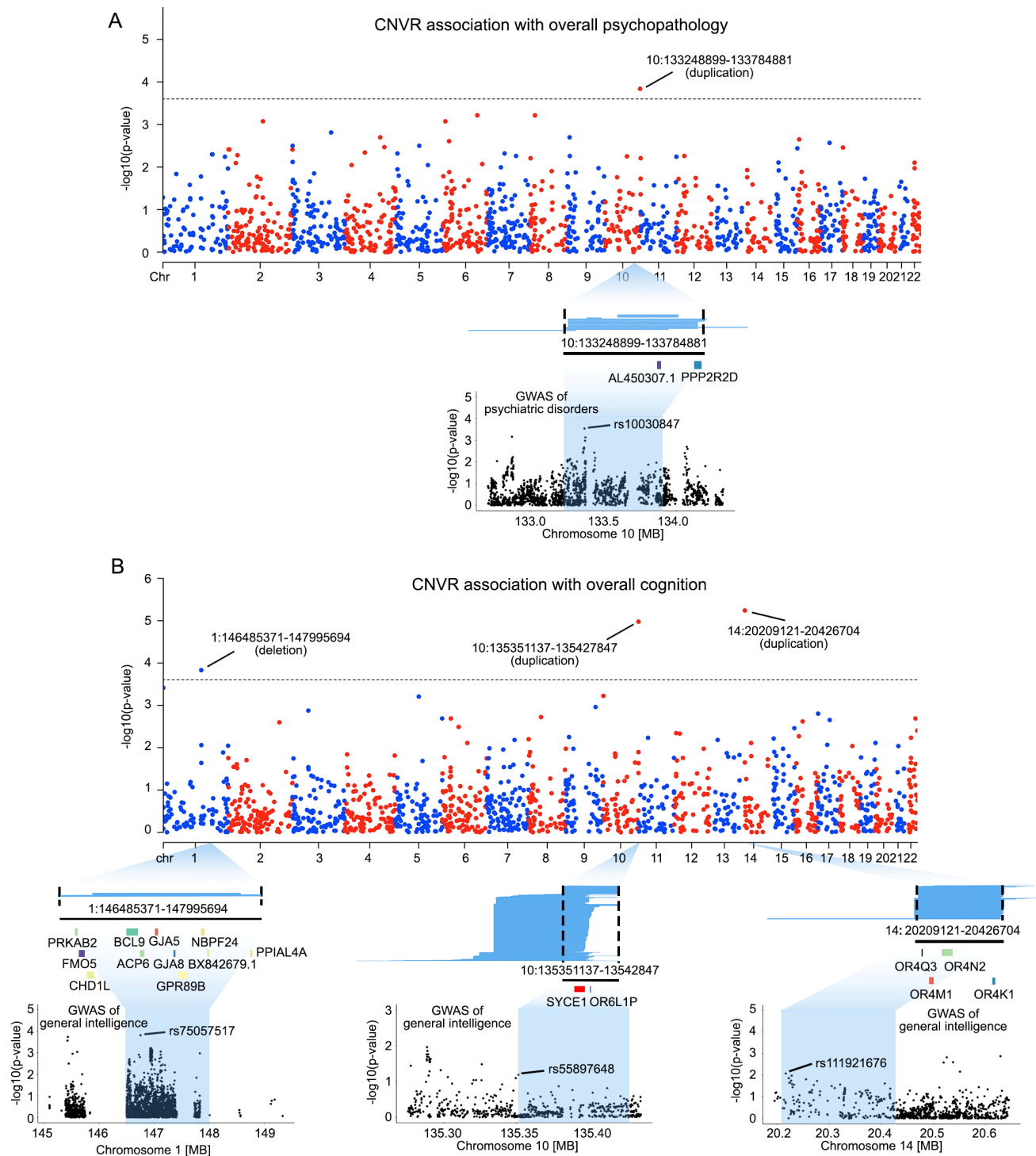
**Figure 1. The landscape of the CNV calls in the ABCD<sup>®</sup> study.** After conservative quality control, we report 5760 CNVs among 8564 individuals from the ABCD<sup>®</sup> study. The figure illustrates the chromosomal distribution of all deletions and duplications called in the ABCD<sup>®</sup> study, in which blue represents deletions, and red represents duplications.

### Genome-wide CNVR association with overall psychopathology and cognition

We used ParseCNV2<sup>40</sup> to identify CNVRs associated with variations in overall psychopathology and cognitive function, including one CNVR associated with overall psychopathology (Figure 2A) and three CNVRs associated with cognition (Figure 2B). Specifically, a duplication at 10q26.3 (i.e.,



chr10:133248899-133784881) was significantly associated with overall psychopathology ( $p=1.45 \times 10^{-4}$ , Figure 2A and Supplementary Table 1). Upon CNV decomposition, we identified 16 individuals carrying duplications that contributed to this significant CNVR association signal. Notably, two genes overlap with this genomic region, including *PPP2R2D* and the long non-coding RNA *AL450307.1* (Figure 2A). Annotation revealed that common variants within *PPP2R2D* have been linked to schizophrenia<sup>47</sup>, and methylation modifications of the genomic region have been implicated in dimensional psychopathology in youth<sup>48</sup>. In a GWAS of 12 psychiatric disorders published by Romero et al.<sup>49</sup>, the strongest association signal was for rs10030847 ( $p=0.0003$ , Figure 2A).



**Figure 2. Genome-wide CNVRs associations with overall psychopathology and cognitive function.** (A) Genome-wide CNVR association with overall psychopathology. Manhattan plot illustrates significant CNVR associations with the overall psychopathology measure at Bonferroni-corrected  $p < 2.5 \times 10^{-4}$  (i.e.,  $5 \times 10^{-4}/2$ ), the recommended threshold for genome-wide significance in the ParseCNV2 platform. Blue lines under the Manhattan plot represent the 16 duplications contributing to the significant CNVR association. The black line and the genes below them indicate the genomic position of genes

encompassed by the CNVR significantly associated with psychopathology. The Manhattan plot at the bottom illustrates the GWAS association of psychiatric disorders with common genetic variants (i.e., SNPs) within the identified CNVR, as reported by Romero et al<sup>49</sup>. (B) Genome-wide CNVR association with global cognitive function. Three CNVRs were found significantly associated with the overall cognitive function at Bonferroni-corrected  $p < 2.5 \times 10^{-4}$  (i.e.,  $5 \times 10^{-4}/2$ ). Blue lines under the Manhattan plot represent the CNVs that contribute to each significant CNVR association. The black lines and the genes below them indicate the genomic position of genes encompassed by the CNVR significantly associated with global cognitive function. The Manhattan plot at the bottom illustrates GWAS SNP associations with general intelligence within each CNVR, as reported by Savage et al.<sup>50</sup>.

Three CNVRs were significantly associated with general cognition, including chr1:146485371-147995694 deletion at 1q21.1 ( $p = 1.48 \times 10^{-4}$ ), chr10:135351137-135427847 duplication at 10q26.3 ( $p = 1.05 \times 10^{-5}$ ), and 14:20209121-20426704 duplication at 14q11.2 ( $p = 5.75 \times 10^{-6}$ , Figure 2B and Supplementary Table 2).

For the deletion at chr1:146485371-147995694, we identified 2 individuals carrying duplications that were associated with this CNVR. Notably, both individuals exhibited poor performance across multiple cognitive domains, particularly in cognitive flexibility, working memory, and language-related functions (Supplementary Figure 4). Regional log R ratio and B allele frequency plots confirmed that copy number loss was present at 1q22.1 in both individuals (Supplementary Figure 5). This CNVR encompasses several genes important for cognitive development, such as *PRKAB2*, *GJA8* and *CHD1L*<sup>51-53</sup> (Figure 2B). Of note, 1q21.1 is a known pathogenic CNV associated with neurodevelopmental disorders and intellectual disability<sup>22,54,55</sup>, providing a validation of the methodological approach. In a GWAS of general intelligence<sup>50</sup>, the top SNP in this CNVR was rs75057517 ( $p = 0.0001$ , Figure 2B), indicating that common variants in the CNVR may also impact general intelligence.

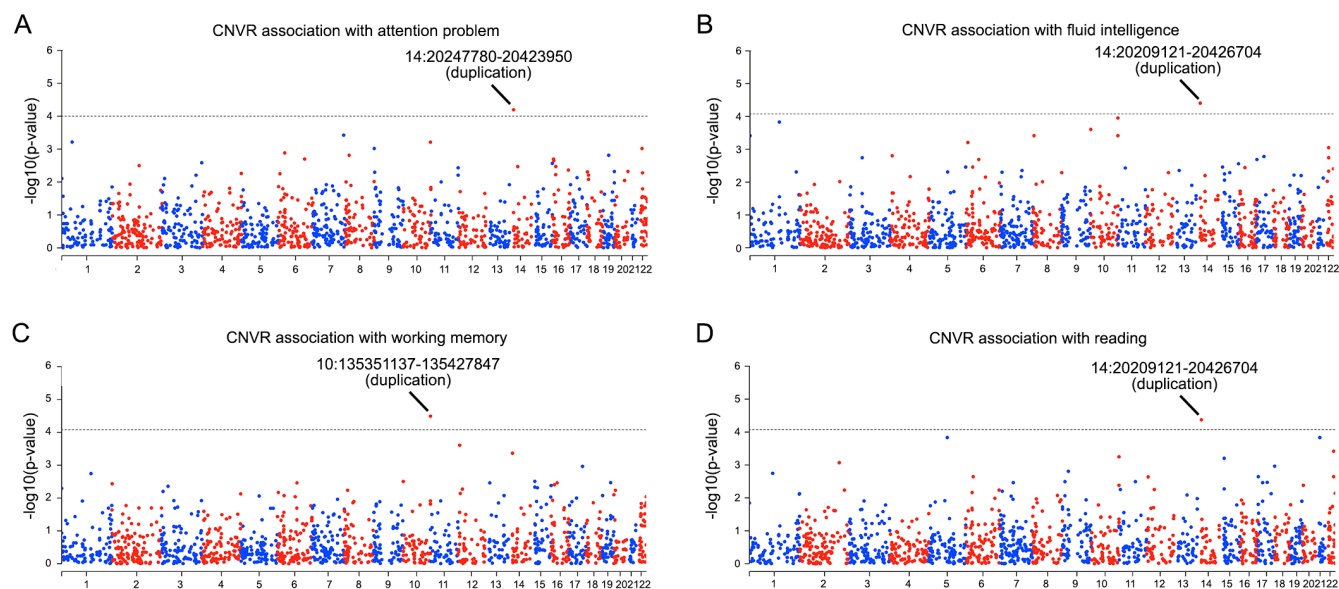
For the duplication at chr10:135351137-135427847, we identified 717 individuals carrying copy number gains associate with this CNVR (Figure 2B). This region encompasses two protein-coding genes, *SYCE1* and *OR6L1P*. Notably, duplications in this CNVR have been previously reported to be a contributing factor for intellectual disability<sup>56</sup>. The top SNP in the CNVR in the general intelligence GWAS was rs55897648 ( $p = 0.06$ , Figure 2B).

For the association of duplication at 14:20209121-20426704, we observed 350 individuals with duplications associated with this CNVR located at 14q11.2. This genomic region contains four olfactory receptor family genes, including *OR4Q3*, *OR4N2*, *OR4M1* and *OR4K1* (Figure 2B). Variations in the copy number of olfactory genes have been identified as significantly associated with cognitive decline in patients with Alzheimer's disease<sup>57</sup>. The top SNP in the CNVR in the general intelligence GWAS was rs111921676 ( $p = 0.008$ , Figure 2B).

### **Genome-wide CNVR association with dimensions of psychopathology and cognition**

Across 8 dimensions of psychopathology, we found that the 14q11.2dup CNVR was significantly associated with attention problems ( $p = 6.39 \times 10^{-5}$ , Figure 3A, Supplementary Figure 6 and Supplementary Tables 3-10) and also had a marginal association with somatic complaints ( $p = 1.45 \times 10^{-4}$ , Figure 3A and Supplementary Table 5). Across 8 domains of cognitive function, the 14q11.2dup CNVR was also significantly associated with fluid intelligence ( $p = 3.91 \times 10^{-5}$ , Figure 3B and Supplementary Table 11) and reading ( $p = 4.24 \times 10^{-5}$ , Figure 3C and Supplementary Table 15). In addition, the CNVR at

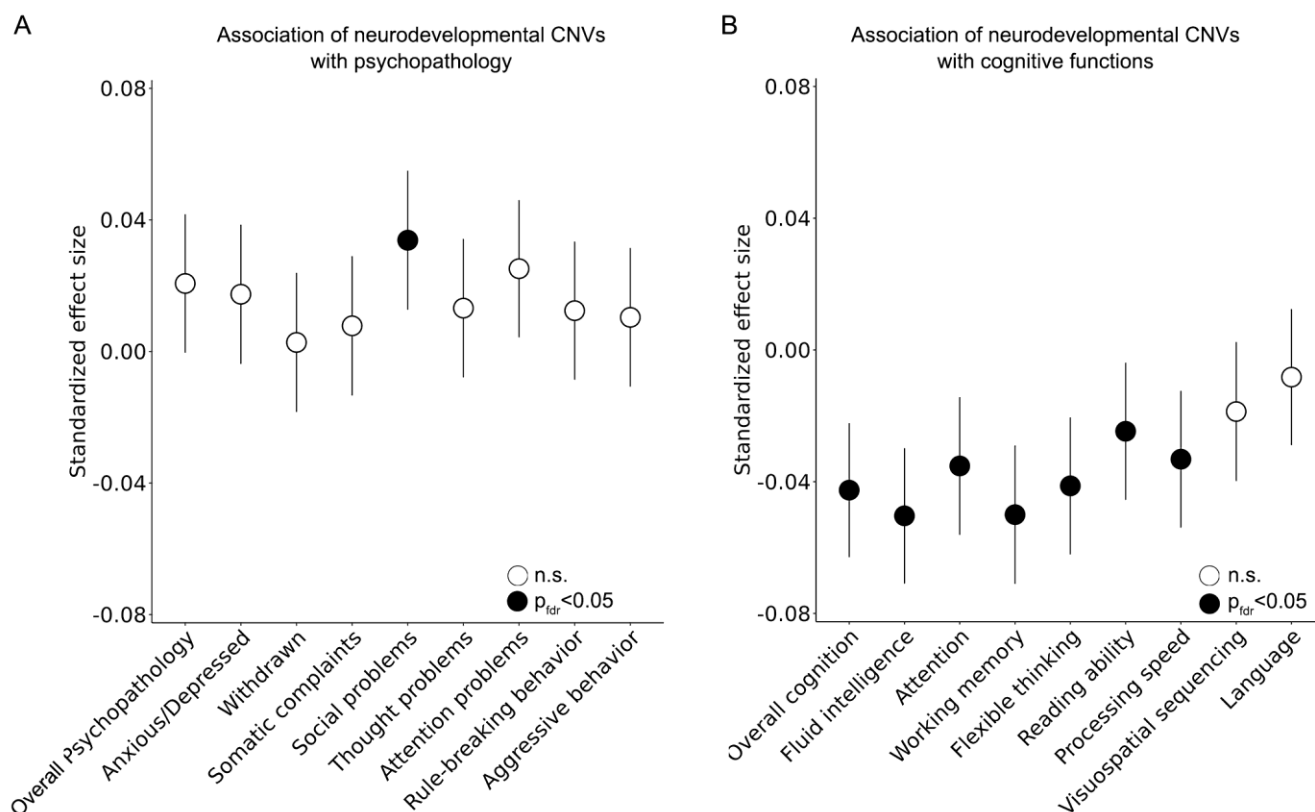
chr10:135351137-135427847 was significantly associated with working memory ( $p=3.21 \times 10^{-5}$ , Figure 3D and Supplementary Table 13). For further details, see Supplementary Figure 7 and Supplementary Tables 11-18.



**Figure 3. Genome-wide CNVRs associations with dimensions of psychopathology and cognitive function.** A) Across 8 dimensions of psychopathology, we found duplication of 14:20247780-20423950 at 14q11.2 was significantly associated with attention problems at Bonferroni-corrected  $p < 1 \times 10^{-4}$  (i.e.,  $5 \times 10^{-4}$  divided by 5, which is the effective number of tests across 8 dimensions of psychopathology). B-D) Across 8 domains of cognitive functions, we found associations between CNVRs and fluid intelligence (B), working memory (C) and reading ability (D) after multiple testing correction ( $p < 5 \times 10^{-4}$  divided by 6, which is the effective number of tests across 8 dimensions of cognitive function). Dashed line indicates the threshold of genome-wide significance for each phenotype.

### Association of neurodevelopmental CNVs with psychopathology and cognition

To complement the genome-wide scan, we specifically looked for the presence of 48 deletions and 36 duplications that have previously been identified as recurrent CNVs associated with genetic risk for neurodevelopmental disorders (Supplementary Table 19 and Supplementary Figure 8)<sup>22</sup>. Among the 4111 individuals from the ABCD<sup>®</sup> study, we identified 20 deletions among 30 ABCD<sup>®</sup> individuals who showed at least 50% reciprocal overlap<sup>58</sup> with 9 out of 48 neurodevelopmental deletions (Supplementary Table 19). 64 individuals carried 22 duplications showing at least 50% reciprocal overlap with 10 of 36 neurodevelopmental duplications (Supplementary Table 19). Of note, 59 out of 93 individuals carrying neurodevelopmental CNVs had either a deletion at 15q11.2, or a duplication at 15q11.2 or 16p13.11 (Supplementary Table 19). Next, we separately examined whether individuals carrying neurodevelopmental CNVs predicted each psychopathological and cognitive phenotype in the ABCD<sup>®</sup> study (see Methods). Our findings revealed that individuals with neurodevelopmental CNVs had high scores in the social domain of psychopathology (FDR-corrected  $p=1.55 \times 10^{-2}$ , Figure 4A, Supplementary Figure 9 and Supplementary Table 20). Moreover, these individuals showed poorer cognitive performance across various domains, including fluid intelligence ( $p_{\text{FDR}}=1.47 \times 10^{-5}$ ), attention ( $p_{\text{FDR}}=1.75 \times 10^{-3}$ ), working memory ( $p_{\text{FDR}}=1.47 \times 10^{-5}$ ), flexible thinking ( $p_{\text{FDR}}=2.35 \times 10^{-4}$ ), reading ability ( $p_{\text{FDR}}=2.59 \times 10^{-2}$ ) and processing speed ( $p_{\text{FDR}}=2.65 \times 10^{-3}$ , Figure 4B, Supplementary Figure 9 and Supplementary Table 20).



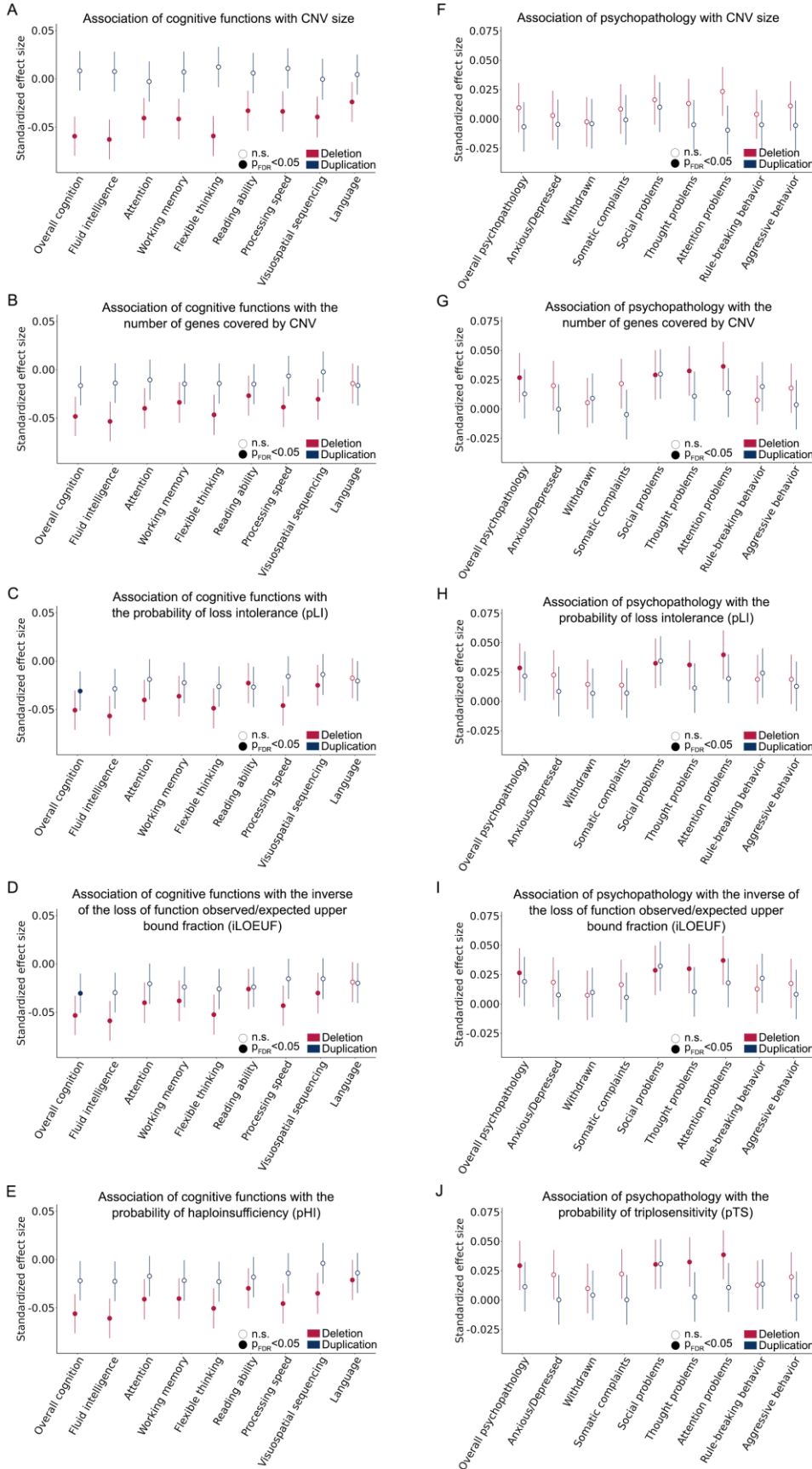
**Figure 4. Associations of psychopathology and cognitive function with individuals carrying known pathogenic recurrent CNVs associated with neurodevelopmental disorders.** (A) Association of overall and dimensional psychopathology with neurodevelopmental CNVs. Individuals with neurodevelopmental CNVs had significantly elevated psychopathology in the social domain at FDR-corrected  $p < 0.05$ . (B) Association of global and dimensional cognitive function with neurodevelopmental CNVs. Individuals with neurodevelopmental CNVs had lower cognitive performance across domains including fluid intelligence, attention, working memory, flexible thinking, reading ability and processing speed ( $p_{FDR} < 0.05$ ). Points in the dot plots indicate the value of a given predictor variable's standardized effect size and error bars indicate 95% CIs for models of cognition and psychopathology.

### CNV risk score association with cognition and psychopathology

At the individual level, we quantified the burden of deletions based on five metrics: the size of deletions (mean=48.89kb [0-6.14Mb]), the number of genes covered (mean=0.13 [0-46]), pLI (mean=0.02 [0-12.77]), iLOEUF (mean=0.17 [0-77.89]) and pHI (mean=0.05 [0-24.27], see Methods and Supplementary Figure 10). For duplications, the CNV risk score at the participant level was determined by the size of duplications (mean=182.51Kb [0-11.10Mb]), the number of genes encompassed (mean=1.26 [0-51]), pLI (mean=0.11 [ranging 0-14.42]), iLOEUF (mean=1.23 [0-86.46]) and pTS (mean=0.295 [0-26.77]). Note that individuals who have no CNVs have CNV risk scores of 0.

To assess the cumulative effect of CNVs on psychopathology and cognition, regression models were performed with each CNV risk score across 8564 individuals, for global and dimensional psychopathology and cognition measures. For deletions, we observed that all domains of the CNV risk score were

significantly associated with global cognition and all cognitive subdomains with the exception of language processing ( $p_{FDR}<0.05$ , Figure 5 and Supplementary Tables 21 and 22; significant subdomains included fluid intelligence, attention, working memory, flexible thinking, reading ability, processing speed, visuospatial sequencing). Deletion CNV risk scores with the exception of CNV size were also significantly associated with overall psychopathology as well as the social, thought, and attention domains ( $p_{FDR}<0.05$ , Figure 5 and Supplementary Tables 23 and 24). For duplications, iLOEUF and pLI risk scores were significantly associated with decreased overall cognition. Duplication risk scores were not significantly associated with domains of psychopathology after correction for multiple comparisons.



**Figure 5. Associations of psychopathology and cognitive function with CNV risk scores.** To determine the cumulative effect of CNVs on psychopathology and cognition, we used regression models to test associations with global and dimensional cognitive functions for each of 5 CNV risk scores across 8564 individuals, including the size of CNV (A), the number of genes overlapping with the CNV (B), the probability of loss intolerance (C), the inverse of the loss of function observed/expected upper bound fraction (D), dosage sensitivity (E). Similarly, we also tested associations with global and dimensions of psychopathology for each of the 5 CNV risk scores (F-J). Points in the dot plots indicate the value of a given predictor variable's effect size and error bars indicate 95% CIs for models of cognition and psychopathology.

## Discussion

The present study delineates the landscape of CNVs in the ABCD<sup>®</sup> study and characterizes CNVRs associated with child psychopathology and cognitive function. Individuals in the ABCD<sup>®</sup> study carrying recurrent CNVs previously implicated in neurodevelopmental disorders exhibited greater psychopathology in the social domain and lower cognitive performance particularly in the domain of working memory. Aggregated CNV risk scores for deletions were associated with social, thought, and attention domains of psychopathology, as well as with nearly all cognitive domains. Additionally, duplication CNV risk scores derived from constraint scores (pLI and iLOEUF) were associated with lower overall cognitive function. These results underscore the significant impact of genomic structural changes on psychopathology and cognitive development in childhood. Furthermore, our study highlights the sensitivity of CNV risk scores as a measure to characterize the complex genetic underpinnings of neurodevelopmental phenotypes.

Previous genetic studies have traditionally focused on identifying genomic structural variants associated with the risk of categorical psychiatric disorders using case-control designs<sup>59,60</sup>. However, case-control designs may not adequately capture the complexity of psychiatric symptoms characterized by high levels of comorbidity and ambiguous diagnostic boundaries<sup>19,21,61</sup>. In our current study, we sought to characterize CNVRs associated with dimensional psychopathology and cognitive outcomes in childhood. This dimensional approach allowed us to explore how gene dosage imbalance affects dimensional psychopathology in community population that was not ascertained for mental health problems. As compared to pathogenic CNVs discovered in case-control studies, CNVs with a higher frequency in the general population are more likely to be inherited, and may modulate psychopathology and cognitive development without necessarily leading to clinically diagnosed mental illness or intellectual disability.

The colocalization of SNP and CNV association signals revealed weak evidence of common genetic variants (from previous GWAS of psychiatric disorders and cognition) within CNVRs. In the context of previous evidence of enriched GWAS signal within pathogenic CNVs<sup>31</sup>, this finding has several potential explanations. First, it is possible that SNPs tag a genetic haplotype with an inherited CNV, such that the GWAS SNP effect is capturing the effect of this CNV. Second, common genetic polymorphisms with subtle effects, acting as genetic modifiers of complex phenotypes, could act cumulatively with CNVs and environmental factors to modulate variability in psychopathology and cognitive ability during childhood. Third, both CNVs and SNPs may influence phenotypes through regulatory effects on nonadjacent genomic regions, which complicates the interpretation of simple overlap between CNVRs and GWAS summary statistics. Fourth, the psychiatric GWAS data we referenced originated from a meta-analytic study pooling results across 12 psychiatric disorders defined by a traditional dichotomous diagnosis framework<sup>49</sup>. Consequently, the genetic liability for adult psychiatric disorders may not fully reflect the



genetic architecture underlying subclinical psychiatric symptoms during adolescence. Similarly for the intelligence GWAS, the results were derived from previous genome-wide meta-analytic studies evaluating overall general intelligence in adults<sup>50</sup>. However, cognitive abilities undergo significant developmental changes during childhood and adolescence, characterized by the rapid growth of neural circuits underlying various cognitive processes<sup>36,62</sup>. Consequently, the genetic basis of cognitive function in childhood may diverge from that in adulthood, such that existing GWAS fail to fully capture the complexity and nuances of cognitive function during adulthood.

In our investigation of CNVRs associated with psychopathology measures derived from the Child Behavioral Checklist, we identified 2 CNVRs that met conservative genome-wide threshold of statistical significance. A copy number gain at 10q26.3 was associated with higher overall psychopathology, and a copy number gain at 14q11.2 was associated specifically with the attention domain. Of note the CNVR on chromosome 10 associated with overall psychopathology did not meet our threshold for statistical significance for any specific domain, although a marginal association was observed with the somatic domain. The fact that relatively few CNVRs were identified illustrates that ABCD<sup>®</sup>, despite being a large study by the standards of developmental neuroscience, is relatively underpowered to fully characterize the complex architecture of CNV associations. A related challenge is that the Child Behavioral Checklist, which is based on parent report of items on a 3 point scale, provides a relatively limited depth of phenotypic characterization<sup>37</sup>. More sensitive phenotypic characterization by aggregating measures across future ABCD<sup>®</sup> time points may yield increased power to identify CNVRs associated with developmental psychopathology.

In our investigation of CNVRs associated with cognitive measures, we identified 3 CNVRs that met genome-wide significance after correction for multiple comparisons, which partially overlapped with the CNVRs associated with psychopathology. The CNVR chr14:20247780-20423950 associated with the attention domain of psychopathology largely overlapped with chr14:20209121-20426704, which was associated with global cognition as well as fluid intelligence and reading domains (reciprocal overlap: 80.97%). Furthermore, the genes encompassed by both regions were fully overlapping. One plausible explanation for this observation is that the variant encompasses pleiotropic genes that influence cognition and psychopathology, which are known to be inter-related<sup>63</sup>. On chromosome 10, we found psychopathological and cognitive measures were significantly associated with duplications of two CNVRs at 10q26.3, which encompasses *PPP2R2D* and *SYCE1*. Genetic variants and modifications of *PPP2R2D* have been reported to impact psychopathology and neural development<sup>47,48,64</sup>, and *SYCE1* variants have been associated with Alzheimer's Disease<sup>65</sup>. In addition, we recovered a genome-wide significant association between lower general cognitive function and a deletion at chr1:146485371-147995694. Fully overlapping with a known genomic hotspot at 1q12.1, a susceptibility locus for neurodevelopmental disorders<sup>55,66</sup>, the recovery of this CNVR provides important validation for the ParseCNV2 methodology. Individuals with 1q12.1 deletion are known to exhibit a range of physical and mental disorders across multiple scales, including congenital heart abnormality<sup>67</sup>, ligamentous laxity or joint hypermobility<sup>68</sup>, intellectual disability<sup>54</sup>, microcephaly<sup>55</sup>, developmental delay<sup>55</sup>, and cognitive deficits<sup>54</sup>.

Extending our analysis to other known recurrent, pathogenic CNVs associated with neurodevelopmental disorders<sup>22</sup>, we found that individuals carrying neurodevelopmental CNVs exhibited lower performance across almost all cognitive domains, a finding that is consistent with recent adult studies<sup>69</sup>. However, with the exception of deficits in the social domain, individuals carrying neurodevelopmental CNVs did not have significant differences in measures of psychopathology (Figure 4). The lack of more consistent effects on psychopathology may be indicative of heterogeneity in psychiatric presentation of neurodevelopmental

CNVs, as compared to relative homogeneity of the cognitive impact across domains. It is also possible that the cognitive impact is more apparent during childhood (or more readily measured with structured batteries). Some individuals with current cognitive deficits may eventually progress to clinically diagnosed psychiatric disorders, as additional psychiatric manifestations may unfold on a developmental timeline across adolescence and adulthood, depending on interactions with environmental and other genetic factors (e.g., risk due to common genetic variants).

Compared to identifying specific CNVRs or querying known recurrent CNVs, CNV risk scores offer a complementary approach based on the expected functional consequences aggregated across CNVs, particularly in terms of encompassed genes' tolerance to loss-of-function mutations and dosage sensitivity<sup>22,23</sup>. In our analysis of the association between CNV risk scores and psychopathology, we observed that deletion risk scores were associated with variations in overall psychopathology as well as social, thought, and attention domains in particular. However, similar associations were not observed with any duplication risk scores, despite duplications at 10q26.3 and 14q11.2 being associated with psychiatric symptoms in the genome-wide CNVR association analysis. One potential explanation for this discrepancy is that the effect sizes of duplications is generally lower and also more heterogenous compared to the effect sizes of deletions<sup>70,71</sup>. Interestingly, unlike other components<sup>70</sup> of CNV risk scores, we found that the size of deletions did not significantly correlate with any psychopathological dimensions. This underscores the potential for improved CNV risk scores, possibly attuned to specific phenotypic associations, to be explored in future investigations.

Several methodological limitations should be noted in the present study. First, our CNV association results were constrained by the limited sample size. Small samples are more susceptible to false discoveries in genetic analysis<sup>72</sup>, particularly with rare genetic variants that may occur only once in a small population. However, given our strict quality control and the fact that the CNVR with the smallest sample size recovered a known pathogenic CNV, false discovery due to low sample size is unlikely to be an issue for the present study. More relevantly, small samples may overlook genetic loci influenced by the nuanced variability of phenotypes among individuals<sup>73</sup>. It is essential to collect larger child cohorts to replicate the current genetic findings, explore additional genomic structural variants related to cross-diagnostic psychiatric symptoms, and investigate how genetic events interact with environmental factors to modulate mental and cognitive development. Second, a technical limitation concerns the genotype platform used for inferring CNVs. Although this approach is generally considered reliable, it lacks SNP probes for detecting CNVs in repeat-enriched regions due to the complex properties of genomic architecture<sup>32,74</sup>. Employing advanced whole-genome sequencing platforms could enhance the detection of high-fidelity CNVs<sup>75,76</sup>. However, genotype array platforms remain the most cost-efficient option for calling reliable CNVs at the current stage. Third, while ABCD<sup>®</sup> used epidemiological sample strategies, the sample is not truly "population representative" and is likely to skew towards higher-functioning children. Consequently, children with severely affected CNVs were less likely to participate in the ABCD<sup>®</sup> study, leading to a decrease in power to detect CNVRs in the present study (e.g., compared to a clinically-ascertained sample) and an underrepresentation of the contribution of CNVs to neurodevelopmental phenotypes.

Cumulatively, the present results represent a significant advance in our understanding of CNVs in the ABCD<sup>®</sup> study. In sharing detected CNVs and analysis code for use by other researchers, this resource will be valuable in future work aimed at understanding CNVs' impact on interindividual variability in complex traits including neurocognitive development and child psychopathology.

## **Acknowledgements**

The research was funded by R01MH132934 and R01MH133843. Data used in the preparation of this article were obtained from the Adolescent Brain Cognitive Development<sup>SM</sup> (ABCD<sup>®</sup>) Study (<https://abcdstudy.org>), held in the NIMH Data Archive (NDA). The ABCD Study<sup>®</sup> is supported by the National Institutes of Health and additional federal partners under award numbers U01DA041048, U01DA050989, U01DA051016, U01DA041022, U01DA051018, U01DA051037, U01DA050987, U01DA041174, U01DA041106, U01DA041117, U01DA041028, U01DA041134, U01DA050988, U01DA051039, U01DA041156, U01DA041025, U01DA041120, U01DA051038, U01DA041148, U01DA041093, U01DA041089, U24DA041123, U24DA041147. A full list of supporters is available at <https://abcdstudy.org/federal-partners.html>. A listing of participating sites and a complete listing of the study investigators can be found at [https://abcdstudy.org/consortium\\_members/](https://abcdstudy.org/consortium_members/). ABCD<sup>®</sup> consortium investigators designed and implemented the study and/or provided data but did not necessarily participate in the analysis or writing of this report. This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH or ABCD consortium investigators.

## **Conflict of interests**

AFA-B receives consulting income from Octave Bioscience. AFA-B and JS hold equity in and serve on the board of Centile Bioscience.

## **Data Availability**

CNV calls will be made available on NDAR (<https://nda.nih.gov/study.html?id=2589>), and code to derive CNVs and perform quality control is publicly available on Github (<https://github.com/BGDlab>).

## References

1. Thapar A, Cooper M, Rutter M. Neurodevelopmental disorders. *Lancet Psychiatry*. Apr 2017;4(4):339-346. doi:10.1016/S2215-0366(16)30376-5
2. Bolte S, Neufeld J, Marschik PB, Williams ZJ, Gallagher L, Lai MC. Sex and gender in neurodevelopmental conditions. *Nat Rev Neurol*. Mar 2023;19(3):136-159. doi:10.1038/s41582-023-00774-6
3. Parenti I, Rabaneda LG, Schoen H, Novarino G. Neurodevelopmental Disorders: From Genetics to Functional Pathways. *Trends Neurosci*. Aug 2020;43(8):608-621. doi:10.1016/j.tins.2020.05.004
4. Takumi T, Tamada K. CNV biology in neurodevelopmental disorders. *Curr Opin Neurobiol*. Feb 2018;48:183-192. doi:10.1016/j.conb.2017.12.004
5. Glessner JT, Wang K, Cai G, et al. Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature*. May 28 2009;459(7246):569-73. doi:10.1038/nature07953
6. Pinto D, Pagnamenta AT, Klei L, et al. Functional impact of global rare copy number variation in autism spectrum disorders. *Nature*. Jul 15 2010;466(7304):368-72. doi:10.1038/nature09146
7. Lionel AC, Crosbie J, Barbosa N, et al. Rare copy number variation discovery and cross-disorder comparisons identify risk genes for ADHD. *Sci Transl Med*. Aug 10 2011;3(95):95ra75. doi:10.1126/scitranslmed.3002464
8. Harich B, van der Voet M, Klein M, et al. From Rare Copy Number Variants to Biological Processes in ADHD. *Am J Psychiatry*. Sep 1 2020;177(9):855-866. doi:10.1176/appi.ajp.2020.19090923
9. Marshall CR, Howrigan DP, Merico D, et al. Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nat Genet*. Jan 2017;49(1):27-35. doi:10.1038/ng.3725
10. Niemi MEK, Martin HC, Rice DL, et al. Common genetic variants contribute to risk of rare severe neurodevelopmental disorders. *Nature*. Oct 2018;562(7726):268-271. doi:10.1038/s41586-018-0566-4
11. Cortese S, Solmi M, Michelini G, et al. Candidate diagnostic biomarkers for neurodevelopmental disorders in children and adolescents: a systematic review. *World Psychiatry*. Feb 2023;22(1):129-149. doi:10.1002/wps.21037
12. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature*. Oct 8 2009;461(7265):747-53. doi:10.1038/nature08494
13. Conrad DF, Pinto D, Redon R, et al. Origins and functional impact of copy number variation in the human genome. *Nature*. Apr 1 2010;464(7289):704-12. doi:10.1038/nature08516
14. Merikangas AK, Corvin AP, Gallagher L. Copy-number variants in neurodevelopmental disorders: promises and challenges. *Trends Genet*. Dec 2009;25(12):536-44. doi:10.1016/j.tig.2009.10.006
15. Stefansson H, Meyer-Lindenberg A, Steinberg S, et al. CNVs conferring risk of autism or schizophrenia affect cognition in controls. *Nature*. Jan 16 2014;505(7483):361-6. doi:10.1038/nature12818
16. Bassett AS, Chow EW. Schizophrenia and 22q11.2 deletion syndrome. *Curr Psychiatry Rep*. Apr 2008;10(2):148-57. doi:10.1007/s11920-008-0026-1
17. Gur RE, Roalf DR, Alexander-Bloch A, McDonald-McGinn DM, Gur RC. Pathways to understanding psychosis through rare-22q11.2DS-and common variants. *Current opinion in genetics & development*. 2021;68:35-40.
18. Jung B, Ahn K, Justice C, et al. Rare copy number variants in males and females with childhood attention-deficit/hyperactivity disorder. *Mol Psychiatry*. Mar 2023;28(3):1240-1247. doi:10.1038/s41380-022-01906-y
19. Conway CC, Forbes MK, Forbush KT, et al. A hierarchical taxonomy of psychopathology can transform mental health research. *Perspectives on psychological science*. 2019;14(3):419-436.
20. Caspi A, Moffitt TE. All for one and one for all: Mental disorders in one dimension. *American Journal of Psychiatry*. 2018;175(9):831-844.
21. Zald DH, Lahey BB. Implications of the Hierarchical Structure of Psychopathology for Psychiatric Neuroimaging. *Biol Psychiatry Cogn Neurosci Neuroimaging*. May 2017;2(4):310-317. doi:10.1016/j.bpsc.2017.02.003

22. Brownstein CA, Douard E, Mollon J, et al. Similar Rates of Deleterious Copy Number Variants in Early-Onset Psychosis and Autism Spectrum Disorder. *Am J Psychiatry*. Nov 1 2022;179(11):853-861. doi:10.1176/appi.ajp.21111175
23. Alexander-Bloch A, Huguet G, Schultz LM, et al. Copy Number Variant Risk Scores Associated With Cognition, Psychopathology, and Brain Structure in Youths in the Philadelphia Neurodevelopmental Cohort. *JAMA Psychiatry*. Jul 1 2022;79(7):699-709. doi:10.1001/jamapsychiatry.2022.1017
24. Volkow ND, Koob GF, Croyle RT, et al. The conception of the ABCD study: From substance use to a broad NIH collaboration. *Dev Cogn Neurosci*. Aug 2018;32:4-7. doi:10.1016/j.dcn.2017.10.002
25. Wang K, Li M, Hadley D, et al. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res*. Nov 2007;17(11):1665-74. doi:10.1101/gr.6861907
26. Colella S, Yau C, Taylor JM, et al. QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. *Nucleic Acids Res*. 2007;35(6):2013-25. doi:10.1093/nar/gkm076
27. Uban KA, Horton MK, Jacobus J, et al. Biospecimens and the ABCD study: Rationale, methods of collection, measurement and early data. *Dev Cogn Neurosci*. Aug 2018;32:97-106. doi:10.1016/j.dcn.2018.03.005
28. Baurley JW, Edlund CK, Pardamean CI, Conti DV, Bergen AW. Smokescreen: a targeted genotyping array for addiction research. *BMC Genomics*. Feb 27 2016;17:145. doi:10.1186/s12864-016-2495-7
29. Delaneau O, Howie B, Cox AJ, Zagury JF, Marchini J. Haplotype estimation using sequencing reads. *Am J Hum Genet*. Oct 3 2013;93(4):687-96. doi:10.1016/j.ajhg.2013.09.002
30. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*. Jun 2009;5(6):e1000529. doi:10.1371/journal.pgen.1000529
31. Huguet G, Schramm C, Douard E, et al. Measuring and Estimating the Effect Sizes of Copy Number Variants on General Intelligence in Community-Based Samples. *JAMA Psychiatry*. May 1 2018;75(5):447-457. doi:10.1001/jamapsychiatry.2018.0039
32. Alkan C, Coe BP, Eichler EE. Genome structural variation discovery and genotyping. *Nat Rev Genet*. May 2011;12(5):363-76. doi:10.1038/nrg2958
33. International HapMap C, Altshuler DM, Gibbs RA, et al. Integrating common and rare genetic variation in diverse human populations. *Nature*. Sep 2 2010;467(7311):52-8. doi:10.1038/nature09298
34. Sanders SJ, He X, Willsey AJ, et al. Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. *Neuron*. 2015;87(6):1215-1233.
35. Barch DM, Albaugh MD, Avenevoli S, et al. Demographic, physical and mental health assessments in the adolescent brain and cognitive development study: Rationale and description. *Dev Cogn Neurosci*. Aug 2018;32:55-66. doi:10.1016/j.dcn.2017.10.010
36. Karcher NR, Barch DM. The ABCD study: understanding the development of risk for mental and physical health outcomes. *Neuropsychopharmacology*. Jan 2021;46(1):131-142. doi:10.1038/s41386-020-0736-6
37. Achenbach TM, Verhulst F. Achenbach system of empirically based assessment (ASEBA). *Burlington, Vermont*. 2010;
38. Luciana M, Bjork JM, Nagel BJ, et al. Adolescent neurocognitive development and impacts of substance use: Overview of the adolescent brain cognitive development (ABCD) baseline neurocognition battery. *Dev Cogn Neurosci*. Aug 2018;32:67-79. doi:10.1016/j.dcn.2018.02.006
39. Gershon RC, Wagster MV, Hendrie HC, Fox NA, Cook KF, Nowinski CJ. NIH toolbox for assessment of neurological and behavioral function. *Neurology*. Mar 12 2013;80(11 Suppl 3):S2-6. doi:10.1212/WNL.0b013e3182872e5f
40. Glessner JT, Li J, Liu Y, et al. ParseCNV2: efficient sequencing tool for copy number variation genome-wide association studies. *Eur J Hum Genet*. Mar 2023;31(3):304-312. doi:10.1038/s41431-022-01222-7

41. Glessner JT, Bick AG, Ito K, et al. Increased frequency of de novo copy number variants in congenital heart disease by integrative analysis of single nucleotide polymorphism array and exome sequence data. *Circ Res*. Oct 24 2014;115(10):884-896. doi:10.1161/CIRCRESAHA.115.304458
42. Li YR, Glessner JT, Coe BP, et al. Rare copy number variants in over 100,000 European ancestry subjects reveal multiple disease associations. *Nat Commun*. Jan 14 2020;11(1):255. doi:10.1038/s41467-019-13624-1
43. Glessner JT, Li J, Wang D, et al. Copy number variation meta-analysis reveals a novel duplication at 9p24 associated with multiple neurodevelopmental disorders. *Genome Med*. Nov 30 2017;9(1):106. doi:10.1186/s13073-017-0494-1
44. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. May 2020;581(7809):434-443. doi:10.1038/s41586-020-2308-7
45. Collins RL, Glessner JT, Porcu E, et al. A cross-disorder dosage sensitivity map of the human genome. *Cell*. Aug 4 2022;185(16):3041-3055 e25. doi:10.1016/j.cell.2022.06.036
46. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. Aug 18 2016;536(7616):285-91. doi:10.1038/nature19057
47. McLaughlin RL, Schijven D, Van Rheenen W, et al. Genetic correlation between amyotrophic lateral sclerosis and schizophrenia. *Nature communications*. 2017;8(1):14774.
48. Spindola LM, Santoro ML, Pan PM, et al. Detecting multiple differentially methylated CpG sites and regions related to dimensional psychopathology in youths. *Clin Epigenetics*. Oct 21 2019;11(1):146. doi:10.1186/s13148-019-0740-z
49. Romero C, Werme J, Jansen PR, et al. Exploring the genetic overlap between twelve psychiatric disorders. *Nat Genet*. Dec 2022;54(12):1795-1802. doi:10.1038/s41588-022-01245-2
50. Savage JE, Jansen PR, Stringer S, et al. Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat Genet*. Jul 2018;50(7):912-919. doi:10.1038/s41588-018-0152-6
51. Nagy S, Maurer GW, Hentze JL, Rose M, Werge TM, Rewitz K. AMPK signaling linked to the schizophrenia-associated 1q21.1 deletion is required for neuronal and sleep maintenance. *PLoS Genet*. Dec 2018;14(12):e1007623. doi:10.1371/journal.pgen.1007623
52. Dou D, Zhao H, Li Z, et al. CHD1L Promotes Neuronal Differentiation in Human Embryonic Stem Cells by Upregulating PAX6. *Stem Cells Dev*. Nov 15 2017;26(22):1626-1636. doi:10.1089/scd.2017.0110
53. Yoshikawa S, Vila A, Segelken J, et al. Zebrafish connexin 79.8 (Gja8a): A lens connexin used as an electrical synapse in some neurons. *Dev Neurobiol*. May 2017;77(5):548-561. doi:10.1002/dneu.22418
54. Mefford HC, Sharp AJ, Baker C, et al. Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *N Engl J Med*. Oct 16 2008;359(16):1685-99. doi:10.1056/NEJMoa0805384
55. Brunetti-Pierri N, Berg JS, Scaglia F, et al. Recurrent reciprocal 1q21.1 deletions and duplications associated with microcephaly or macrocephaly and developmental and behavioral abnormalities. *Nat Genet*. Dec 2008;40(12):1466-71. doi:10.1038/ng.279
56. Ghasemi Firouzabadi S, Vameghi R, Kariminejad R, et al. Analysis of Copy Number Variations in Patients with Autism Using Cytogenetic and MLPA Techniques: Report of 16p13.1p13.3 and 10q26.3 Duplications. *Int J Mol Cell Med*. Fall 2016;5(4):236-245.
57. Shaw CA, Li Y, Wiszniewska J, et al. Olfactory copy number association with age at onset of Alzheimer disease. *Neurology*. Apr 12 2011;76(15):1302-9. doi:10.1212/WNL.0b013e3182166df5
58. Castellani CA, Melka MG, Wishart AE, et al. Biological relevance of CNV calling methods using familial relatedness including monozygotic twins. *BMC Bioinformatics*. Apr 21 2014;15:114. doi:10.1186/1471-2105-15-114
59. Malhotra D, Sebat J. CNVs: harbingers of a rare variant revolution in psychiatric genetics. *Cell*. Mar 16 2012;148(6):1223-41. doi:10.1016/j.cell.2012.02.039

60. Mollon J, Almasy L, Jacquemont S, Glahn DC. The contribution of copy number variants to psychiatric symptoms and cognitive ability. *Mol Psychiatry*. Apr 2023;28(4):1480-1493. doi:10.1038/s41380-023-01978-4
61. Calkins ME, Merikangas KR, Moore TM, et al. The Philadelphia Neurodevelopmental Cohort: constructing a deep phenotyping collaborative. *J Child Psychol Psychiatry*. Dec 2015;56(12):1356-1369. doi:10.1111/jcpp.12416
62. Feldstein Ewing SW, Bjork JM, Luciana M. Implications of the ABCD study for developmental neuroscience. *Dev Cogn Neurosci*. Aug 2018;32:161-164. doi:10.1016/j.dcn.2018.05.003
63. Chavez-Baldini U, Nieman DH, Keestra A, et al. The relationship between cognitive functioning and psychopathology in patients with psychiatric disorders: a transdiagnostic network analysis. *Psychological Medicine*. 2023;53(2):476-485.
64. Panicker N, Coutman M, Lawlor-O'Neill C, Kahl RG, Roselli S, Verrills NM. Ppp2r2a knockout mice reveal that protein phosphatase 2A regulatory subunit, PP2A-B55 $\alpha$ , is an essential regulator of neuronal and epidermal embryonic development. *Frontiers in Cell and Developmental Biology*. 2020;8:358.
65. Parrado A. P4-294: Identification of Genome-Wide Significant Associations in the Alzheimer's Disease Sequencing Project Dominican Families with Variants in UGT2B28, AXIN2, SYCE1, BCAT1, FAM13A, AND DQ594761. *Alzheimer's & Dementia*. 2016;12:P1146-P1146.
66. Harvard C, Strong E, Mercier E, et al. Understanding the impact of 1q21.1 copy number variant. *Orphanet J Rare Dis*. Aug 8 2011;6:54. doi:10.1186/1750-1172-6-54
67. Christiansen J, Dyck JD, Elyas BG, et al. Chromosome 1q21.1 contiguous gene deletion is associated with congenital heart disease. *Circ Res*. Jun 11 2004;94(11):1429-35. doi:10.1161/01.RES.0000130528.72330.5c
68. Niederhoffer KY, Fahiminiya S, Eydoux P, et al. Diagnosis of Van den Ende-Gupta syndrome: Approach to the Marden-Walker-like spectrum of disorders. *Am J Med Genet A*. Sep 2016;170(9):2310-21. doi:10.1002/ajmg.a.37831
69. Kendall KM, Rees E, Escott-Price V, et al. Cognitive performance among carriers of pathogenic copy number variants: analysis of 152,000 UK Biobank subjects. *Biological psychiatry*. 2017;82(2):103-110.
70. Douard E, Zeribi A, Schramm C, et al. Effect Sizes of Deletions and Duplications on Autism Risk Across the Genome. *Am J Psychiatry*. Jan 1 2021;178(1):87-98. doi:10.1176/appi.ajp.2020.19080834
71. Auwerx C, Lepamets M, Sadler MC, et al. The individual and global impact of copy-number variants on complex human traits. *The American Journal of Human Genetics*. 2022;109(4):647-668.
72. Hong EP, Park JW. Sample size and statistical power calculation in genetic association studies. *Genomics Inform*. Jun 2012;10(2):117-22. doi:10.5808/GI.2012.10.2.117
73. Moonesinghe R, Khoury MJ, Liu T, Ioannidis JP. Required sample size and nonreplicability thresholds for heterogeneous genetic associations. *Proc Natl Acad Sci U S A*. Jan 15 2008;105(2):617-22. doi:10.1073/pnas.0705554105
74. Guo Y, Sheng Q, Samuels DC, et al. Comparative study of exome copy number variation estimation tools using array comparative genomic hybridization as control. *Biomed Res Int*. 2013;2013:915636. doi:10.1155/2013/915636
75. de Ligt J, Boone PM, Pfundt R, et al. Detection of clinically relevant copy number variants with whole-exome sequencing. *Hum Mutat*. Oct 2013;34(10):1439-48. doi:10.1002/humu.22387
76. Tan R, Wang Y, Kleinstein SE, et al. An evaluation of copy number variation detection tools from whole-exome sequencing data. *Hum Mutat*. Jul 2014;35(7):899-907. doi:10.1002/humu.22537