medRxiv preprint doi: https://doi.org/10.1101/2024.08.02.24311302; this version posted August 6, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Fetal Gene Regulatory Gene Deletions are Associated with Poor Cognition and Altered Cortical Morphology in Schizophrenia and Community-Based Samples

Jennifer K. Forsyth[1,2], Jinhan Zhu[1], Ariana S. Chavannes[1], Zachary H. Trevorrow[1], Mahnoor Hyat[1], Sam A. Sievertsen[1], Sophie Ferreira-Ianone[1], Matthew P. Conomos[3], Keith H. Nuechterlein[4,6], Robert F. Asarnow[4], Michael F. Green[4,5], Katherine H. Karlsgodt[4,6], Diana O. Perkins[7], Tyrone D. Cannon[8], Jean M. Addington[9], Kristen S. Cadenhead[10], Barbara A. Cornblatt[11], Matcheri S. Keshavan[12], Daniel H. Mathalon[13], William S. Stone[12], Ming T. Tsuang[10], Elaine F. Walker[14], Scott W. Woods[15], Katherine L. Narr[4,16], Sarah C. McEwen[4], Charles H. Schleifer[4], Cindy M. Yee[4,6], Caroline K. Diehl[6], Anika Guha[6], Gregory A. Miller[4,6], Aaron F. Alexander-Bloch[17], Jakob Seidlitz[17], Richard A. I. Bethlehem[18], Roel A. Ophoff[4,19†] & Carrie E. Bearden[4,6†]

1 Department of Psychology, University of Washington, WA, USA

2 Institute for Public Health Genetics, University of Washington, WA, USA

3 Department of Biostatistics, University of Washington, WA, USA

4 Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles, CA, USA

5 VA Greater Los Angeles Healthcare System, Los Angeles, CA, USA

6 Department of Psychology, University of California, Los Angeles, CA, USA

7 Department of Psychiatry, University of North Carolina, Chapel Hill, NC, USA

8 Department of Psychology, Yale University, New Haven, CT, USA

9 Department of Psychiatry, University of Calgary, Calgary, Alberta, Canada

10 Department of Psychiatry, University of California, San Diego, CA, USA

11 Department of Psychiatry, Donald and Barbara Zucker School of Medicine at Hofstra/Northwell, Hempstead, NY, USA

12 Department of Psychiatry, Harvard Medical School at Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2024.08.02.24311302; this version posted August 6, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

13 Department of Psychiatry, University of California, and San Francisco Veterans Affairs Medical Center, San Francisco, CA, USA

14 Department of Psychology, Emory University, Atlanta, GA, USA

15 Department of Psychiatry, Yale University, New Haven, CT, USA

16 Department of Neurology, University of California, Los Angeles, CA, USA

17 Department of Psychiatry, University of Pennsylvania, Philadelphia, PA, USA

18 Department of Psychology, University of Cambridge, Cambridge, UK

19 Department of Human Genetics, University of California, Los Angeles, CA, USA

[†]Senior authors

***Corresponding authors:**

Jennifer Forsyth, Ph.D.

Department of Psychology, University of Washington

Email: jenforsy@uw.edu

Carrie E. Bearden, Ph.D.

Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles

Email: cbearden@mednet.ucla.edu

medRxiv preprint doi: https://doi.org/10.1101/2024.08.02.24311302; this version posted August 6, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

ABSTRACT

Schizophrenia spectrum disorders (SSDs) are characterized by substantial clinical and genetic heterogeneity. Multiple recurrent copy number variants (CNVs) increase risk for SSDs; however, how known risk CNVs and broader genome-wide CNVs influence clinical variability is unclear. The current study examined associations between borderline intellectual functioning or childhood-onset psychosis, known risk CNVs, and burden of deletions affecting genes in 18 previously validated neurodevelopmental gene-sets in 618 SSD individuals. CNV associations were assessed for replication in 235 SSD relatives and 583 controls, and 9,930 youth from the Adolescent Brain Cognitive Development (ABCD) Study. Known SSD- and neurodevelopmental disorder (NDD)-risk CNVs were associated with borderline intellectual functioning in SSD cases (odds ratios (OR) = 7.09 and 4.57, respectively); NDD-risk deletions were nominally associated with childhood-onset psychosis (OR = 4.34). Furthermore, deletion of genes involved in regulating gene expression during fetal brain development was associated with borderline intellectual functioning across SSD cases and non-cases (OR = 2.58), with partial replication in the ABCD cohort. Exploratory analyses of cortical morphology showed associations between fetal gene regulatory gene deletions and altered gray matter volume and cortical thickness across cohorts. Results highlight contributions of known risk CNVs to phenotypic variability in SSD and the utility of a neurodevelopmental framework for identifying mechanisms that influence phenotypic variability in SSDs, as well as the broader population, with implications for personalized medicine approaches to care.

medRxiv preprint doi: https://doi.org/10.1101/2024.08.02.24311302; this version posted August 6, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

INTRODUCTION

One of the greatest barriers to understanding the core biological processes that underlie schizophrenia spectrum disorders (SSD; i.e., schizophrenia and other psychotic disorders) and improving patient outcomes is the clinical and genetic heterogeneity of these disorders. For example, while psychosis onset is most common in late adolescence or early adulthood, a subset of patients experience psychosis as early as childhood[1]. There is also transdiagnostic heterogeneity in developmental course, with some patients manifesting depression, anxiety, and/or developmental disorders prior to psychosis onset[2]. Similarly, while cognitive functioning in SSD patients is, on average, 1 to 2 standard deviations below healthy individuals[3,4], some patients function in the above average range, while others function in the borderline or intellectual disability (ID) range. Importantly, early psychosis onset and poor cognitive function are associated with poor long-term outcomes, including higher rates of disability and a more chronic course of illness[5–8]. Understanding whether severe phenotypes in SSD are linked to distinct genetic risk profiles is a critical question with implications for personalized medicine approaches to care.

One class of genetic variants that may contribute to severe phenotypes in SSD are rare copy number variants (CNVs), in which large stretches of DNA are deleted or duplicated. CNVs can occur across the genome; however, regions flanked by low-copy repeats (LCRs) are particularly vulnerable to misalignment of sister chromatids during meiosis and non-allelic homologous recombination, giving rise to recurrent CNVs[9]. A Psychiatric Genomics Consortium (PGC) study of 21,094 SSD cases and 20,227 controls identified 8 recurrent CNV loci as significantly associated with SSD, with an additional 5 showing suggestive association[10]. Although only a small portion of SSD cases carry these CNVs (e.g., 1.4% of the PGC sample), they can increase risk substantially when present. Notably, multiple SSD-associated CNVs also increase risk for broader neurodevelopmental disorders (NDD), including ID and autism spectrum disorder (ASD)[11–15], indicating pleiotropic effects. Furthermore, SSD- and NDD-risk CNVs were recently associated with lower cognitive ability in the general population[16–18]. Few studies have investigated relationships between known risk CNVs and phenotypic variability within SSD cohorts; however, a recent study found that cognitive performance in patient carriers of known risk

medRxiv preprint doi: https://doi.org/10.1101/2024.08.02.24311302; this version posted August 6, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

CNVs was 0.5-1 SD below noncarriers[19]. NDD-risk CNVs were also associated with childhood-onset SSD in one prior study[20]; however, as childhood-onset psychosis is rare, this has yet to be replicated. Further studies are needed to clarify associations between known risk CNVs and severe phenotypes in SSD.

Beyond established risk CNVs, broader genome-wide deletions may also contribute to severe phenotypes in SSD. In the Genome Aggregation Database (gnomAD) study of 14,891 individuals, rare CNVs were found to yield loss-of-function (LOF) effects on ~5.5 genes per genome, accounting for an estimated 25-29% of rare, protein-truncating events per person[9]. Weak associations have been found between global CNV burden scores, such as total number of deleted genes, and cognitive functioning and psychopathology in some community sample studies,[21,22] but not others[18,23–25]. Importantly, refining global burden scores with genomic annotations may improve predictive power. For example, CNV scores weighted by the intolerance of affected genes to LOF were recently found to have stronger associations with cognitive functioning in general population samples[26–28], as well as in a SSD cohort[18], compared to non-weighted scores. However, it is unknown if other genomic annotations could further improve explanatory power of CNV scores.

One potential strategy is to annotate CNVs by the neurodevelopmental pathway that affected genes participate in. In particular, we previously derived and validated a set of 18 mutually-exclusive neurodevelopmental gene-sets, spanning 17,216 genes, by applying weighted-gene co-expression network analysis to BrainSpan transcriptomic data from 1,061 human brain samples spanning 6 weeks post-conception to 30 years of age[29]. Identifying these neurodevelopment gene-sets leveraged the fact that genes are co-expressed in specific patterns to give cells their stable identity and drive different biological processes[30]. Gene expression patterns change most dynamically in the brain during early development to drive the progression from cell proliferation to neural differentiation, neuronal migration, synapse formation, and circuit refinement[31]. These neurodevelopmental gene-sets showed distinct developmental expression trajectories and cell-type enrichment patterns, and were useful for identifying convergent versus divergent neurodevelopmental processes associated with genetic risk for SSD versus ASD. Specifically, genetic risk variants for SSD were enriched for gene-sets involved in

medRxiv preprint doi: https://doi.org/10.1101/2024.08.02.24311302; this version posted August 6, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

modulating neuronal excitability and postnatal synaptic signaling and plasticity, as well as transcriptional regulation across development. ASD risk variants showed overlapping enrichment for these gene-sets but were additionally enriched for gene-sets involved in early neuronal differentiation and synaptogenesis, as well as in regulating the gene expression changes that drive these processes during fetal development. This indicates that damaging variants in genes involved in fetal neurodevelopment are key contributors to risk for early-onset NDDs. However, it is unknown whether damaging variants in these genes contribute to severe phenotypes in SSD patients.

In the present study, we therefore investigated whether severe phenotypes in SSD (i.e., borderline intellectual functioning and childhood-onset SSD) were associated with known SSD- or NDD-associated risk CNVs, genome-wide deletion burden scores, or a novel set of scores capturing deletion burden in each of our 18, previously validated, neurodevelopmental gene-sets. Primary analyses focused on a within-SSD case, phenotypic variability design. However, analyses for borderline intellectual functioning were extended to non-cases to assess generalizability and replicability. Following associations between borderline intellectual functioning and fetal gene regulatory gene deletions across SSD cases and non-cases, we explored relationships between this gene-set and cortical morphology for subjects with available structural magnetic resonance imaging (MRI) data. Finally, we further assessed replicability of primary findings in the Adolescent Brain Cognitive Development (ABCD) Study® cohort of over 11,000 youth.

METHODS

SSD-Focused Cohort

Data from 1,514 individuals, including 645 with a SSD, 253 relatives of individuals with SSDs, and 616 healthy controls, were harmonized from studies conducted at UCLA, Feinstein Institute for Medical Research, and across 9 sites of the North American Prodrome Longitudinal Study (NAPLS 2 & 3). Genetic ancestry of participants was diverse, in line with the urban setting of most study sites (see Fig. S1). Recruitment foci of contributing studies included recent-onset, clinical high-risk, child- and adolescent-onset, and chronic SSD cohorts. Studies were included if the protocol included collection of

medRxiv preprint doi: https://doi.org/10.1101/2024.08.02.24311302; this version posted August 6, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

DNA samples from SSD patients (e.g., schizophrenia, schizoaffective disorder, schizophreniform disorder, psychotic disorder not otherwise specified, mood disorders with psychosis), structured clinical interviews for diagnosis, and standard cognitive measures. See Supplementary Methods for details.

Clinical, Cognitive, and MRI Data

Presence of a SSD was determined using the Structured Clinical Interview for DSM (SCID)[32] or the Kiddie Schedule for Affective Disorders and Schizophrenia (KSADS)[33]. Age of psychosis onset was defined via the psychosis module of the SCID or KSADS, or a question identifying age when full-blown psychotic symptoms lasted for >1 week. Childhood-onset SSD was defined as psychosis onset before age 13, in line with previous studies[20]. IQ estimates were derived from commonly used, standardized cognitive measures; borderline intellectual functioning was defined as an IQ estimate ≤ 85 (hereafter "borderline IQ"). SSD relative status was determined via the Family Interview for Genetic Studies (FIGS)[34], or because a family member with a SSD had been directly assessed. Specific inclusion and exclusion criteria differed between studies, but SSD relatives and controls had no history of a psychotic disorder. Detailed *n*s by clinical and cognitive assessment measure are provided in Table S1. T1-weighted MRI data were available for a subset of participants and processed using standard pipelines with Freesurfer version 7.1.1. Age- and sex-normalized cortical morphology metrics (i.e., centile scores) were derived using population reference models[35] (herein referred to as "BrainCharts"; see Supplementary Methods for details).

CNV Calling and Quality Control

CNV calling followed standard pipelines[10,28,36]. Briefly, DNA was extracted from whole blood, saliva, or buccal samples, followed by genome-wide genotyping at 700,078 sites using the Illumina Global Screening Array in two batches. CNVs were called in autosomal chromosomes using PennCNV[37] and QuantiSNP[38] with GC-correction. CNV calls were iteratively merged if the gap between neighboring calls was <20% of the total length using the PennCNV clean_cnv.pl script. CNVs were retained if they had a minimum PennCNV confidence score and QuantiSNP Log Bayes Factor score of

medRxiv preprint doi: https://doi.org/10.1101/2024.08.02.24311302; this version posted August 6, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

10, were concordant in direction (i.e., deletion or duplication) with ≥50% overlap between PennCNV and QuantiSNP and at least one algorithm with a confidence score ≥15, and were a minimum of 10 SNPs and 20 kilobases (kb) in length. CNVs overlapping ≥50% with centromeres, telomeres, or segmental duplications were excluded. Rare CNVs were defined as those with <50% overlap of any CNV found with a population maximum frequency ≥1% in gnomAD[9]. Samples with non-matching self-reported versus genetically-imputed sex (*n*=6), identified as duplicates (*n*=11), <98% genotyping rate (*n*=26), or outliers (>3 SD from median) in Log R Ratio SD (LRR SD), absolute waviness factor, B Allele Frequency standard deviation (BAF SD), or total deletion or duplication length (*n*=30) were excluded. Batch 1 and batch 2 were genotyped in hg19 and hg38 coordinates, respectively. CNV processing and annotation were completed in the native reference space of each batch and data across batches was combined at the analysis variable level.

CNV Annotations

CNVs associated with SSD[10], ASD (significant loci in SSC + AGP cohorts from [39], with >1 affected patient per deletion vs. duplication direction), developmental disorders (DD; downloaded May 2020 from https://www.deciphergenomics.org/[40]), or NDDs (significant CNVs from [12]) were compiled and lifted over to hg38 as needed (see Extended Table 1). Disorder-associated CNVs were identified if a call overlapped ≥40% with one of these known loci with concordant direction, except for NRXN1 deletions, which were called if any NRXN1 exon was overlapped by a deletion[14,41].

All CNVs passing quality control (QC) were annotated for overlap with exons for any protein-coding gene in RefSeq[42] (i.e., “NM_” prefix). Total number of genes with one or more exons spanned by deletions and number of genes with exons spanned by deletions for each of the 18 mutually-exclusive BrainSpan neurodevelopmental gene-sets described above (see Supplementary Methods and Extended Table 2 for detailed summary) were summed to generate total and gene-set-specific deletion scores for each subject, respectively. Additionally, the “loss-of-function observed/expected upper bound fraction” (LOEUF) metric from gnomAD, describing the ratio of observed versus expected LOF variants in 141,456 humans per protein-coding gene, was converted to percentiles across genes,

medRxiv preprint doi: https://doi.org/10.1101/2024.08.02.24311302; this version posted August 6, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

ordered such that genes with higher percentiles had lower LOEUF scores (i.e., greater mutational constraint)[43]. For each subject, LOEUF percentiles were summed across all deletion-spanning genes to generate a total LOEUF score, capturing total deletion burden weighted by affected genes' LOF-intolerance. Exploratory analyses were also conducted for deletion LOEUF scores within each neurodevelopmental gene-set.

Statistical Analyses

Logistic mixed models were first conducted using the GENESIS package in R[44] to compare rates of any rare deletion or duplication and known SSD- or broader NDD-risk (i.e., SSD-, ASD-, DD-, or NDD-associated) CNVs between SSD cases versus non-cases (i.e., controls and SSD relatives).

Logistic mixed models were then conducted among SSD cases to examine associations between childhood-onset psychosis or borderline IQ and 3 types of CNV scores: 1) presence of known SSD- or broader NDD-risk CNVs; 2) number of genes deleted in each neurodevelopmental gene-set; and 3) two genome-wide CNV metrics previously associated with neurobehavioral phenotypes - total number of deleted genes[21] and total LOEUF score[28]. To minimize model convergence failures and low-confidence coefficient estimates, only neurodevelopmental gene-sets with ≥3 subjects with deletions affecting the gene-set were analyzed. A linear mixed model was also conducted to test differences in IQ estimates among childhood-onset versus later-onset SSD cases. Covariates for sex, 10 ancestry principal components (PCs), and a genetic relatedness matrix (GRM) were included in all models. Age at testing was included as a covariate for borderline IQ analyses. Ancestry PCs and the GRM were generated using PC-AiR[45] and PC-Relate[46] from the GENESIS package, respectively. FDR correction was applied across all genetic scores analyzed that achieved model convergence relative to each severe SSD-related trait. All analyses used two-sided tests for significance.

Following initial associations between neurodevelopmental gene-sets and borderline IQ in SSD cases, parallel logistic mixed models were run: 1) excluding SSD cases with known SSD- or NDD-risk CNVs; 2) in non-cases only (i.e., SSD relatives and controls); and 3) across SSD, SSD relative, and

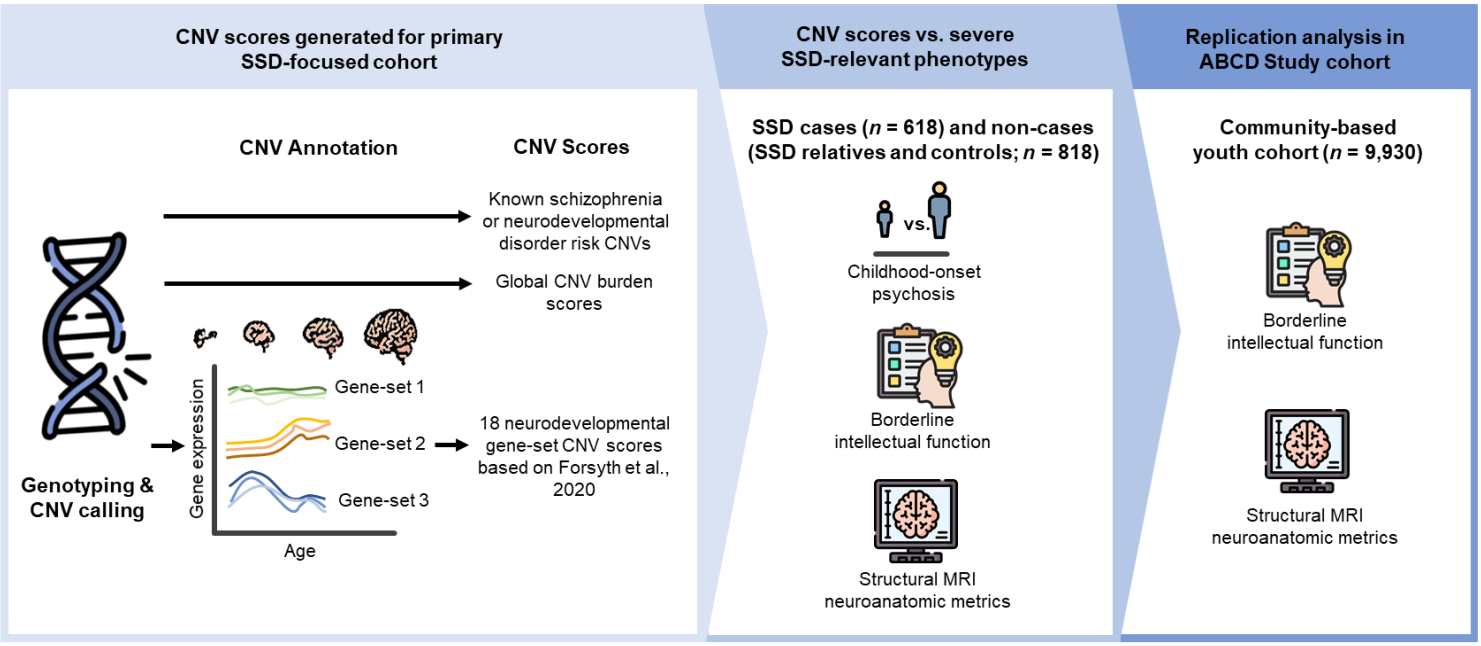
medRxiv preprint doi: https://doi.org/10.1101/2024.08.02.24311302; this version posted August 6, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

control subjects without known SSD- or NDD-risk CNVs. Analyses incorporating controls and SSD relatives included covariates for diagnostic group.

Replication in ABCD Study

ABCD is a study of over 11,000 youth in the United States with diverse demographic characteristics and longitudinal clinical, cognitive, behavioral, and neuroimaging data. Overall cognitive functioning estimates were available via the National Institutes of Health (NIH) Cognitive Toolbox for 11,471 subjects at baseline, when youth were ~9-10 years of age. Samples were genotyped using the Affymetrix Smokescreen array[47]. CNV calling and annotation followed the procedures described above for the primary SSD-focused cohort. Quality control (QC) procedures were similar with additional steps implemented to account for quality issues depending on genotyping plate (see Supplementary Methods). A schematic of the overall study analysis flow is shown in Figure 1.

Figure 1. Schematic of flow of analyses in the primary schizophrenia spectrum disorder (SSD) cohort and replication assessment of key copy number variant (CNV) score associations in the Adolescent Brain Cognitive Development (ABCD) Study cohort.



medRxiv preprint doi: https://doi.org/10.1101/2024.08.02.24311302; this version posted August 6, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

RESULTS

CNV Rates and Known Pathogenic CNVs in SSD Cases vs. Non-Cases

Data from up to 618 SSD cases, 235 SSD relatives, and 583 controls were included in analyses following QC (Table 1). Rare, large deletions or duplications (>20 kb, population frequency <1%) were identified in 544 SSD (88.0%), 208 SSD relative (88.5%), and 498 control (85.4%) subjects. The odds of SSD case status based on presence of rare deletions was not significant (OR = 1.25, 95%CI [0.99,1.58], p = .061), but was significant for deletions spanning genes (OR = 1.39, 95%CI [1.09,1.78], p = .008). There was no significant association between rare duplications and SSD case status (OR = 0.92, 95%CI [0.72,1.17], p = .49), including for duplications spanning genes (OR = 1.03, 95%CI [0.83,1.31], p = .74).

Table 1. Demographic, clinical, and CNV burden characteristics across groups for included subjects

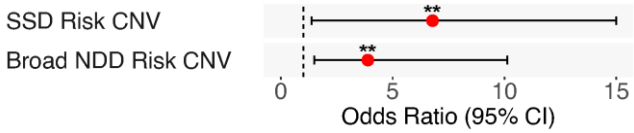
	SSD Subjects	SSD Relatives	Controls
N Subjects (% Female)	618 (30.1%)	235 (55.3%)	583 (50.6%)
Self-Identified Race			
White *n* (%)	280 (45.3%)	129 (54.9%)	338 (58.0%)
Black/African American *n* (%)	123 (19.9%)	17 (7.2%)	91 (26.9%)
Asian *n* (%)	60 (9.7%)	19 (3.1%)	73 (12.5%)
American Indian / Alaska Native / Native Hawaiian *n* (%)	16 (2.6%)	27 (11.5%)	17 (2.9%)
>1 *n* (%)	54 (8.7%)	30 (25.5%)	53 (9.1%)
Other/Unknown *n* (%)	85 (13.8%)	13 (2.1%)	12 (2.1%)
Hispanic/Latino *n* Yes/No/Unknown (% Yes)	143/388/87 (23.1%)	77/145/13 (32.8%)	119/456/8 (20.4%)
N with Rare Deletions (%)[+]	402 (65.0%)	146 (62.1%)	350 (60.0%)
N with Rare, Genic Deletions (%)	214 (34.6%)	64 (27.2%)	164 (28.1%)
N with Rare Duplications (%)[+]	411 (66.5%)	150 (63.8%)	388 (66.6%)
N with Rare, Genic Duplications (%)	324 (52.4%)	107 (45.5%)	291 (49.9%)
Mean IQ ± SD (Range, % Borderline IQ)	99.1 ± 16.0 (50-145, 20.0%)	105.1 ± 16.9 (55-145, 13.8%)	110.9 ± 13.6 (67-142, 4.5%)
Mean Age of Cognitive Testing ± SD	23.82 (10.28)	35.90 (18.36)	24.56 (12.30)
Mean Age Psychosis Onset ± SD (Range, % Child-Onset)	19.93 ± 5.54 (3-48, 6.4%)	*NA*	*NA*

+Across groups, 597 subjects (41.6%) had both rare deletions and duplications meeting QC criteria, 186 subjects (13.0%) had neither rare deletions or duplications

medRxiv preprint doi: https://doi.org/10.1101/2024.08.02.24311302; this version posted August 6, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Known SSD- and broader NDD-risk CNVs were identified in 9 (1.5%) and 16 (2.6%) SSD subjects, respectively, compared to 2 (0.2%) and 7 (0.9%) non-cases, respectively, revealing significantly increased rates among SSD cases for both SSD- (OR = 6.78, 95%CI [1.37,33.60], *p* = .019), and broad NDD-risk CNVs (OR = 3.89, 95%CI [1.49,10.12], *p* = .005; Fig. 2). Recurrent risk CNVs observed in multiple SSD subjects included NRXN1 deletions, 15q11.2 deletions, 16p12.1 deletions, and 1q21.1 TAR deletions (see Extended Table 3).

Figure 2. Odds ratio for SSD case status associated with known risk CNVs. **FDR q < 0.05, corrected for number of independent variables tested. Max OR confidence interval shown = 15.



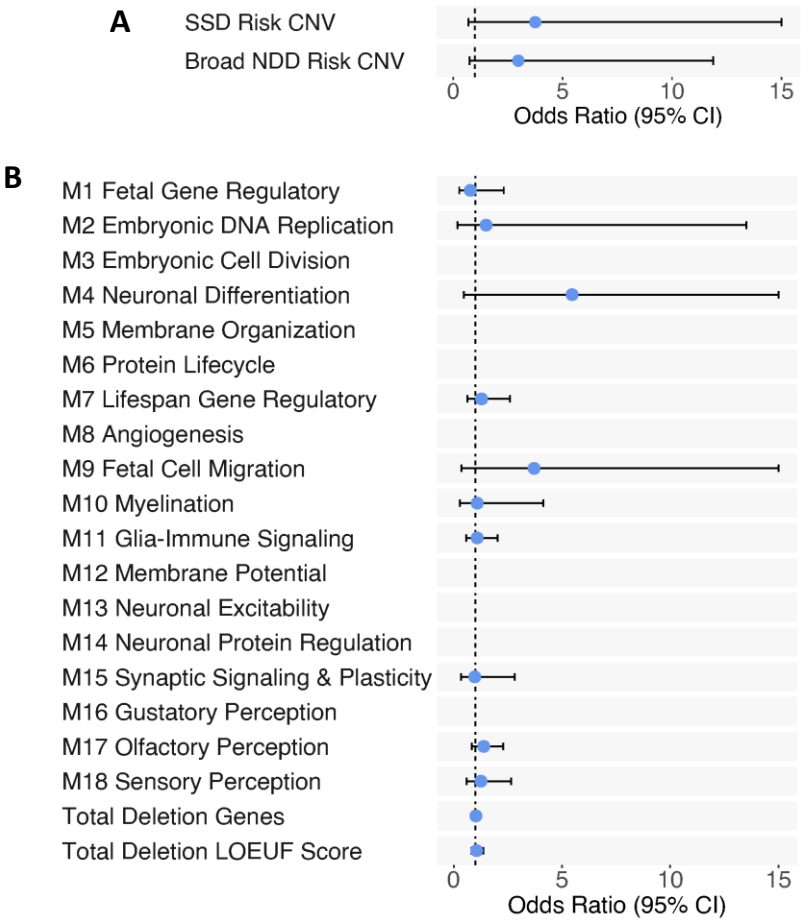
Child-Onset Psychosis and CNV Scores

Among SSD cases, 39 had childhood-onset psychosis (6.3%), with a mean onset at 8.56 years (*SD* = 2.77). The mean IQ estimate for childhood-onset cases (mean = 94.8, *SD* = 16.3) was significantly lower than for later-onset cases (mean = 99.6, *SD* = 15.8), *p* = .025.

SSD- and broader NDD-risk CNVs were found in 5.1% (2/39) and 7.7% (3/39) of childhood-onset cases, respectively, versus 1.2% (7/569) and 2.3% (13/569) of later-onset cases, respectively. Although elevated in childhood-onset psychosis, these rates were not significantly different (OR = 3.75, 95%CI [0.70,20.27], *p* = .124, FDR q = .703; OR = 2.98, 95%CI[0.75,11.89], *p* = .122, FDR q = .703, respectively; Fig. 3A). There were no significant associations between deletions in any neurodevelopmental gene-set and childhood-onset SSD, nor for global CNV scores, *ps* = *ns* (Fig. 3B).

Figure 3. Odds ratio (OR) for childhood-onset psychosis in SSD subjects associated with (A) presence of known risk CNVs or (B) number of genes deleted per neurodevelopmental gene-set, total number of genes deleted, and total LOEUF score of deleted genes. Max OR confidence interval shown = 15; see Extended Table 4 for full statistics

medRxiv preprint doi: https://doi.org/10.1101/2024.08.02.24311302; this version posted August 6, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



.

Borderline IQ and CNV Scores

One hundred and twenty SSD subjects (20.0%) had an IQ estimate in the borderline IQ range or lower. Of these, 4.2% had a SSD-risk CNV and 5.8% had a NDD-risk CNV, compared to 0.8% and 1.9% of cases without borderline IQ, respectively. These rates were significantly elevated for both SSD- (OR = 7.09, 95%CI[1.60,31.38], *p* = .01, FDR q = .033) and NDD-risk CNVs (OR = 4.57, 95%CI[1.48,14.11], *p* = .008, FDR q = .033; Fig. 4A).

Deletion burden in the M1 gene-set involved in regulating fetal gene expression during neuronal differentiation (OR = 2.27, 95%CI[1.34,3.84], p = .002, FDR q = .033), M6 gene-set involved in regulating protein lifecycle (OR = 6.53, 95%CI[1.39,30.66], p = .017, FDR q = .042), and M11 glia and immune signaling gene-set (OR = 1.82, 95%CI[1.21,2.75], p = .004, FDR q = .033) were also associated with borderline IQ, as was total number of deleted genes (OR = 1.11, 95%CI[1.04,1.48], p =

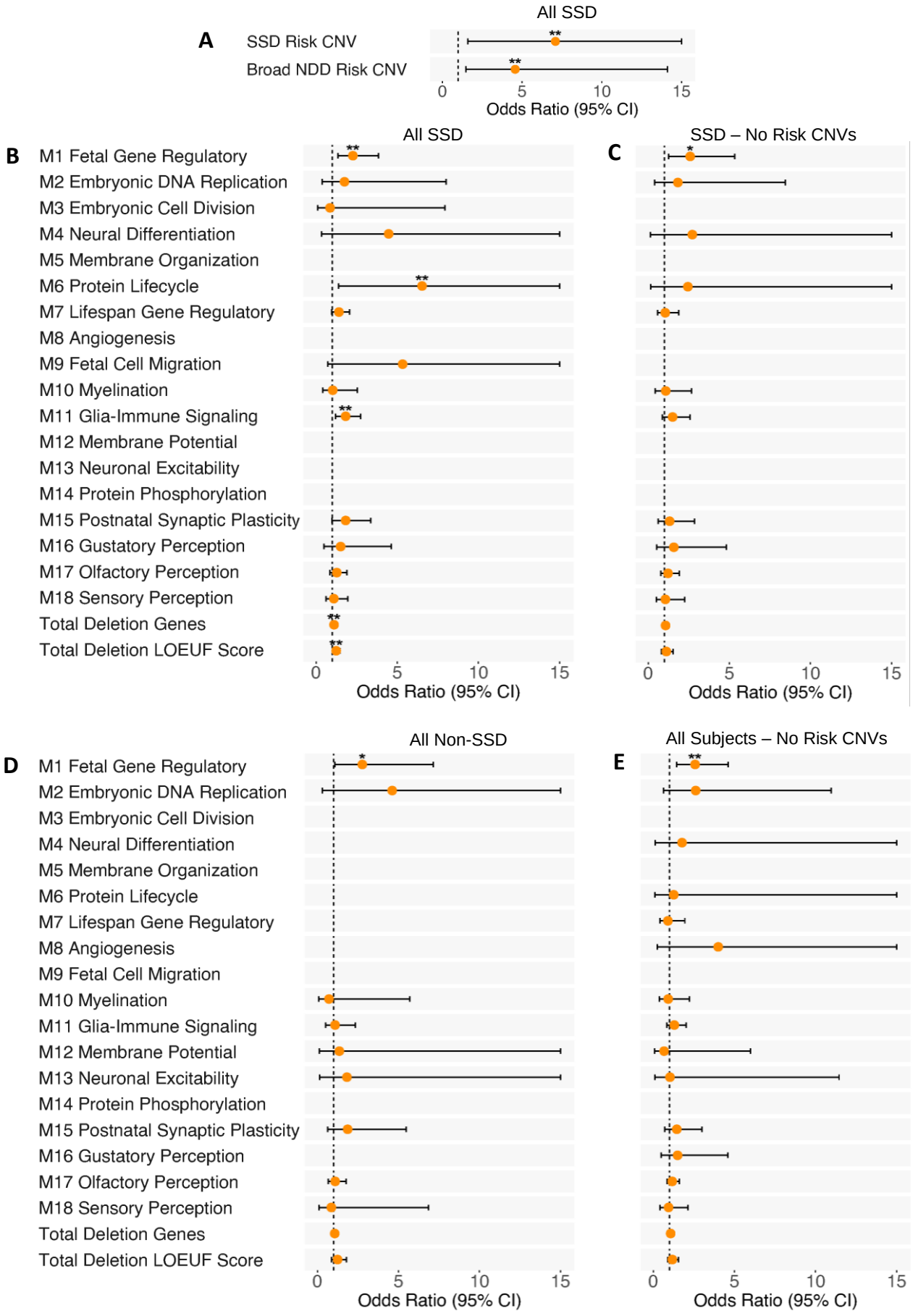
medRxiv preprint doi: https://doi.org/10.1101/2024.08.02.24311302; this version posted August 6, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

.009, FDR q = .033; Fig. 3B), and total LOEUF score (OR = 1.24, , 95%CI[1.04,1.48], p = .017, FDR q = .042).

Importantly, SSD- and NDD-risk CNVs are generally large, multi-gene CNVs[10,39,48], and often affect genes in the M1, M6, and M11 gene-sets, as well as LOF-intolerant genes (see Table S3). To determine whether the above associations exist beyond effects of known risk CNVs, we re-ran analyses excluding subjects with known risk CNVs. Among these subjects, only the M1 fetal gene regulatory gene-set remained associated with borderline IQ (OR = 2.59, 95%CI[1.25,5.34], p = .010, FDR q = .132; Fig 4C). Although this did not survive multiple-testing correction, the effect size was similar, suggesting this was due to reduced power rather than lower effect size. Consistent with this, extending analyses to non-SSD cases revealed a nominal association between M1 deletions and borderline IQ, with similar effect size (OR = 2.77, 95%CI[1.07,7.14], p = .035, FDR q = .301; Fig 4D). NDD-risk CNVs were also nominally associated with borderline IQ in non-cases (OR = 7.47, 95%CI[1.00,55.81], *p* = .050, FDR q = .301; Extended Table 5). Maximizing statistical power for gene-set analyses across SSD cases and non-cases without known risk CNVs revealed a significant association between M1 deletions and borderline IQ (OR = 2.58, 95%CI[1.44,4.62], p = .001, FDR q = .023; Fig 4E). Exploratory analyses for neurodevelopmental gene-sets weighted by LOEUF score showed a similar pattern of results (Fig. S2).

Figure 4. (A) Odds ratio (OR) for borderline intellectual functioning in SSD subjects associated with known risk CNVs. OR for number of genes deleted per neurodevelopmental gene-set, total number of genes deleted, and total deletion LOEUF score for (B) all SSD subjects and (C) for SSD subjects, excluding those with known risk CNVs. OR for borderline intellectual functioning and number of genes deleted per neurodevelopmental gene-set, total number of genes deleted, and total LOEUF score of deleted genes in (D) non-SSD subjects, and (E) SSD subjects, SSD relatives, and controls, excluding subjects with known risk CNVs. *p < 0.05; ** FDR q < 0.05, corrected for number of CNV scores tested per analysis group. Max OR confidence interval shown = 15; see Extended Table 5 for full statistics.

medRxiv preprint doi: https://doi.org/10.1101/2024.08.02.24311302; this version posted August 6, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



medRxiv preprint doi: https://doi.org/10.1101/2024.08.02.24311302; this version posted August 6, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Thus, beyond known risk CNVs, deletion of genes involved in regulating gene expression during fetal brain development increased likelihood of borderline IQ across SSD cases and non-cases. Indeed, among SSD subjects without a known risk CNV, 3.6% carried M1 gene-spanning deletions, 42.9% of whom were in the borderline IQ range compared to 18.5% of SSD subjects without M1 gene-spanning deletions. Similarly, across all subjects without known risk CNVs, 3.8% carried deletions affecting M1 genes, 26.4% of whom had borderline IQ compared to 11.6% without an M1 gene-spanning deletion.

Sensitivity Analyses

Restricting analyses to 438 SSD subjects with more narrowly defined schizophrenia or schizophreniform diagnoses revealed a highly similar pattern of associations for borderline IQ and child-onset psychosis, indicating that this pattern exists across the psychosis spectrum (see Supplementary Results, Fig. S3-4).

Exploratory Analyses of Known Risk Deletions vs. Duplications and CNVs of Uncertain Significance

The small number of SSD subjects with known-risk CNVs limited power to look at individual loci or SSD-risk duplications as a class. However, exploratory analyses of SSD- or NDD-risk deletions and NDD-risk duplications, separately, showed a nominal association between NDD-risk deletions and childhood-onset psychosis (OR = 4.34, 95%CI[2.91,5.78], p = .045); SSD-risk deletions showed a similar effect size that was not significant (OR = 4.34, 95%CI[2.62,6.07], p = .095). Analyses for borderline IQ similarly suggested stronger effects for SSD- and NDD-risk deletions (OR = 10.28, 95%CI[8.66,11.90], p = .005 and OR = 4.89, 95%CI[3.64,6.14], p = .013, respectively) compared to NDD-risk duplications (OR = 3.09, 95%CI[0.60,5.59], p = .37).

Exploratory analyses of rare CNVs >500 kb (i.e., potential pathogenic variants of “uncertain significance”[49]) showed no significant associations with borderline IQ or childhood-onset psychosis in SSD. See Supplementary Results for details.

medRxiv preprint doi: https://doi.org/10.1101/2024.08.02.24311302; this version posted August 6, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Exploratory Analyses of M1 Fetal Gene Regulatory Deletions and Structural MRI Deviation

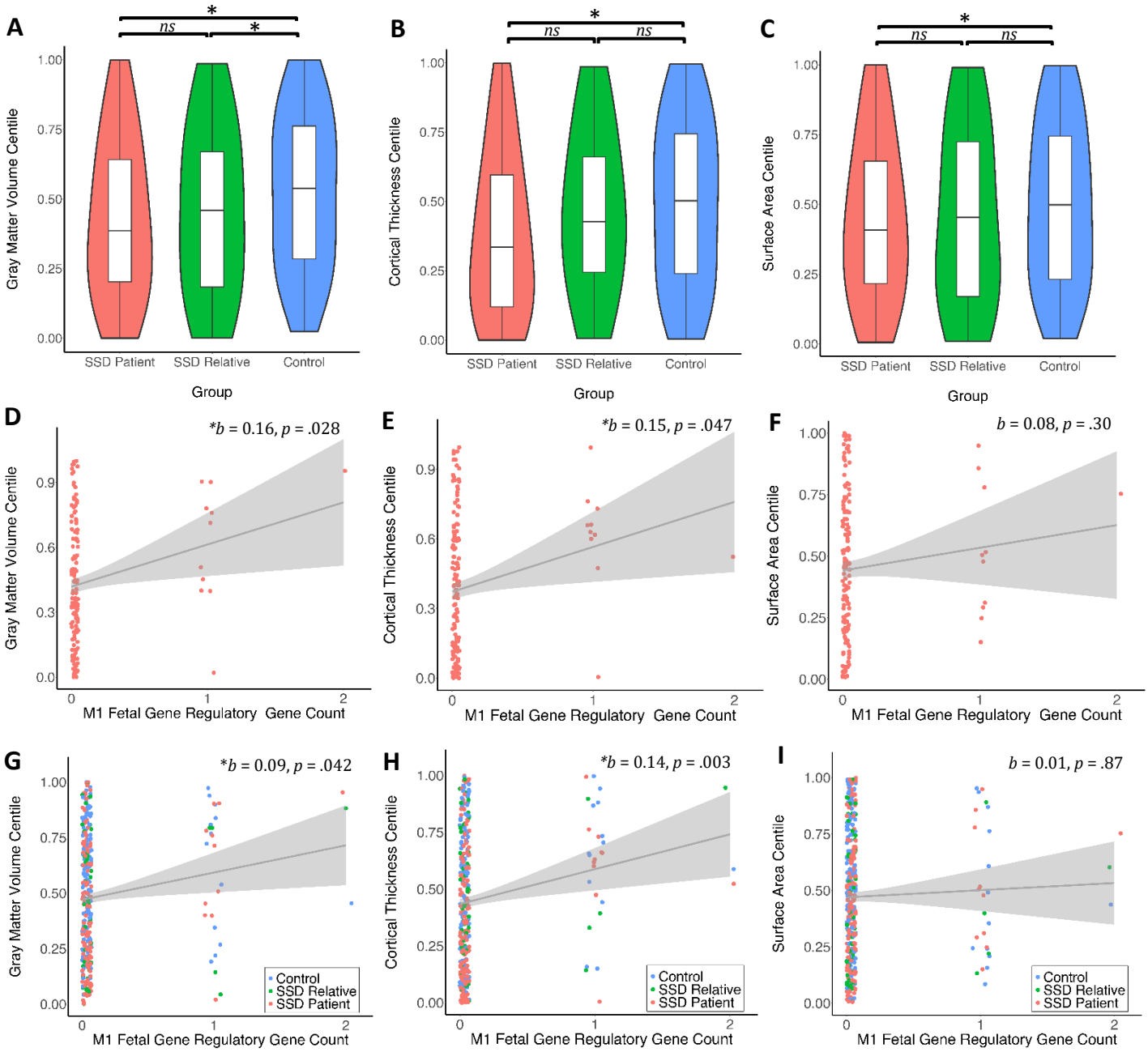
Reductions in cortical gray matter volume, surface area, and particularly cortical thickness, are well established in SSD[50]; however, SSD patients also show greater heterogeneity than controls across these measures[51], which could reflect variability in genetic risk profiles. As an exploratory follow-up analysis, we therefore examined whether M1 deletions are associated with cortical alterations. Leveraging the recently developed BrainCharts tool, which generates centile scores accounting for sex and age-normative changes in neuroanatomic metrics across development[35], we first assessed for differences in these metrics across groups, and then examined associations with M1 deletion burden.

Approximately half the cohort had QC-passing T1 structural MRI data (SSD n = 288, SSD relative n = 101, Control n = 308). There were significant group differences across gray matter volume, $F(2,692) = 12.87$, $p = .000003$, cortical thickness, $F(2,692) = 11.46$, $p = .00001$, and surface area centiles, $F(2,692) = 4.66$, $p = .01$ (Fig. 4A-C), with SSD cases showing reduced centiles across metrics compared to controls (ps < .02). Relatives were intermediate, differing significantly from controls for gray matter volume centile ($p = .007$), but not cortical thickness ($p = .60$) or surface area centile ($p = .10$), nor from SSD subjects for any metric (ps > .055).

Excluding subjects with known risk CNVs, which are associated with macrocephaly for some loci and microcephaly for others[52], M1 deletions in SSD subjects were associated with significantly increased gray matter volume ($b = 0.16$, 95%CI[0.02,0.29], $p = .028$) and cortical thickness centile ($b = 0.15$, 95%CI[0.002,0.30], $p = .047$); but not surface area centile ($b = 0.08$, 95%CI[-0.07,0.30], $p = .30$; Fig. 5D-F). Extending this analysis to all cases and non-cases without known risk CNVs showed a weaker but still significant association for gray matter volume ($b = 0.09$, 95%CI[0.003,0.17], $p = .042$), a similar association with cortical thickness centile ($b = 0.14$, 95%CI[0.05,0.24], $p = .003$), and no association for surface area centile ($b = 0.01$, 95%CI[-0.08,0.10], $p = .87$; Fig. 5G-I). M1 deletions weighted by LOEUF score were not associated with centiles in this half of the sample (ps = *ns*; Fig. S5).

medRxiv preprint doi: https://doi.org/10.1101/2024.08.02.24311302; this version posted August 6, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Figure 5. Group differences in gray matter volume (A), cortical thickness (B), and surface area (C) centiles. Excluding subjects with known risk CNVs, relationship between M1 deletions and cortical centile scores for (D-F) SSD subjects and (G- I) across all subjects. To facilitate interpretability, relationships are plotted using raw centile and gene count values; reported effect sizes and p-values are from the mixed models. *$p < 0.05$.



medRxiv preprint doi: https://doi.org/10.1101/2024.08.02.24311302; this version posted August 6, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Replication Analyses in ABCD

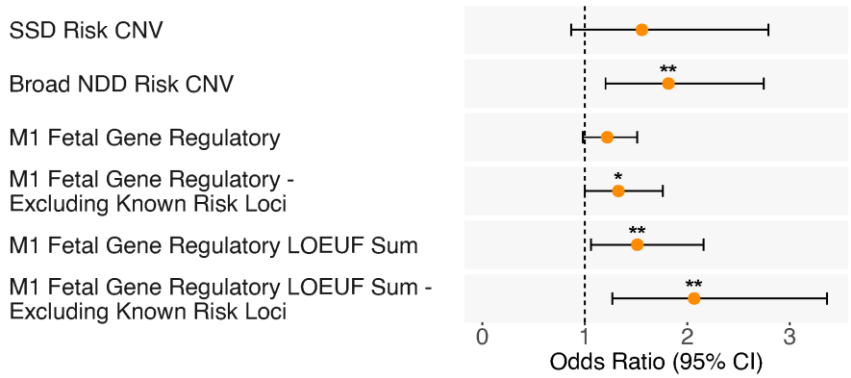
In ABCD, 9,930 youth had QC-passing CNV calls. Rates of rare, large deletions and duplications were lower CNV in ABCD compared to the SSD cohort (Table 2), which may reflect noisier Affymetrix Array data (Table S3). Rates of borderline IQ were significantly elevated among carriers of NDD-risk CNVs (OR = 1.82, 95%CI[1.20,2.74], p = .005, FDR q = .014) and were non-significantly elevated in carriers of SSD-risk CNVs (OR = 1.56, 95%CI[0.87,2.79], p = .14, FDR q = .14; Figure 6). Likelihood of borderline IQ was not significantly associated with M1 gene deletions when subjects with NDD-risk CNVs were included (OR = 1.23, 95%CI[0.98,1.51], p = .072, FDR q = .086); but was nominally associated when subjects with NDD-risk CNVs were excluded (OR = 1.33, 95%CI[1.00,1.76], p = .049, FDR q = .074, respectively). When M1 deletions were weighted by LOF intolerance, they were significantly associated with borderline IQ, both including (OR = 1.51, 95%CI[1.06,2.16], p = .023, FDR q = .046), or excluding subjects with NDD-risk CNVs (OR = 2.07, 95%CI[1.27,3.37], p = .003, FDR q = .014). Effect sizes were attenuated overall in ABCD compared to the SSD-focused cohort.

Table 2. Baseline demographic, clinical, and CNV burden characteristics for included ABCD subjects.

ABCD Cohort After Quality Control	
N Subjects (% Female)	9930 (47.17%)
Self-Identified Race	
White *n* (%)	6494 (65.4%)
Black/African American *n* (%)	1653 (16.6%)
Asian *n* (%)	180 (1.8%)
American Indian / Alaska Native / Native Hawaiian *n* (%)	52 (0.5%)
>1 *n* (%)	1024 (10.3%)
Other/Unknown *n* (%)	527 (5.3%)
Hispanic/Latino *n* (%)	1882 (19.0%)
N with Included Deletions (%)	2834 (28.54%)
N with Included Duplications (%)	3527 (35.52%)
N with Included Genic Deletions (%)	1426 (14.36%)
N with Included Genic Duplications (%)	2963 (29.84%)
N with SSD / NDD Risk CNV Hits (% Subjects)	80 (0.81%) / 154 (1.55%)
Mean NIH-TB Total Composite IQ ± SD (% Borderline IQ)	100.09 ± 17.87 (19.1%)
Mean Age of NIH-TB Testing (SD)	9.92 (0.62)

medRxiv preprint doi: https://doi.org/10.1101/2024.08.02.24311302; this version posted August 6, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

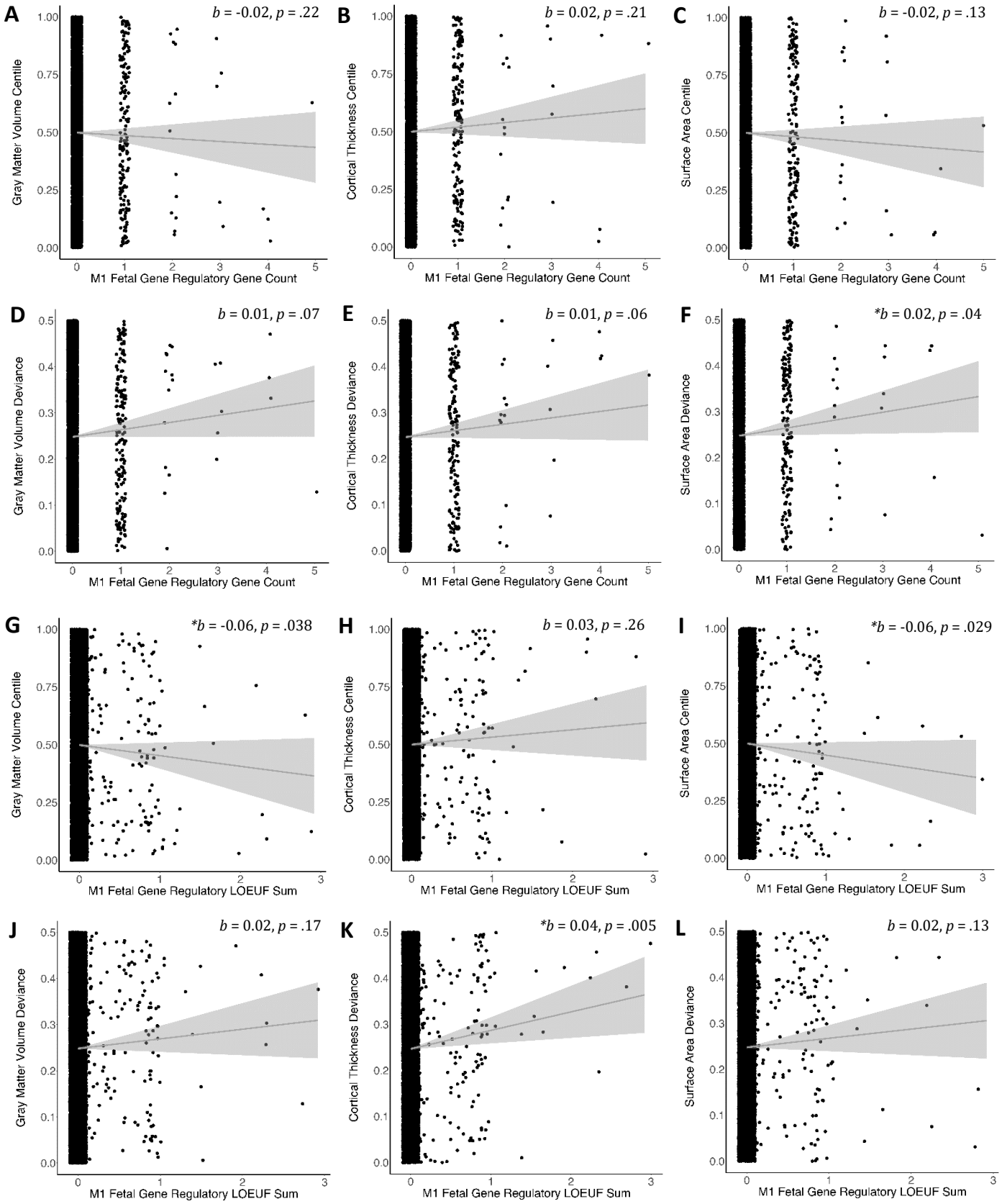
Figure 6. Odds ratio (OR) for borderline intellectual functioning in ABCD study subjects associated with SSD-risk CNVs, broad NDD-risk CNVs, number of M1 genes deleted, and M1 LOEUF score, including or excluding subjects with known risk CNVs. **FDR $q < .05$, corrected for number of CNV scores tested.



In the 9,374 youth with QC-passing MRI data, excluding subjects with NDD-risk CNVs, number of deleted M1 genes was not associated with centile scores in a consistent direction, *p*s = *ns* (Fig. 7A-C), but was associated with deviance in surface area centile (i.e., both larger and smaller surface area; $b = .02$, 95%CI[0.001,0.03], $p = .041$, Fig. 7F). Furthermore, M1 deletions weighted by LOF intolerance were associated with lower gray matter volume ($b = -0.06$, 95%CI[-0.11,-0.003], $p = .038$; Fig. 7D) and surface area centiles ($b = -0.06$, 95%CI[-0.11,-0.006], $p = .029$; Fig. 7F), as well as with deviance in cortical thickness centile ($b = .04$, 95%CI[0.01,0.07], $p = .005$, Fig. 7K). Thus, in late childhood-aged participants, M1 deletions weighted by LOF intolerance were associated smaller age-normalized gray matter volume and surface area and with increased deviation from age-normalized thickness in either direction (i.e., relatively less or greater cortical thickness).

Figure 7. Associations between M1 deletion gene count and (A) gray matter volume, (B) cortical thickness, or (C) surface area centile or deviance in these centile scores (D-F) in ABCD study subjects. Parallel associations between M1 LOEUF score and (G) gray matter volume, (H) cortical thickness, or (I) surface area centile or deviance in these centile scores (J-L). Relationships are plotted using raw centile and M1 scores; reported effect sizes and p-values are from the mixed models. *$p < 0.05$.

medRxiv preprint doi: https://doi.org/10.1101/2024.08.02.24311302; this version posted August 6, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



medRxiv preprint doi: https://doi.org/10.1101/2024.08.02.24311302; this version posted August 6, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

DISCUSSION

Here, we investigated relationships between severe SSD-relevant phenotypes and CNV burden scores in two cohorts and found, for the first time, that the deletion of fetal gene regulatory genes that drive early brain development was associated with severe phenotypes in both SSD individuals and the broader population. Thus, known SSD- and NDD-risk CNVs were strongly associated with impaired cognitive functioning in individuals with SSDs, while NDD-risk deletions were nominally associated with childhood-onset psychosis. However, beyond the effects of known risk CNVs, deletions impacting genes that regulate the gene expression changes that orchestrate early neuronal differentiation and development also increased risk for impaired cognitive functioning and were also associated with altered cortical morphology. These effects were observed across SSD patients, relatives, and controls, and in the community-based ABCD youth cohort. These results demonstrate the utility of a neurodevelopmental framework for identifying biological processes that shape phenotypic variability across the diagnostic spectrum and suggest that disruptions to early neuronal development, in particular, contribute to severe SSD-relevant phenotypes.

Established risk CNVs for SSD and broader NDDs are often large, affect regions of the genome with high genic content, and frequently span genes that are intolerant to LOF. Our observation that 1.5% and 2.6% of SSD patients carried a CNV previously associated with SSD or NDDs more broadly, respectively, is in line with prior estimates[10,19]. We also observed non-significantly elevated rates of SSD- and NDD-risk CNVs in patients with childhood-onset psychosis, and a nominal association between childhood-onset psychosis and NDD-risk deletions. This suggests that known NDD-risk deletions, in particular, may increase likelihood of this rare and severe clinical presentation; however, this requires replication. To our knowledge, no other studies have reported the rate of SSD-risk CNVs in childhood-onset patients and only one prior study reported rates of broader NDD-associated CNVs in childhood-onset patients[20]. In this prior study, NDD-risk deletions and duplications were analyzed together and found in 11.4% of 126 individuals with childhood-onset schizophrenia. The slightly lower rate of 7.7% of childhood-onset psychosis subjects with a NDD-risk CNV in the current study compared

medRxiv preprint doi: https://doi.org/10.1101/2024.08.02.24311302; this version posted August 6, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

to [20] may reflect differences in patient characteristics, ascertainment strategy, genotyping platform and/or statistical noise.

Our finding that known SSD- and broader NDD-risk CNVs were associated with 7- and 4.5-fold increased risk for borderline IQ, respectively, in SSD cases is consistent with pleiotropic associations of many of these CNVs with ID and DD[13]. Similar but attenuated associations between NDD-risk CNVs and borderline IQ were also found among non-cases and in the ABCD study, consistent with other community samples[16,22]. Our results add to a growing literature that cognitive impairment is elevated among individuals with SSD- and NDD-risk CNVs[18,19]. As genetic diagnoses can facilitate access to specialized services[53] and poor cognitive function is associated with a more chronic course of illness in SSD, this suggests that genetic testing for CNVs could be useful when a patient with a SSD presents with poor cognitive functioning. However, as only 5.8% of patients with borderline IQ had a known risk CNV, this highlights the importance of considering other potential contributing factors, which may include common variants (i.e., polygenic contribution) and other damaging variants.

Indeed, beyond known risk CNVs, our novel neurodevelopmental pathway annotation of genome-wide CNV scores revealed that deletions affecting M1 fetal gene regulatory genes were associated with borderline IQ and altered cortical morphology in SSD cases, as well as in non-cases. Notably, this neurodevelopmental gene-set was previously found to be strongly enriched for genetic variants associated with ASD[29]. Genes in this module are enriched for fetal-specific expression and hub genes include the histone demethylase enzyme, *KDM5B*, and the transcription factors, *SOX11* and *SOX4,* which are known to critically regulate neurogenesis, neuronal differentiation, and dendritic morphogenesis[54–56]. Deleted M1 genes in SSD patients with borderline IQ outside known risk loci included *SEMA3C*, *MACROD2*, *CDKAL1*, and *ZNF568*. *SEMA3C* encodes a glycoprotein that is secreted as an axonal guidance cue for developing neurons and facilitates nervous-system patterning during embryonic development[57,58]. *MACROD2* encodes a deacetylase involved in DNA repair and chromatin structure that is expressed predominantly during fetal brain development[59], and is a candidate ASD-risk gene[60], although it has yet to be associated with ASD in large-scale genome-wide studies. *ZNF568* is a transcription factor essential for maintaining the neural stem cell pool during fetal

medRxiv preprint doi: https://doi.org/10.1101/2024.08.02.24311302; this version posted August 6, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

development and is associated with brain size in mice and humans[61–63]. *CDKAL1* encodes a transfer RNA modifying enzyme that has been shown to be important for accurate protein translation, and has ubiquitous tissue expression, including in brain[64,65]. Common variants spanning *CDKAL1* are robustly associated with Type 2 Diabetes[64]. Follow-up studies are needed to clarify the role of these genes in cognitive functioning and cortical morphology. Nevertheless, these findings suggest that deletions of genes involved in regulating the gene expression changes that shape early neurogenesis and synapse formation can disrupt cortical development and cognitive functioning in individuals with SSDs, as well as the broader population.

Interestingly, associations between poor cognitive functioning, M1 deletions, and known risk CNVs were attenuated in the youth ABCD cohort compared to the SSD-focused cohort. While the direction of associations remained the same, only NDD-risk CNVs and M1 deletions weighted by LOF-intolerance were significantly associated with borderline intellectual functioning in ABCD. Interestingly, M1 deletions weighted by LOF-intolerance were associated with smaller gray matter volume and surface area in ABCD, and both increases and decreases in age-normalized cortical thickness. This is consistent with a recent finding that high LOEUF-based CNV scores were associated with extreme neuroanatomic centiles, in either direction, in the Philadelphia Neurodevelopmental Cohort and effects of known pathogenic CNVs on cortical thickness in both directions[26]. The stronger association between poor cognitive function and M1 deletions in ABCD when LOF-intolerance information was incorporated is consistent with prior studies weighting genome-wide deletion scores by LOF-intolerance in other general population cohorts[26–28]. Attenuated associations for known risk CNVs and M1 deletions in ABCD may reflect lower quality CNV calls and/or lower heritability of cognitive functioning in childhood (h^2 = ~0.4) compared to adulthood (h^2 = ~0.80)[66–68]. Indeed, cognitive abilities increase substantially between childhood and adulthood, as neural circuits undergo postnatal refinement[69,70], such that the full phenotypic effects of risk CNVs and M1 deletions may not be evident until later in development. Differences in cognitive assessment battery (i.e., NIH-TB) and/or the specific genes affected by deletions in ABCD compared to the SSD-focused cohort may have also contributed to attenuated effects in ABCD.

medRxiv preprint doi: https://doi.org/10.1101/2024.08.02.24311302; this version posted August 6, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Several limitations to the present study should be noted. First, many SSD patients were recruited through studies focused on early-onset SSD and/or on identifying risk factors associated with future development of psychosis. Psychotic diagnoses included in the primary SSD-cohort were therefore broad, which could result in noisier effect size estimates for associations between phenotypes and CNV scores. However, including broader diagnoses offers opportunities to understand relationships between genetic and clinical variability across SSD diagnoses. Indeed, analyses in subjects with narrowly-defined schizophrenia or schizophreniform diagnoses showed highly similar CNV-phenotype association patterns, suggesting that the observed associations exist across the psychosis spectrum. Second, relatives and controls in the SSD-focused cohort were often recruited to be demographically similar to clinical high-risk or early-onset SSD subjects, such that they were assessed prior to passing through the primary risk period for developing a SSD. Some non-cases with known risk CNVs could therefore still develop a SSD in the future; however, this would mean that our effect size estimates for known risk CNVs versus SSD case status are conservative. Third, cognitive, clinical, and MRI data acquisition methods differed between contributing studies in the SSD-focused cohort. Post-hoc harmonization was necessary to achieve a sufficiently powered sample for within-case genotype-phenotype analyses but may have contributed noise to our models. Nevertheless, we were able to replicate associations between key CNV scores and cognitive and cortical metrics across non-cases, as well as in the youth-based ABCD cohort, providing independent support across studies and assessment methods. Finally, we had limited power to detect associations between clinical variability and deletion burden in small neurodevelopmental gene-sets. Relatedly, while the neurodevelopmental gene-sets used to annotate deletions were previously validated for shared biological function using independent datasets (e.g., gene ontology, known protein-protein interactions, and transcriptional co-regulation) and offer an important opportunity to examine relationships between severe phenotypes and deletion burden in a set of mutually exclusive, genome-wide gene-sets, in reality, many genes are involved in multiple biological functions. It is possible that parsing CNV scores by gene-sets derived through other methods could reveal stronger associations between CNV scores and severe phenotypes. Analyses in larger datasets that can accommodate higher multiple testing correction and

medRxiv preprint doi: https://doi.org/10.1101/2024.08.02.24311302; this version posted August 6, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

are better powered for testing associations in small gene-sets will be necessary to identify the optimal biological parcellation of CNV risk scores for patient stratification.

In summary, known NDD-risk CNVs were associated with poor cognitive function among individuals with SSDs in the current study, as well as in non-SSD individuals. Deletion of genes involved in regulating gene expression changes necessary for early neuron development and differentiation were also shown, for the first time, to be associated with poor cognitive function, as well as altered cortical morphology. The implication of a fetal development-specific etiologic process in poor cognitive function for a subset of SSD individuals may help explain why traditional treatment strategies involving antipsychotic and/or psychotherapy initiation after illness onset (i.e., usually in late adolescence or adulthood) have poorer efficacy among patients with poor cognitive function[71,72]. This suggests that the development of alternate interventions that can be administered earlier in development may be necessary to optimize therapeutic effects for some patients. Notably, CNV score associations were found in two diverse ancestry cohorts, highlighting the potential advantage of investigating functionally-defined variants for identifying generalizable genotype-phenotype relationships, relative to studying variants defined based on population frequency (i.e., single nucleotide polymorphisms (SNPs)). Overall, our findings demonstrate the utility of a neurodevelopmental lens for understanding how damaging variants contribute to phenotypic variability in SSD and the broader population, and offers a promising direction for nominating specific genes that may contribute to severe phenotypes in SSD for future investigation.

Author Contributions: JKF conceived the study with support from CEB, RAO, and KHN. CEB, KHN, RFA, MFG, KHK, DOP, TDC, JMA, KSC, BAC, MSK, DHM, WSS, MTT, EFW, SWW, KLN, SCM, CMY, and GAM obtained funding for, and oversaw, original data collection and curation. CKD and AG contributed to data collection. JKF obtained funding for genotyping and data harmonization. JKF oversaw data curation during the data harmonization phase with support from CEB, KHN, and RAO, and contributions from SAS, ASC, SFI, ZHT, JZ, and MH. JKF led genetic data processing with assistance from JZ, MH, ZHT, and SAS. JKF, SFI, ASC, ZHT, SAS, CHS, and KLN contributed to

medRxiv preprint doi: https://doi.org/10.1101/2024.08.02.24311302; this version posted August 6, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

neuroimaging processing and analysis with support from JS and AAB for centiles generation. JKF carried out formal analyses with support from CEB, RAO, KHN, and MPC. Code contributions were made by JKF, JZ, and MPC for genetic data processing and analyses, and JS and RAI for centiles generation. JKF led manuscript write-up in consultation with CEB and RAO. All authors contributed to manuscript review and editing.

Disclosures: RAIB, JS and AFA-B are co-founders of Centile Bioscience. RAIB and JS additionally serve as directors for Centile Bioscience Inc. and Ltd.

Acknowledgements: The authors thank Evan Eichler for his feedback and advice on analyses. This work was supported by NIMH (K08 MH118577 to JKF; MH101506 to KHK; R01 MH37705, R01 MH110544 and P50 MH066286 to KHN; U01 MH082004 to DOP; U01 MH082004 to TDC; U01 MH081988 to EFW; R37 MH085953, U01 MH081902, P50 MH066286, R01 MH129858, U01 MH124639, R01 MH123575, and U01 MH119736 to CEB; F31 MH119786 to CKD; F31 MH124421 to AG), the Brain & Behavior Research Foundation (NARSAD Young Investigator Award to JKF), the National Center for Advancing Translational Sciences UCLA CTSI Grant UL1TR001881 (Award to CEB & JKF), the UCLA Brain Research Institute (Postdoctoral Award to JKF), and the Shear Family Foundation. Figure 1 (i.e., the schematic of analyses) was created using images from Flaticon.com.

medRxiv preprint doi: https://doi.org/10.1101/2024.08.02.24311302; this version posted August 6, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

REFERENCES

1. Forsyth, J. K. & Asarnow, R. F. Genetics of Childhood-onset Schizophrenia 2019 Update. *Child Adolesc. Psychiatr. Clin. N. Am.* (2019) doi:10.1016/j.chc.2019.08.007.
2. Forsyth, J. K. & Bearden, C. E. Rethinking the First Episode of Schizophrenia: Identifying Convergent Mechanisms During Development and Moving Toward Prediction. *Am. J. Psychiatry* **180**, 792–804 (2023).
3. McCutcheon, R. A., Keefe, R. S. E. & McGuire, P. K. Cognitive impairment in schizophrenia: aetiology, pathophysiology, and treatment. *Mol. Psychiatry* **28**, 1902–1918 (2023).
4. Rajji, T. K., Ismail, Z. & Mulsant, B. H. Age at onset and cognition in schizophrenia: meta-analysis. *Br. J. Psychiatry* **195**, 286–293 (2009).
5. Rabinowitz, J., Levine, S. Z. & Häfner, H. A population based elaboration of the role of age of onset on the course of schizophrenia. *Schizophr. Res.* **88**, 96–101 (2006).
6. Díaz-Caneja, C. M. *et al.* Predictors of outcome in early-onset psychosis: a systematic review. *NPJ Schizophr* **1**, 14005 (2015).
7. Legge, S. E. *et al.* Clinical indicators of treatment-resistant psychosis. *Br. J. Psychiatry* **216**, 259–266 (2020).
8. Green, M. F., Kern, R. S. & Heaton, R. K. Longitudinal studies of cognition and functional outcome in schizophrenia: implications for MATRICS. *Schizophr. Res.* **72**, 41–51 (2004).
9. Collins, R. L. *et al.* A structural variation reference for medical and population genetics. *Nature* **581**, 444–451 (2020).
10. Marshall, C. R. *et al.* Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nat. Genet.* **49**, 27–35 (2017).
11. Rees, E. *et al.* Analysis of Intellectual Disability Copy Number Variants for Association With Schizophrenia. *JAMA Psychiatry* **73**, 963–969 (2016).
12. Coe, B. P. *et al.* Refining analyses of copy number variation identifies specific genes associated with developmental delay. *Nat. Genet.* **46**, 1063–1071 (2014).
13. Kirov, G. CNVs in neuropsychiatric disorders. *Hum. Mol. Genet.* **24**, R45-9 (2015).

medRxiv preprint doi: https://doi.org/10.1101/2024.08.02.24311302; this version posted August 6, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

14. Birnbaum, R., Mahjani, B., Loos, R. J. F. & Sharp, A. J. Clinical Characterization of Copy Number Variants Associated With Neurodevelopmental Disorders in a Large-scale Multiancestry Biobank. *JAMA Psychiatry* **79**, 250–259 (2022).
15. Birnbaum, R. & Weinberger, D. R. Genetic insights into the neurodevelopmental origins of schizophrenia. *Nat. Rev. Neurosci.* **18**, 727–740 (2017).
16. Kendall, K. M. *et al.* Cognitive Performance Among Carriers of Pathogenic Copy Number Variants: Analysis of 152,000 UK Biobank Subjects. *Biol. Psychiatry* **82**, 103–110 (2017).
17. Kendall, K. M. *et al.* Cognitive performance and functional outcomes of carriers of pathogenic copy number variants: analysis of the UK Biobank. *Br. J. Psychiatry* **214**, 297–304 (2019).
18. Thygesen, J. H. *et al.* Genetic copy number variants, cognition and psychosis: a meta-analysis and a family study. *Mol. Psychiatry* **26**, 5307–5319 (2021).
19. Hubbard, L. *et al.* Rare Copy Number Variants Are Associated With Poorer Cognition in Schizophrenia. *Biol. Psychiatry* **90**, 28–34 (2021).
20. Ahn, K. *et al.* High rate of disease-related copy number variations in childhood onset schizophrenia. *Mol. Psychiatry* **19**, 568–572 (2014).
21. Guyatt, A. L. *et al.* Association of copy number variation across the genome with neuropsychiatric traits in the general population. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **177**, 489–502 (2018).
22. Mollon, J., Almasy, L., Jacquemont, S. & Glahn, D. C. The contribution of copy number variants to psychiatric symptoms and cognitive ability. *Mol. Psychiatry* **28**, 1480–1493 (2023).
23. MacLeod, A. K. *et al.* Genetic copy number variation and general cognitive ability. *PLoS One* **7**, e37385 (2012).
24. Bagshaw, A. T. M. *et al.* No effect of genome-wide copy number variation on measures of intelligence in a New Zealand birth cohort. *PLoS One* **8**, e55208 (2013).
25. van Scheltinga, A. F. T. *et al.* Schizophrenia genetic variants are not associated with intelligence. *Psychol. Med.* **43**, 2563–2570 (2013).
26. Alexander-Bloch, A. *et al.* Copy Number Variant Risk Scores Associated With Cognition,

medRxiv preprint doi: https://doi.org/10.1101/2024.08.02.24311302; this version posted August 6, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Psychopathology, and Brain Structure in Youths in the Philadelphia Neurodevelopmental Cohort. *JAMA Psychiatry* **79**, 699–709 (2022).

27. Huguet, G. *et al.* Genome-wide analysis of gene dosage in 24,092 individuals estimates that 10,000 genes modulate cognitive ability. *Mol. Psychiatry* (2021) doi:10.1038/s41380-020-00985-z.
28. Huguet, G. *et al.* Measuring and Estimating the Effect Sizes of Copy Number Variants on General Intelligence in Community-Based Samples. *JAMA Psychiatry* **75**, 447–457 (2018).
29. Forsyth, J. K. *et al.* Synaptic and Gene Regulatory Mechanisms in Schizophrenia, Autism, and 22q11.2 Copy Number Variant-Mediated Risk for Neuropsychiatric Disorders. *Biol. Psychiatry* **87**, 150–163 (2020).
30. Parikshak, N. N., Gandal, M. J. & Geschwind, D. H. Systems biology and gene networks in neurodevelopmental and neurodegenerative disorders. *Nat. Rev. Genet.* **16**, 441–458 (2015).
31. Kang, H. J. *et al.* Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483–489 (2011).
32. First, Michael B., Spitzer, Robert L, Gibbon Miriam, and Williams, Janet B.W. *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Patient Edition. (SCID-I/P)*. (Biometrics Research, New York State Psychiatric Institute, New York, 2002).
33. Chambers, W. J. *et al.* The assessment of affective disorders in children and adolescents by semistructured interview. Test-retest reliability of the schedule for affective disorders and schizophrenia for school-age children, present episode version. *Arch. Gen. Psychiatry* **42**, 696–702 (1985).
34. Maxwell, M. E. Family Interview for Genetic Studies (FIGS): a manual for FIGS. *, Intramural Research Program, National Institute of ….*
35. Bethlehem, R. A. I. *et al.* Brain charts for the human lifespan. *Nature* **604**, 525–533 (2022).
36. Huang, A. Y. *et al.* Rare Copy Number Variants in NRXN1 and CNTN6 Increase Risk for Tourette Syndrome. *Neuron* **94**, 1101-1111.e7 (2017).
37. Wang, K. *et al.* PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.* **17**, 1665–1674 (2007).

medRxiv preprint doi: https://doi.org/10.1101/2024.08.02.24311302; this version posted August 6, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

38. Colella, S. *et al.* QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. *Nucleic Acids Res.* **35**, 2013–2025 (2007).

39. Sanders, S. J. *et al.* Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron* **87**, 1215–1233 (2015).

40. Firth, H. V. *et al.* DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am. J. Hum. Genet.* **84**, 524–533 (2009).

41. Lowther, C. *et al.* Molecular characterization of NRXN1 deletions from 19,263 clinical microarray cases identifies exons important for neurodevelopmental disease expression. *Genet. Med.* **19**, 53–61 (2017).

42. O'Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733-45 (2016).

43. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).

44. Gogarten, S. M. *et al.* Genetic association testing using the GENESIS R/Bioconductor package. *Bioinformatics* **35**, 5346–5348 (2019).

45. Conomos, M. P., Miller, M. B. & Thornton, T. A. Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness. *Genet. Epidemiol.* **39**, 276–293 (2015).

46. Conomos, M. P., Reiner, A. P., Weir, B. S. & Thornton, T. A. Model-free Estimation of Recent Genetic Relatedness. *Am. J. Hum. Genet.* **98**, 127–148 (2016).

47. Baurley, J. W., Edlund, C. K., Pardamean, C. I., Conti, D. V. & Bergen, A. W. Smokescreen: a targeted genotyping array for addiction research. *BMC Genomics* **17**, 145 (2016).

48. Cooper, G. M. *et al.* A copy number variation morbidity map of developmental delay. *Nat. Genet.* **43**, 838–846 (2011).

49. Kaminsky, E. B. *et al.* An evidence-based approach to establish the functional and clinical significance of copy number variants in intellectual and developmental disabilities. *Genet. Med.* **13**, 777–784 (2011).

medRxiv preprint doi: https://doi.org/10.1101/2024.08.02.24311302; this version posted August 6, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

50. van Erp, T. G. M. *et al.* Cortical Brain Abnormalities in 4474 Individuals With Schizophrenia and 5098 Control Subjects via the Enhancing Neuro Imaging Genetics Through Meta Analysis (ENIGMA) Consortium. *Biol. Psychiatry* (2018) doi:10.1016/j.biopsych.2018.04.023.

51. Alnæs, D. *et al.* Brain Heterogeneity in Schizophrenia and Its Association With Polygenic Risk. *JAMA Psychiatry* (2019) doi:10.1001/jamapsychiatry.2019.0257.

52. Sønderby, I. E. *et al.* Effects of copy number variations on brain structure and risk for psychiatric illness: Large-scale studies from the ENIGMA working groups on CNVs. *Hum. Brain Mapp.* **43**, 300–328 (2022).

53. Morris, E., Inglis, A. & Austin, J. Psychiatric genetic counseling for people with copy number variants associated with psychiatric conditions. *Clin. Genet.* **102**, 369–378 (2022).

54. Bergsland, M., Werme, M., Malewicz, M., Perlmann, T. & Muhr, J. The establishment of neuronal properties is controlled by Sox4 and Sox11. *Genes Dev.* **20**, 3475–3486 (2006).

55. Hoshiba, Y. *et al.* Sox11 Balances Dendritic Morphogenesis with Neuronal Migration in the Developing Cerebral Cortex. *J. Neurosci.* **36**, 5775–5784 (2016).

56. Zhou, Q. *et al.* Inhibition of the histone demethylase Kdm5b promotes neurogenesis and derepresses Reln (reelin) in neural stem cells from the adult subventricular zone of mice. *Mol. Biol. Cell* **27**, 627–639 (2016).

57. Limoni, G. & Niquille, M. Semaphorins and Plexins in central nervous system patterning: the key to it all? *Curr. Opin. Neurobiol.* **66**, 224–232 (2021).

58. Ruediger, T. *et al.* Integration of opposing semaphorin guidance cues in cortical axons. *Cereb. Cortex* **23**, 604–614 (2013).

59. Ito, H. *et al.* Biochemical and Morphological Characterization of a Neurodevelopmental Disorder-Related Mono-ADP-Ribosylhydrolase, MACRO Domain Containing 2. *Dev. Neurosci.* **40**, 278–287 (2018).

60. Wang, T., Zhao, P. A. & Eichler, E. E. Rare variants and the oligogenic architecture of autism. *Trends Genet.* **38**, 895–903 (2022).

61. Yang, P. *et al.* A placental growth factor is silenced in mouse embryos by the zinc finger protein

medRxiv preprint doi: https://doi.org/10.1101/2024.08.02.24311302; this version posted August 6, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

ZFP568. *Science* **356**, 757–759 (2017).

62. Chien, H.-C. *et al.* Targeted disruption in mice of a neural stem cell-maintaining, KRAB-Zn finger-encoding gene that has rapidly evolved in the human lineage. *PLoS One* **7**, e47481 (2012).
63. Al-Naama, N., Mackeh, R. & Kino, T. C2H2-Type Zinc Finger Proteins in Brain Development, Neurodevelopmental, and Other Neuropsychiatric Disorders: Systematic Literature-Based Analysis. *Front. Neurol.* **11**, 32 (2020).
64. Ghosh, C. *et al.* Involvement of Cdkal1 in the etiology of type 2 diabetes mellitus and microvascular diabetic complications: a review. *J. Diabetes Metab. Disord.* **21**, 991–1001 (2022).
65. Torres, A. G., Batlle, E. & Ribas de Pouplana, L. Role of tRNA modifications in human diseases. *Trends Mol. Med.* **20**, 306–314 (2014).
66. Plomin, R. & Deary, I. J. Genetics and intelligence differences: five special findings. *Mol. Psychiatry* **20**, 98–108 (2015).
67. Haworth, C. M. A. *et al.* The heritability of general cognitive ability increases linearly from childhood to young adulthood. *Mol. Psychiatry* **15**, 1112–1120 (2010).
68. Davis, O. S. P., Haworth, C. M. A. & Plomin, R. Dramatic increase in heritability of cognitive development from early to middle childhood: an 8-year longitudinal study of 8,700 pairs of twins. *Psychol. Sci.* **20**, 1301–1308 (2009).
69. Casey, B. J., Tottenham, N., Liston, C. & Durston, S. Imaging the developing brain: what have we learned about cognitive development? *Trends Cogn. Sci.* **9**, 104–110 (2005).
70. Luna, B., Garver, K. E., Urban, T. A., Lazar, N. A. & Sweeney, J. A. Maturation of cognitive processes from late childhood to adulthood. *Child Dev.* **75**, 1357–1372 (2004).
71. de Bartolomeis, A. *et al.* Differential cognitive performances between schizophrenic responders and non-responders to antipsychotics: correlation with course of the illness, psychopathology, attitude to the treatment and antipsychotics doses. *Psychiatry Res.* **210**, 387–395 (2013).
72. Millgate, E. *et al.* Neuropsychological differences between treatment-resistant and treatment-responsive schizophrenia: a meta-analysis. *Psychol. Med.* **52**, 1–13 (2022).