medRxiv preprint doi: https://doi.org/10.1101/2024.07.16.24310489; this version posted July 16, 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 The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

CLINICAL AND COGNITIVE PHENOTYPING OF COPY NUMBER VARIANTS PATHOGENIC FOR NEURODEVELOPMENTAL DISORDERS FROM A MULTI-ANCESTRY BIOBANK

AUTHORS:

Nina Zaks[1], Behrang Mahjani[2,3], Abraham Reichenberg[2,3,4], Rebecca Birnbaum[2,5]

AFFILIATIONS:

1. Department of Child and Adolescent Psychiatry, NYU Langone Health
2. Department of Psychiatry, Icahn School of Medicine at Mount Sinai
3. Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai
4. Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai
5. Department of Genetic and Genomics Sciences, Icahn School of Medicine at Mount Sinai

Correspondence:
Rebecca Birnbaum, MD
Departments of Psychiatry and Genetics and Genomics Sciences
Icahn School of Medicine at Mount Sinai
1399 Park Avenue, Room 305B
New York, NY 10029
212-585-4657
rebecca.birnbaum@mssm.edu

MANUSCRIPT WORD COUNT: 3,624

KEYWORDS: Copy Number Variants, Neurodevelopmental Disorders, Schizophrenia, Neurocognition, Biobank

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.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/10.1101/2024.07.16.24310489; this version posted July 16, 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 The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

ABSTRACT

Background: Rare copy number variants (CNVs) are pathogenic for neurodevelopmental disorders (NDDs) and effect neurocognitive impairment. In aggregate, NDD CNVs may present in up to 2% of population cohorts with implications for neuropsychiatric disease risk and cognitive health. However, analyses of NDD CNVs in biobanks or population cohorts have been hindered by limited clinical or cognitive phenotypes, and a lack of ancestral diversity. In the current proof-of-concept study, NDD CNV carriers were recontacted from Bio*Me*, a multi-ancestry biobank derived from the Mount Sinai healthcare system, to enable 'deep phenotyping' beyond electronic health record outcomes.

Methods: From Bio*Me* biobank, 892 adult participants were recontacted, including 335 harboring NDD CNVs, 217 with schizophrenia and 340 neurotypical controls as comparators. Clinical and cognitive assessments were administered to each recruited participant.

Results: Seventy-three participants completed study assessments (mean age=48.8 years; 66% female; 36% African, 26% European, 34% Hispanic), or 8% of the recontacted subset, including 30 NDD CNV carriers across 15 loci. Among NDD CNV carriers, assessments indicated 40% with mood and anxiety disorders, 30% with learning disorders, and 13% with a history of special education. NDD CNV carriers were significantly cognitively impaired compared to controls on digit span backwards (Beta=-1.76, FDR=0.04) and digit span sequencing (Beta=-2.01, FDR=0.04).

Conclusions: Feasibility of "recall-by-genotype" from a multi-ancestry biobank was established for NDD CNV carriers, along with comparator groups. The current study corroborated past reports of NDD CNVs effects of cognitive impairment, while elucidating clinical phenotypes for recalled individuals. Future "recall-by-genotype" studies may further facilitate clinical characterization of disease-relevant genomic variants.

medRxiv preprint doi: https://doi.org/10.1101/2024.07.16.24310489; this version posted July 16, 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. It is made available under a CC-BY-NC-ND 4.0 International license .

INTRODUCTION

Rare copy number variants (CNVs), genomic microdeletions or microduplications greater than one kilobase, are known to be pathogenic for multiple neurodevelopmental disorders (NDDs), including schizophrenia, autism spectrum disorder, intellectual disability, and developmental delay.[1-5] NDD CNVs are variable in penetrance and expressivity, ranging from unaffected to severely affected, with reports indicating that genome-wide polygenic risk and CNV burden may influence penetrance.[6-8] Overall, pathogenic NDD CNVs likely underlay up to 20% of ASD cases, 14% of cases of developmental delay and intellectual disability, and a smaller fraction, 1-3%, of schizophrenia cases.[9-13]

NDD CNVs are also known to effect neurocognition, a quantitative trait, that may independently result in functional impairments and mediate neuropsychiatric disease risk.[9, 14] Variable neurocognitive impairments of NDD CNV carriers have been reported in clinically-ascertained cohorts, including for 22q11.2 deletion underlying velocardiofacial syndrome, perhaps the most investigated NDD CNV.[14, 15] Yet, population-cohort and biobank studies have demonstrated that NDD CNVs also effect neurocognition in neurotypical, healthy controls.[16-18] Within an Icelandic population cohort, 167 controls harboring neuropsychiatric CNVs were impaired in cognitive performance across neurocognitive domains, intermediary between controls without CNVs and individuals with idiopathic schizophrenia.[16] Effect size varied by CNV locus, and six NDD CNV loci were associated with verbal or performance IQ with large effects (0.73-3.51SD). A more recent UK biobank analysis of ~150,000 participants queried the effect of NDD CNVs at 53 loci on seven cognitive tests, finding impairment across cognitive domain, with a reduction from 0.1 to 0.5SD compared to non-NDD CNV carriers, varying by CNV locus.[17] In a more recent analysis of ~420,000 UK biobank participants without a neurodevelopmental disorder, 24 of 33 NDD CNVs were associated with reduced performance on at least one of seven cognitive tests, including processing speed, general intelligence, working memory and executive function.[18]

Though individual NDD CNVs are rare, their combined prevalence may be up to 2% of population cohorts.[17-21] Therefore, in populations unselected for neuropsychiatric disorders, studies of neurocognitive effects of NDD CNVs may have important public health implications for cognitive health, including early cognitive interventions, and may elucidate factors mediating neuropsychiatric disease risk. However, previous studies of neurocognition of NDD CNVs in research biobanks derived from healthcare systems or population cohorts have several limitations. First, many studies include individuals of mostly European ancestry, precluding generalizability and limiting elucidation of cognitive effects in diverse ancestries.[16-18] Second, some neurocognitive domains have not been assessed, such as social cognition, which encompasses mentalizing, cognitive empathy, emotional perception and processing, often impaired in multiple NDDs, contributing to unique variance in functional outcomes.[22, 23] Third, many biobanks or population cohorts do not contain neurocognitive data, and the feasibility of recontacting individuals for secondary assessments has not been clearly established.

A recent analysis of the Bio*Me* biobank, a multi-ancestry biobank derived from the Mount Sinai healthcare system, with genotype data linked to electronic health records (EHR) of primarily adults with a mean age of 50 years, reported 2.5% prevalence of NDD CNVs among approximately

medRxiv preprint doi: https://doi.org/10.1101/2024.07.16.24310489; this version posted July 16, 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. It is made available under a CC-BY-NC-ND 4.0 International license .

25,000 participants.[24] For NDD CNV carriers, enrichments for congenital disorders and major depressive disorder were identified, as well as associations with obesity and increased body mass index. However, the previous report was limited to EHR outcomes, so some phenotypes relevant to NDD CNVs were not reported, including neurocognition, diagnoses from structured scales, and retrospective developmental histories. Therefore, in the current study, adult NDD CNV carriers were recalled from the Bio*Me* biobank, by implementing a novel, 'genotype-first' recruitment strategy, for expanded 'deep phenotyping'. The clinical and neurocognitive assessments of NDD CNV carriers were compared to individuals with idiopathic schizophrenia and neurotypical controls, each without NDD CNVs, also recalled from the biobank and included as comparators, as per previous reports.[16-18]

METHODS AND MATERIALS

IRB Approval: The Bio*Me* Biobank is an EHR-linked biobank of ~55,000 participants derived from the Mount Sinai Health System (New York, NY), with recruitment from clinics across various specialties since 2007, predominantly outpatient medical clinics, as approved by the Institutional Review Board (IRB) of the Mount Sinai School of Medicine.[24, 25] As per initial IRB-approval, Bio*Me* participants consented to be 'recontacted' for future research studies. For the current study, the IRB approved Bio*Me* data access and chart review, and recontact of participants for study assessments, but without return of genetic information nor neurocognitive results. The current study was conducted from November 12, 2020, until March 23, 2023.

Study Recruitment: Within the Bio*Me* biobank, three groups of adults, ages 18-65 were identified for recontact with the following criteria: (i) NDD-CNVs: Harboring NDD CNVs.[24] (ii) Schizophrenia: At least two documented International Classification of Diseases (ICD10) codes for schizophrenia or schizoaffective disorder, and without NDD CNVs. (iii) Neurotypical controls: Without ICD-10 code for neuropsychiatric nor neurodevelopmental disorders and without NDD CNVs. An upper age limit of age 65 was applied to minimize age-related cognitive decline.

Biobank participants in each group were contacted by mail, email or phone (Supplementary Methods). Individuals who responded were screened for study eligibility, with three exclusion criteria: 1) active, unstable, severe medical illness for which there is no current medical care; 2) substance use disorder or alcohol use disorder, active in the past 6 months; 3) a known medical or neurological condition that could affect cognition. Prior to study assessment, chart review of medical and psychiatric history was performed (by board-certified psychiatrist, RB), to confirm schizophrenia or control status, and to phenotype NDD CNV carriers, whether unaffected or affected by neuropsychiatric disorder, as per accessible EHR records (Supplementary Methods).

Clinical and Cognitive Assessments: Due to the COVID-19 pandemic, study assessments were performed remotely, rather than in-person (Supplementary Methods). A clinical assessment included the Mini International Neuropsychiatric Interview (MINI v 7.0.2) brief structured diagnostic interview for DSM-V, as well as a medical, psychiatric, and retrospective developmental history (administered by one board-certified psychiatrist, RB).[26] A cognitive assessment (administered by NZ) included multiple cognitive domains (Table 1, Supplementary Methods). The reliability and validity of each cognitive test as well as evidence to support remote administration has been previously demonstrated for use in schizophrenia spectrum disorders and in other populations (Supplementary Methods).[27-34]

medRxiv preprint doi: https://doi.org/10.1101/2024.07.16.24310489; this version posted July 16, 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. It is made available under a CC-BY-NC-ND 4.0 International license .

Statistical Analyses: Cognitive raw scores were used as outcomes for linear regression analyses and group comparisons, with age, sex and ancestry included as covariates. All analyses were performed using R, version 4.0.4. Statistical significance was reported using the Benjamini-Hochberg false discovery rate (FDR) for multiple testing.

RESULTS

Study Recruitment: From the Bio*Me* biobank, 892 participants were recontacted for study participation (Figure 1): 335 individuals harboring NDD CNVs, 217 individuals with schizophrenia, and 340 neurotypical controls. There was an initial response rate of 18% (15% of NDD CNV carriers, 20% of individuals with schizophrenia, 19% of controls), a yield of 160 participants. However, 53 individuals were not screened for eligibility, due to subsequent lack of response or an expressed lack of interest, including 26 from the control group. Of 107 individuals who were screened for eligibility, 13% were identified to be ineligible, due to active substance use, or illness that could confound cognitive measures, with the highest rate of exclusion within the schizophrenia group. A total of 93 individuals were consented, of which an additional 20 were excluded after clinical assessment, 18 due to group diagnostic misclassification (study assessment did not corroborate chart review). Thus, 73 individuals, or 8% of the overall recontacted subset, were retained and completed the study, including 30 NDD CNV carriers (9% of recontacted), 20 individuals with schizophrenia (9.2% of recontacted), and 23 neurotypical controls (6.8% of recontacted).

Utilization of the healthcare system did not bias recruitment, an average of 125 encounters for the study sample, compared to an average of 133 encounters for others recontacted (t=0.30, p=0.76). Proximity to the most recent clinical encounter correlated with successful recruitment, as enrolled study participants were seen on average 12 weeks prior to recontact, compared to the recontacted subset not enrolled, who had a clinical encounter on average 57 weeks prior to recontact (t=7.1, $p=1.1x10^{-10}$). A female skew in the overall biobank (62% female) increased for the recontacted NDD CNV carriers (65% female) and further, for the 30 NDD CNV carriers recruited (87% female) (Table 2, Supplementary Table 1). There was an ancestry skew in the those recontacted for the current study compared to the overall biobank, within the schizophrenia group, relatively increased in African (47%) and decreased in European (8%) ancestries, and the control group, relatively decreased in European (15%) and increased in Hispanic (40%) (Supplementary Table 1).

Of 73 participants who completed study assessments, with a mean age of 48.8 years, there was no significant difference in age across the three comparator groups (p=0.58) (Table 2, Supplementary Table 1). In contrast to the NDD CNV group (87% female), the schizophrenia group was mostly male (45% female), resulting in a significant difference in sex across the groups (p=0.003). Self-reported ancestry (overall 36% African, 26% European, 34% Hispanic) did not differ significantly across groups overall (p=0.55) or by ancestry group, African (p=0.68), European (p=0.16) and Hispanic (p=0.49).

Study Sample, NDD-CNV Status: Among 30 NDD CNV carriers, 16 harbored duplications and 14 harbored deletions, across 15 unique NDD CNV loci (Supplementary Table 2). Some NDD CNV loci

medRxiv preprint doi: https://doi.org/10.1101/2024.07.16.24310489; this version posted July 16, 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. It is made available under a CC-BY-NC-ND 4.0 International license .

were enriched in the subset that completed assessments compared to the overall biobank including for example, TAR duplication, 15q11.2 deletion and 16p13.11 deletion (Supplementary Table 2). None of the participants with schizophrenia nor neurotypical controls harbored NDD CNVs.

Clinical Assessments (Table 3, Supplementary Table 3): Of 30 NDD CNV carriers, twelve had a mood or anxiety disorder upon initial chart review. The MINI scale did not corroborate EHR-derived diagnoses for three individuals but identified mood or anxiety disorders in an alternative three individuals unreported in EHRs. Therefore overall, 40% of NDD CNV carriers had a mood or anxiety disorder, as validated by MINI, including five with major depressive disorder and two with obsessive compulsive disorder (Supplementary Table 3). One TAR duplication carrier had bipolar disorder with psychotic features, but there were no other cases of psychosis nor schizophrenia among the NDD CNV carriers. One 16p11.2 distal deletion carrier had a history of seizure disorder. Childhood speech delay warranting speech therapy was reported in three NDD CNV carriers, and one 1q21.1 deletion carrier reported global developmental delay including motor skills of walking. Further, formally diagnosed learning disorders were reported by approximately 30% of NDD CNV carriers, and a history of special education was reported for 13%, from elementary school through high school, across multiple years and multiple subject areas.

For 20 individuals with schizophrenia, MINI scale administration confirmed a psychotic disorder. One participant with schizophrenia reported a history of seizure disorder. Fifty percent reported a history of special education and an additional 10% reported extreme difficulty in school, but no diagnosed learning disorder. Twenty percent reported speech therapy during childhood, and one individual reported developmental delay of motor skills, 'walking late'. For 23 neurotypical controls, MINI scale corroborated absence of major psychiatric disorders. None had a history of seizure disorder. There was no history of developmental delay, nor history of speech therapy during childhood.

Overall, the highest level of education differed across group, schizophrenia with the lowest mean of 11.9 years, NDD CNV carriers intermediate with 14.5 years, and controls with the highest of 15.3 years (F-statistic=6.24, p=0.003) with pairwise significant differences between schizophrenia and NDD CNVs (p=0.02), and schizophrenia and controls (p=0.003), but not between NDD CNVs and controls (p=0.68). As for occupational status, within the NDD CNV and control groups, each participant reported a history of employment without disability, however 20% within the schizophrenia group reported disability and lack of employment. All NDD CNV carriers and controls reported the ability to live independently, but the schizophrenia group had a lower functional status, with 20% living in supervised residences or unable to live independently.

The study also elucidated medical phenotypes qualitatively, by EHR review and clinical interview (Supplementary Table 4), in contrast to a previous ICD-code based analysis.[24] Some medical phenotypes known to be significant from previous biobank and case reports were observed: obesity (TAR duplication and 16p11.2 distal deletion), hypertension (16p13.11 duplication and 16p11.2 deletion), and neuropathy (17p12 (HNPP) deletion), as well as cardiac anomaly (22q12. deletion). Three NDD CNV carriers reported congenital anomalies, two cardiac anomalies

medRxiv preprint doi: https://doi.org/10.1101/2024.07.16.24310489; this version posted July 16, 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. It is made available under a CC-BY-NC-ND 4.0 International license .

(22q11.2 deletion and 2q13 (NPHP1) deletion) and one renal anomaly (15q11.2 deletion), none within schizophrenia and controls. No participants in the current study were of short stature.

Neurocognitive Assessments: To maximize power, NDD CNV carriers were collapsed across loci, into one group. Each of seven cognitive tests was regressed against group while adjusting for age, sex, and ancestry. There were no significant differences between NDD CNV carriers and controls (when aggregating across all 15 NDD CNV loci), though for each of seven tests, NDD CNV carriers trended in intermediate directionality, impaired compared to controls and higher performing than schizophrenia (Figure 2A). The schizophrenia group performed significantly worse in cognitive performance compared to the NDD CNV group on the digit span forward (Beta=-2.04, FDR=0.03), digit span backwards (Beta=-1.96, FDR=0.03), and HVLT-R, Total Recall (Beta=-5.04, FDR=0.01) (Figure 2A, Supplementary Table 5). The schizophrenia group also performed significantly worse than controls on digit span forward (Beta=-3.10, FDR= 4.56x10-4), digit span backwards (Beta=-3.24, FDR= 2.79x10-4), digit span sequencing (Beta=-3.07, FDR= 4.56x10-4), and HVLT-R Total Recall (Beta=-6.12, FDR= 4.56x10-4), (Supplementary Table 5). The performance of each NDD CNV carrier across seven cognitive domains is summarily ranked (Supplementary Table 6, Supplementary Figure 1), with varying cognitive performance by locus, notably the 22q11.2 deletion carrier among the worst performing of the NDD CNV group and 1q21.1 duplication the highest performing.

In an alternative analysis, the NDD CNV group was subset to loci included in previously-reported UK Biobank (UKBB) analyses of NDD CNVs, thereby excluding four loci (15q13.3 (CHRNA7) duplication, 17p12 deletion, and 2q13 NPHP1) deletion/duplication), resulting in 22 NDD CNV carriers of 11 unique loci (Supplementary Table 7).[17, 18] Cognitive analyses of this more stringent subset of 22 NDD CNV carriers again found intermediate performance, with controls performing significantly higher than NDD CNVs on two tests: digit span backwards (Beta=1.76, FDR=0.04) and digit span sequencing (Beta=2.01, FDR=0.04), and schizophrenia significantly impaired compared to NDD CNVs on HVLT-R Total Immediate Recall (Beta=-4.5, FDR=0.05) (Figure 2B, Supplementary Table 5).

In the current sample, ancestry was correlated with social cognition (r^2=0.36, p=6.5x10^{-7}) and sex correlated with both social cognition (r^2=0.12, p=0.003) and verbal learning (r^2=0.11, p=0.004) (Supplementary Figure 2). Including education as an additional covariate in the analyses, which is correlated with group status as well, reduced the significance of the between group differences for each cognitive test, while maintaining the trend in directionality of between group effects (Supplementary Table 8).

DISCUSSION

The current report describes a novel, genotype-first, targeted recruitment of NDD CNV carriers recalled from Bio*Me*, a multi-ancestry, healthcare-system derived biobank, to enable 'deep phenotyping' beyond EHR outcomes. Notably to date, the advent of research biobanks, with genotype data linked to EHRs, have yielded numerous *in silico* genomic and phenotypic analyses, as well as the development of novel analytic methods; in contrast, however, "recall-by-genotype" studies have been exceedingly rare despite their potential importance in elucidating genomic risk

medRxiv preprint doi: https://doi.org/10.1101/2024.07.16.24310489; this version posted July 16, 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. It is made available under a CC-BY-NC-ND 4.0 International license .

factors.[35-38] As research biobanks (from healthcare-systems and population cohorts) continue to proliferate and increase in scale, opportunities for "recall-by-genotype" will become increasingly apparent across disease and variant class, including neuropsychiatry, though feasibility has not yet been established. In the current, proof-of-concept "recall-by-genotype" study, 15% of recontacted adult NDD CNV carriers responded to recruitment, as well as 20% of recontacted individuals with schizophrenia and 19% of recontacted controls. Within the NDD CNV group, more females than males responded to recontact, furthering a sex bias in biobank participation. Subsequent dropout and study exclusions reduced the study sample to 8% of the initially recontacted subset, including 9% of the initially contacted NDD CNV group. The recontact rates were higher than predicted, given that the healthcare system from which the biobank is derived is not contained, that is biobank participants may receive care from other healthcare systems or relocate geographically after biobank enrollment.

To date, the few published biobank "recall-by-genotype" reports have focused on pathogenic variants for diseases for which actionable therapeutic or prophylactic options may be available, such as for American College of Medical Genetics and Genomics (ACMG) secondary or other findings.[39] For example, 21 Estonia biobank participants who harbor rare, deleterious variants pathogenic for familial hypercholesterolemia within *LDLR, APOB,* or *PCSK9* genes, were recontacted for counseling, intervention with statin treatment and longitudinal follow-up.[40, 41] In another study germane to the current report, the prevalence and penetrance of 31 CNVs pathogenic for neuropsychiatric disorders were analyzed in the Geisinger *MyCode* biobank, among 90,595 individuals, mostly adults of European ancestry. Though neurocognitive outcomes were not within the EHRs, 66.4% of CNV carriers having at least one EHR code for a neuropsychiatric disorder, including mood and anxiety disorders.[20] The genetic findings of CNVs for nine loci were returned to a subset of 141 Geisinger *MyCode* participants, with overall positive reactions reported to the disclosure of this genetic information. Less common however, such as herein described, are "recall-by-genotype" studies for secondary research, without immediate actionability or specific return of results. Notably the current pilot study *did not* return CNV results to adult participants, in contrast to the Geisinger *MyCode* CNV study, as per local IRB protocol. Additional studies are needed to further benchmark "recall-by-genotype", to further elucidate feasibility in recontact, including limitations, as well as alternative approaches for genetic disclosure. For example, future "recall-by-genotype" study designs may recall participants for longitudinal follow-up studies, rather than cross-sectional, as herein described, or recontact family members for further query of genomic variants of interest.

The clinical phenotypes of 30 NDD CNV carriers across 15 unique CNV loci was under-powered for statistical associations (as previously reported for ICD codes for NDD CNV carriers in this biobank), but rather, permitted an in-depth, qualitative evaluation. Overall, 70% of the 30 NDD-CNV carriers harbored at least one neuropsychiatric or developmental phenotype, including 40% with mood or anxiety disorders, 13% with speech or motor developmental delays, 27% with learning disorders, and 13% with a history of special education. There was one case of bipolar disorder with psychosis, but no other psychotic disorders identified among the NDD CNV carriers, even for carriers of the CNV loci reported to confer schizophrenia risk at especially high effect (i.e. 1q21.1 deletion, 16p11.2distal_deletion, 22q11.2 deletion). Overall these findings

medRxiv preprint doi: https://doi.org/10.1101/2024.07.16.24310489; this version posted July 16, 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. It is made available under a CC-BY-NC-ND 4.0 International license .

corroborate past reports of variable penetrance of NDD CNVs, and also highlights the potential role of NDD CNVs in mood and anxiety disorders, suggested in some past reports, but with less well-established evidence than for neurodevelopmental disorders.[9, 42, 43] Interestingly, half of the mood or anxiety disorders identified during the study assessment were discordant with EHRs, and furthermore EHRs did not contain the developmental history elicited. This suggests potential shortcomings of solely relying on EHR-data for characterizing genetic variants within biobanks and indicates a role for recall of individuals for supplemental clinical evaluations.

Corroborating past reports, cognitive analyses indicated that NDD CNV carriers were significantly impaired compared to neurotypical controls on tests of executive function and working memory (digit span backwards and digit span sequencing), but higher performing than individuals with schizophrenia on immediate recall (verbal learning).[16-18] As a validation of study measures, individuals with schizophrenia were significantly impaired compared to neurotypical controls on tests of working memory and executive function, attention and immediate recall, concordant with reports of individuals with schizophrenia performing (on average one standard deviation) below controls.[16, 44] The cognitive analyses of a multi-ancestry cohort was more diverse than many past reports of NDD CNV carriers of predominantly European ancestry.[16-18] The study also tested social cognition (not included in previous UKBB analyses), albeit no group differences were detected. A modest correlation of social cognition with ancestry was observed. Education level differed across groups, as expected, a confounding variable associated with group status and cognitive outcome. Comparatively, previously reported UKBB analyses did not include years of education as a covariate in analyzing cognitive effects of NDD CNVs.[17, 18]

This study had several limitations. The clinical assessments were qualitative, under-powered for statistical associations, or for evaluation of NDD CNV pathogenicity. While the cognitive analyses yielded some significant between group differences, other differences may not have been detected due to sample size. Sex and ancestry were included as covariates in pooled analyses, but the study was not powered for sex or ancestry stratified analyses. Remote assessments were conducted due to the COVID pandemic, so future recall studies may incorporate in-person assessments including physical or neurological exams, and further phenotyping. Population normative data for cognitive tests were derived from in-person testing. Some of the clinical assessments relied on participant self-report, subject to recall bias. The study excluded some conditions known to result in cognitive impairment, but did not control for other potential confounders, such as psychotropic medication burden. Cognitive analyses included a subset of known, recurrent NDD CNV loci, but overall genome-wide CNV burden, rare non-recurrent CNVs was not considered. Cognitive analyses were not stratified by CNV locus, and included individuals both affected and unaffected by neuropsychiatric and developmental phenotypes.

Overall, the current study demonstrated feasibility of “recall-by-genotype” from a healthcare system derived biobank, for individuals harboring NDD CNVs of rare frequency, as well as individuals from comparator groups, schizophrenia and controls, while identifying some factors influencing recruitment, such as time lapse since most recent healthcare encounter. The proof-of-concept, recall study supplements past biobank and other reports of NDD CNVs, by expanding the range of phenotypes assessed, including retrospective developmental history and cognitive

medRxiv preprint doi: https://doi.org/10.1101/2024.07.16.24310489; this version posted July 16, 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. It is made available under a CC-BY-NC-ND 4.0 International license .

data, in a multi-ancestry context. Interestingly, the current "deep phenotyping" identified clinical features for some participants that were discordant with EHR-derived diagnoses. Cognitive results corroborated past reports, and with NDD CNVs across multiple loci effecting impairment of executive function and working memory. Future studies may further implement "recall-by-genotype" strategies in sequenced cohorts to identify individuals harboring NDD CNVs or other pathogenic variants, for more thorough clinical characterization and follow-up research opportunities.

medRxiv preprint doi: https://doi.org/10.1101/2024.07.16.24310489; this version posted July 16, 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. It is made available under a CC-BY-NC-ND 4.0 International license .

ACKNOWLEDGEMENTS:
The study was supported by K23MH112955 (PI: Birnbaum). Dr. Birnbaum is also supported by R21MH137536. Dr. Mahjani is supported by a grant from the Beatrice and Samuel A. Seaver Foundation.

We thank individuals within the Bio*Me* Biobank for their participation. We thank colleagues from the Institute of Personalized Medicine at the Icahn School of Medicine for facilitating recall of the BioMe biobank participants: Amanda Merkelson, Sheryl Cruz and Alanna Gomez. We are grateful Dr. Zhongyang Zhang for his perusal of the manuscript and comments.

DISCLOSURES:
The authors report no biomedical financial interests or potential conflicts of interest.

medRxiv preprint doi: https://doi.org/10.1101/2024.07.16.24310489; this version posted July 16, 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. It is made available under a CC-BY-NC-ND 4.0 International license .

REFERENCES

1. Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, et al. Strong association of de novo copy number mutations with autism. Science. 2007;316(5823):445-9.
2. Marshall CR, Howrigan DP, Merico D, Thiruvahindrapuram B, Wu W, Greer DS, et al. Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. Nat Genet. 2017;49(1):27-35.
3. Pinto D, Delaby E, Merico D, Barbosa M, Merikangas A, Klei L, et al. Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. Am J Hum Genet. 2014;94(5):677-94.
4. Malhotra D, Sebat J. CNVs: harbingers of a rare variant revolution in psychiatric genetics. Cell. 2012;148(6):1223-41.
5. Mulle JG, Sullivan PF, Hjerling-Leffler J. Editorial overview: Rare CNV disorders and neuropsychiatric phenotypes: opportunities, challenges, solutions. Curr Opin Genet Dev. 2021;68:iii-ix.
6. Kirov G. CNVs in neuropsychiatric disorders. Hum Mol Genet. 2015;24(R1):R45-9.
7. Tansey KE, Rees E, Linden DE, Ripke S, Chambert KD, Moran JL, et al. Common alleles contribute to schizophrenia in CNV carriers. Molecular Psychiatry. 2016;21(8):1085-9.
8. Girirajan S, Rosenfeld JA, Coe BP, Parikh S, Friedman N, Goldstein A, et al. Phenotypic heterogeneity of genomic disorders and rare copy-number variants. N Engl J Med. 2012;367(14):1321-31.
9. Mollon J, Almasy L, Jacquemont S, Glahn DC. The contribution of copy number variants to psychiatric symptoms and cognitive ability. Mol Psychiatry. 2023;28(4):1480-93.
10. Sanders SJ, He X, Willsey AJ, Ercan-Sencicek AG, Samocha KE, Cicek AE, et al. Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. Neuron. 2015;87(6):1215-33.
11. Jiang YH, Yuen RK, Jin X, Wang M, Chen N, Wu X, et al. Detection of clinically relevant genetic variants in autism spectrum disorder by whole-genome sequencing. Am J Hum Genet. 2013;93(2):249-63.
12. Tammimies K, Marshall CR, Walker S, Kaur G, Thiruvahindrapuram B, Lionel AC, et al. Molecular Diagnostic Yield of Chromosomal Microarray Analysis and Whole-Exome Sequencing in Children With Autism Spectrum Disorder. Jama. 2015;314(9):895-903.
13. Miller DT, Adam MP, Aradhya S, Biesecker LG, Brothman AR, Carter NP, et al. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. Am J Hum Genet. 2010;86(5):749-64.
14. Molloy CJ, Quigley C, McNicholas Á, Lisanti L, Gallagher L. A review of the cognitive impact of neurodevelopmental and neuropsychiatric associated copy number variants. Transl Psychiatry. 2023;13(1):116.
15. Moberg PJ, Richman MJ, Roalf DR, Morse CL, Graefe AC, Brennan L, et al. Neurocognitive Functioning in Patients with 22q11.2 Deletion Syndrome: A Meta-Analytic Review. Behav Genet. 2018;48(4):259-70.
16. Stefansson H, Meyer-Lindenberg A, Steinberg S, Magnusdottir B, Morgen K, Arnarsdottir S, et al. CNVs conferring risk of autism or schizophrenia affect cognition in controls. Nature. 2014;505(7483):361-6.

medRxiv preprint doi: https://doi.org/10.1101/2024.07.16.24310489; this version posted July 16, 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. It is made available under a CC-BY-NC-ND 4.0 International license .

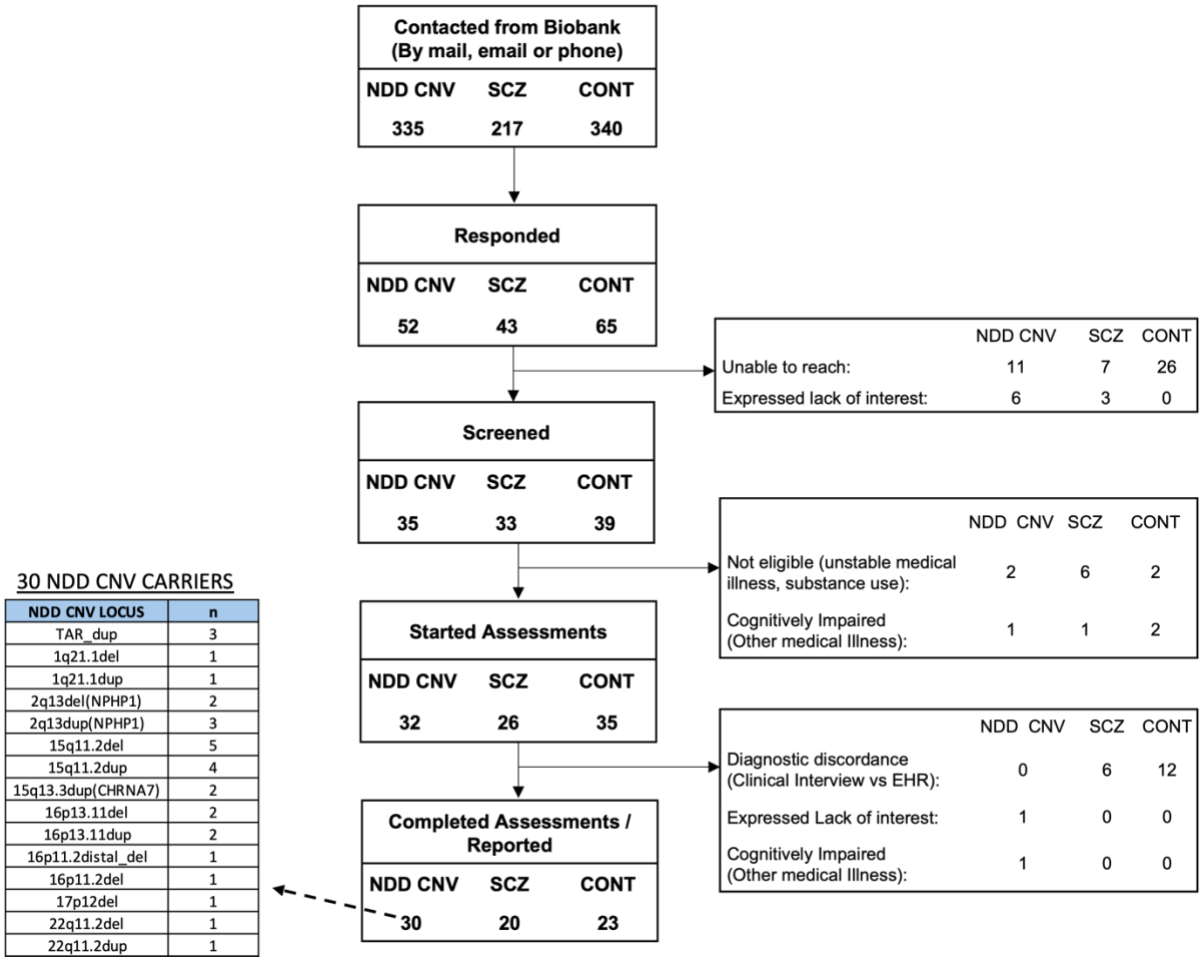
17. Kendall KM, Rees E, Escott-Price V, Einon M, Thomas R, Hewitt J, et al. Cognitive Performance Among Carriers of Pathogenic Copy Number Variants: Analysis of 152,000 UK Biobank Subjects. Biol Psychiatry. 2017;82(2):103-10.
18. Kendall KM, Bracher-Smith M, Fitzpatrick H, Lynham A, Rees E, Escott-Price V, et al. Cognitive performance and functional outcomes of carriers of pathogenic copy number variants: analysis of the UK Biobank. Br J Psychiatry. 2019;214(5):297-304.
19. Crawford K, Bracher-Smith M, Owen D, Kendall KM, Rees E, Pardiñas AF, et al. Medical consequences of pathogenic CNVs in adults: analysis of the UK Biobank. J Med Genet. 2019;56(3):131-8.
20. Martin CL, Wain KE, Oetjens MT, Tolwinski K, Palen E, Hare-Harris A, et al. Identification of Neuropsychiatric Copy Number Variants in a Health Care System Population. JAMA Psychiatry. 2020;77(12):1276-85.
21. Vaez M, Montalbano S, Calle Sánchez X, Georgii Hellberg K-L, Dehkordi SR, Krebs MD, et al. Population-Based Risk of Psychiatric Disorders Associated With Recurrent Copy Number Variants. JAMA Psychiatry. 2024.
22. Lemmers-Jansen I, Velthorst E, Fett AK. The social cognitive and neural mechanisms that underlie social functioning in individuals with schizophrenia - a review. Transl Psychiatry. 2023;13(1):327.
23. Velikonja T, Fett AK, Velthorst E. Patterns of Nonsocial and Social Cognitive Functioning in Adults With Autism Spectrum Disorder: A Systematic Review and Meta-analysis. JAMA Psychiatry. 2019;76(2):135-51.
24. Birnbaum R, Mahjani B, Loos RJF, Sharp AJ. Clinical Characterization of Copy Number Variants Associated With Neurodevelopmental Disorders in a Large-scale Multiancestry Biobank. JAMA Psychiatry. 2022;79(3):250-9.
25. Belbin GM, Cullina S, Wenric S, Soper ER, Glicksberg BS, Torre D, et al. Toward a fine-scale population health monitoring system. Cell. 2021;184(8):2068-83.e11.
26. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry. 1998;59 Suppl 20:22-33.
27. Russell MT, Funsch KM, Springfield CR, Ackerman RA, Depp CA, Harvey PD, et al. Validity of remote administration of the MATRICS Consensus Cognitive Battery for individuals with severe mental illness. Schizophr Res Cogn. 2022;27:100226.
28. Mahon S, Webb J, Snell D, Theadom A. Feasibility of administering the WAIS-IV using a home-based telehealth videoconferencing model. Clin Neuropsychol. 2022;36(3):558-70.
29. Brown T, Zakzanis K. A review of the reliability of remote neuropsychological assessment. Applied Neuropsychology: Adult. 2023:1-7.
30. Lavigne KM, Sauvé G, Raucher-Chéné D, Guimond S, Lecomte T, Bowie CR, et al. Remote cognitive assessment in severe mental illness: a scoping review. Schizophrenia. 2022;8(1):14.
31. Hays R, Henson P, Wisniewski H, Hendel V, Vaidyam A, Torous J. Assessing Cognition Outside of the Clinic: Smartphones and Sensors for Cognitive Assessment Across Diverse Psychiatric Disorders. Psychiatr Clin North Am. 2019;42(4):611-25.
32. Kilciksiz CM, Keefe R, Benoit J, Ongur D, Torous J. Verbal memory measurement towards digital perspectives in first-episode psychosis: A review. Schizophr Res Cogn. 2020;21:100177.

medRxiv preprint doi: https://doi.org/10.1101/2024.07.16.24310489; this version posted July 16, 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. It is made available under a CC-BY-NC-ND 4.0 International license .

33. Atkins AS, Kraus MS, Welch M, Yuan Z, Stevens H, Welsh-Bohmer KA, et al. Remote self-administration of digital cognitive tests using the Brief Assessment of Cognition: Feasibility, reliability, and sensitivity to subjective cognitive decline. Front Psychiatry. 2022;13:910896.
34. Hamner T, Salorio CF, Kalb L, Jacobson LA. Equivalency of In-Person Versus Remote Assessment: WISC-V and KTEA-3 Performance in Clinically Referred Children and Adolescents. J Int Neuropsychol Soc. 2022;28(8):835-44.
35. Zheutlin AB, Dennis J, Karlsson Linnér R, Moscati A, Restrepo N, Straub P, et al. Penetrance and Pleiotropy of Polygenic Risk Scores for Schizophrenia in 106,160 Patients Across Four Health Care Systems. Am J Psychiatry. 2019;176(10):846-55.
36. Sealock JM, Lee YH, Moscati A, Venkatesh S, Voloudakis G, Straub P, et al. Use of the PsycheMERGE Network to Investigate the Association Between Depression Polygenic Scores and White Blood Cell Count. JAMA Psychiatry. 2021;78(12):1365-74.
37. McCarty CA, Chisholm RL, Chute CG, Kullo IJ, Jarvik GP, Larson EB, et al. The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. BMC Med Genomics. 2011;4:13.
38. Mbatchou J, Barnard L, Backman J, Marcketta A, Kosmicki JA, Ziyatdinov A, et al. Computationally efficient whole-genome regression for quantitative and binary traits. Nature Genetics. 2021;53(7):1097-103.
39. Blout Zawatsky CL, Shah N, Machini K, Perez E, Christensen KD, Zouk H, et al. Returning actionable genomic results in a research biobank: Analytic validity, clinical implementation, and resource utilization. Am J Hum Genet. 2021;108(12):2224-37.
40. Alver M, Palover M, Saar A, Läll K, Zekavat SM, Tõnisson N, et al. Recall by genotype and cascade screening for familial hypercholesterolemia in a population-based biobank from Estonia. Genet Med. 2019;21(5):1173-80.
41. Nurm M, Reigo A, Nõukas M, Leitsalu L, Nikopensius T, Palover M, et al. Do Biobank Recall Studies Matter? Long-Term Follow-Up of Research Participants With Familial Hypercholesterolemia. Front Genet. 2022;13:936131.
42. Mahjani B, Birnbaum R, Buxbaum Grice A, Cappi C, Jung S, Avila MN, et al. Phenotypic Impact of Rare Potentially Damaging Copy Number Variation in Obsessive-Compulsive Disorder and Chronic Tic Disorders. Genes (Basel). 2022;13(10).
43. Kendall KM, Rees E, Bracher-Smith M, Legge S, Riglin L, Zammit S, et al. Association of Rare Copy Number Variants With Risk of Depression. JAMA Psychiatry. 2019;76(8):818-25.
44. McCutcheon RA, Keefe RSE, McGuire PK. Cognitive impairment in schizophrenia: aetiology, pathophysiology, and treatment. Mol Psychiatry. 2023;28(5):1902-18.

medRxiv preprint doi: https://doi.org/10.1101/2024.07.16.24310489; this version posted July 16, 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. It is made available under a CC-BY-NC-ND 4.0 International license .

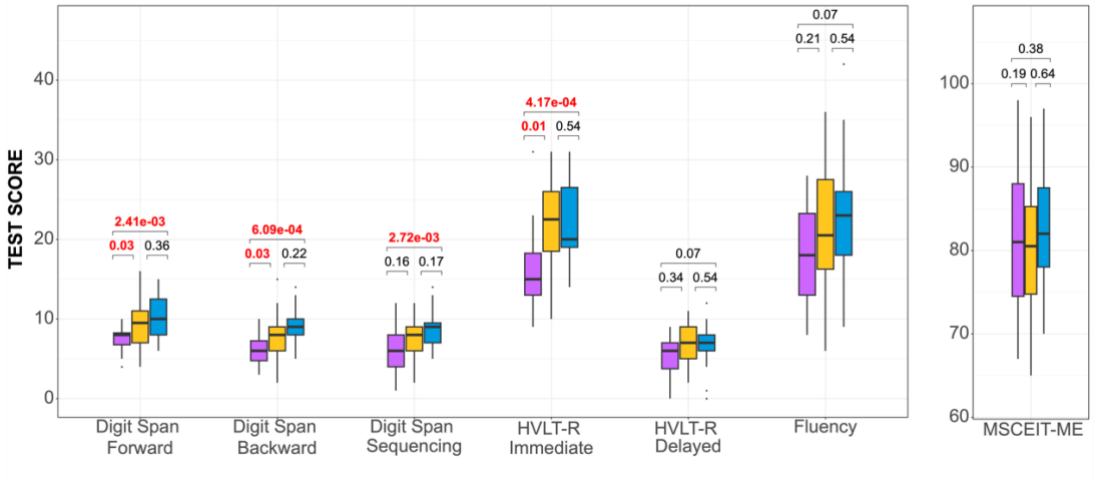
Figure 1: Study Recruitment. Flow chart of recontact of 892 biobank participants across three comparator groups, from initial stage of recontact to recruitment and study assessments. The indented table indicates the CNV loci of the 30 NDD CNV carriers who completed study assessments.



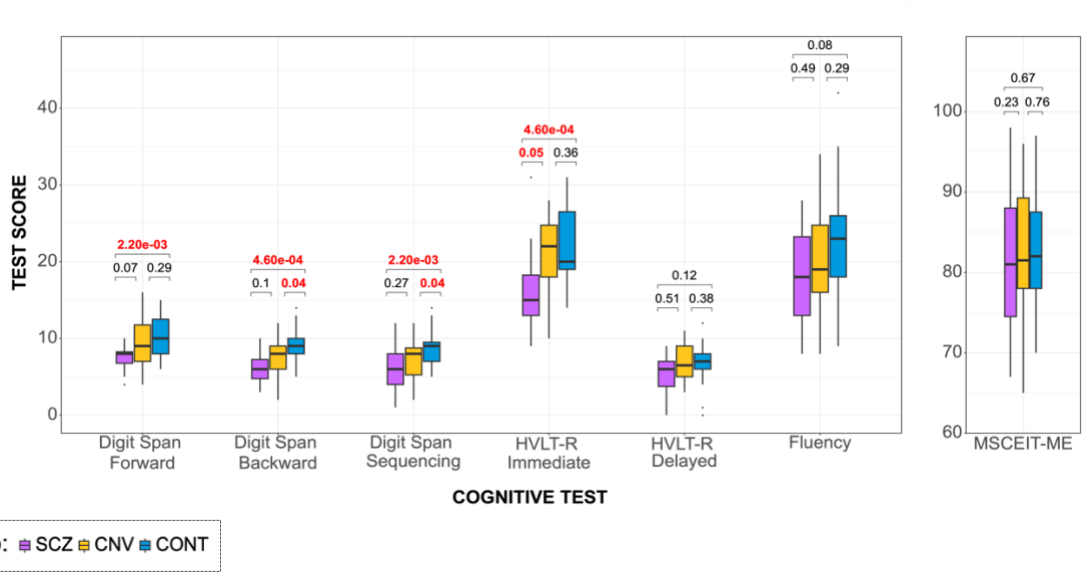
medRxiv preprint doi: https://doi.org/10.1101/2024.07.16.24310489; this version posted July 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 2: Cognitive Performance, by Group. Boxplots of cognitive performance for each of seven cognitive tests for NDD CNV carriers, compared to schizophrenia and controls, for (a) NDD CNV (n=30 individuals) and (b) NDD CNV (n=22 individuals)

(a)



(b)



Notes: x-axis is cognitive test and y-axis is raw cognitive test score. Each cognitive test score was regressed against group status, covarying for age, sex and ancestry. FDR values are indicated for each test. Red indicates statistical significance at FDR>0.05. See Supplementary Table 5 for complete results of cognitive analyses.

medRxiv preprint doi: https://doi.org/10.1101/2024.07.16.24310489; this version posted July 16, 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. It is made available under a CC-BY-NC-ND 4.0 International license .

Table 1: Study Assessment, Overview. Clinical and cognitive assessments administered to each study participant.

CLINICAL ASSESSMENT

Mini International Neuropsychiatric Interview (MINI)
Medical History
Psychiatric History (including medications and hospitalizations)
Developmental History (Developmental Milestones, Learning or Developmental Delay)
Social, Occupational and Educational History

COGNITIVE ASSESSMENT

Test	Domain	Battery
Digit Span Forward	Attention	WAIS-IV
Digit Span Backward	Executive function, Verbal working memory	WAIS-IV
Digit Span Sequencing	Working memory maintenance	WAIS-IV
Category fluency	Processing speed	MCCB
Hopkins Verbal Learning Test - Revised (HVLT-R)	Delayed recall	MCCB
Hopkins Verbal Learning Test - Revised (HVLT-R)	Immediate recall	MCCB
MSCEIT-ME	Social Cognition, Emotional intelligence	MCCB

Notes: MCCB=Matrics Concensus Cognitive Battery, WAIS=Wechsler Adult Intelligence Scale

medRxiv preprint doi: https://doi.org/10.1101/2024.07.16.24310489; this version posted July 16, 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. It is made available under a CC-BY-NC-ND 4.0 International license .

Table 2: Study Sample Demographics. The age, sex, and ancestry of each of three comparator groups who were recontacted and completed study assessments.

			RECONTACTED FROM BIOBANK			COMPLETED STUDY ASSESSMENTS				
	Biobank Participants (n=24,877)	NDD CNV (n=627)	NDD CNV (n=335)	Schizophrenia (n=217)	Controls (n=340)	NDD CNV (n=30)	Schizophrenia (n=20)	Controls (n=23)	Test statistic	P-value
Mean Age, Biobank Enrollment (SD)	50.5 (17.3)	50.4 (17.0)	39.0 (10.8)	45.4 (9.9)	41.9 (10.7)	37.3 [9.2]	41 [9.3]	39.9 [9.4]		
Mean Age at Study Assessment (SD)						47.3 [9.5]	50.2 [10.1]	49.5 [11.2]	0.55	0.58
Female (%)	14,586 (59%)	389 (62%)	219 (65%)	97 (45%)	209 (61%)	26 (87%)	8 (40%)	14 (61%)	11.96	0.003
Ancestry (%)						-	-	-	4.98	0.55
African	7892 (32%)	168 (27%)	95 (28%)	101 (47%)	111 (33%)	10 (33%)	9 (45%)	8 (35%)	0.77	0.68
European	5965 (24%)	209 (33%)	102 (30%)	17 (8%)	52 (15%)	10 (33%)	2 (10%)	7 (30%)	3.73	0.15
Hispanic	8536 (34%)	200 (32%)	109 (33%)	87 (40%)	151 (44%)	9 (30%)	9 (45%)	7 (30%)	1.42	0.49
Other	2484 (10)	51 (8%)	29 (9%)	12 (6%)	26 (8%)	1 (3%)	0 (0)	1 (4%)	-	-

Notes: Test statistics and p-values are for group differences across the three groups that completed study assessments.
Results of ANOVA for mean age (F-Statistic) and Chi-square tests for sex and ancestry are indicated.
Ancestry is self-reported ancestry.

medRxiv preprint doi: https://doi.org/10.1101/2024.07.16.24310489; this version posted July 16, 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. It is made available under a CC-BY-NC-ND 4.0 International license .

Table 3: Prevalence of Clinical Features, Overview. The prevalence of neuropsychiatric and neurodevelopmental clinical features, as assessed across each of three comparator groups.

	NDD CNV (n=30)			SCZ (n=20)		CONT (n=23)	
CLINICAL FEATURE	N	NDD CNV LOCI	EHR Unreported/ Discordant	N	EHR Unreported/ Discordant	N	EHR Unreported/ Discordant
MOOD OR ANXIETY DISORDER	12	TAR dup, 1q21.1 del, 1q21.1 dup, 2q13 (NPHP1) dup, 2q13 (NPHP1) dup, 15q11.2 dup, 15q11.2 dup, 15q13.3 dup (CHRNA7), 16p11.2 distal del, 16p13.11 del, 16p13.11 del, 16p13.11 dup	6	0	0	0	0
SEIZURE DISORDER	1	16p11.2 distal del	1	1	1	0	0
DEVELOPMENTAL DELAY (SPEECH OR MOTOR)	4	1q21.1 del, 2q13del(NPHP1), 15q11.2 dup, 16p11.2 del	1	3	1	0	0
CONGENITAL MALFORMATIONS	3	22q11.2 deletion, 2q13 (NPHP1) deletion, 15q11.2 deletion	1	0	0	0	0
LEARNING DISORDER	8	TAR dup; 1q21.1 del; 2q13 (NPHP1) dup, 2q13 (NPHP1) dup, 15q11.2 dup, 15q11.2 dup, 16p11.2 distal del; 22q11.2 del	8	0	0	0	0
HISTORY OF SPECIAL EDUCATION	4	2q13(NPHP1) dup, 15q11.2dup, 1q21.1del, 22q11.2 del	3	10	8	1	1
EXTREME LEARING DIFFICULTIES (WITHOUT LEARING DISORDER DIAGNOSIS)	4	TAR dup; 15q11.2 del; 16p13.11 del; 22q11.2 dup	4	2	2	0	0

Notes: For each group, the count of participants affected for each clinical feature is indicated. For NDD CNV carriers, NDD CNV loci of affected individuals are Indicated; History of Special Education for one control subject was limited and subject-specific for failed class, without an accompanying diagnosed learning disorder.
del=Deletion; dup=Duplication.