INTRODUCTION

Genomic copy number variations (CNVs) are structural variations that involve deletions and/or duplications of segments of DNA. Their impact is not necessarily harmful, but loss, increase, or disruption of genes is often associated with, and can underlie neurodevelopmental disorders (NDDs).1–6 Rare CNVs have been extensively studied in autism spectrum disorder (ASD),7–9 attention deficit hyperactivity disorder (ADHD),10 schizophrenia (SCZ),11,12 and obsessive-compulsive disorder (OCD).13,14 Various NDDs share genomic structural human disease, including neurodevelopmental disorders (NDDs).1–6 Rare CNVs have been extensively studied in autism spectrum disorder (ASD),7–9 attention deficit hyperactivity disorder (ADHD),10 schizophrenia (SCZ),11,12 and less so, obsessive compulsive disorder (OCD).13,14 Various NDDs share genomic structural
variations, including CNVs that perturb the same genes.\textsuperscript{1,15} For example, deletions or duplications affecting coding sequences of NRXN1 or CNTN6 have been implicated in ASD, ADHD, and SCZ.\textsuperscript{16,17}

A shared genomic etiology across multiple disorders is supported through genome-wide association studies\textsuperscript{18} or analyzing small loss-of-function (LoF) variations.\textsuperscript{19} However, CNV analysis across disorders has been limited; e.g., ADHD and ASD,\textsuperscript{10,19} ASD and SCZ,\textsuperscript{20} reviews on selected well-established genomic disorders such as 16p11.2 deletions and duplications,\textsuperscript{21} or a meta-analysis of CNVs on a single gene (NRXN1)\textsuperscript{15}. To date, there has been no genome-wide study of rare CNVs identified using an identical technology, encompassing ASD, ADHD, OCD, and SCZ. An identical method for interrogating CNVs across multiple disorders increases the chance of finding rare CNVs with cross-disorder implications. These could be missed if multiple technologies with different detection sensitivities were applied.

This project was established to provide a public resource of CNV samples with NDDs mostly from the province of Ontario, Canada, all genotyped on same microarray platform: the Affymetrix CytoScan HD platform, which consists of 1.9 million copy number markers and 750,000 genotype-able single nucleotides polymorphisms. A CNV resource of control population samples was published earlier under the Ontario Population Genomics Platform.\textsuperscript{22}

In creating this data resource, we aimed to: (i) catalogue CNVs that are clinically relevant to each of ASD, ADHD, OCD, and SCZ, and (ii) identify genes and loci with CNVs that are shared among different NDDs. Where available, we analyzed whole genome (WGS) or whole exome sequence (WES) data, in search of variants that were not detected by microarray. The relevant genotypes and CNVs are available in dbGAP (accession number phs001881.v1.p1) and dbVar (accession number nstd173), respectively.

RESULTS
Sample description and detection of CNVs
We analyzed 4,460 samples ascertained for four NDDs; 2,691 (60.3%) were from individuals recruited because of the diagnosis of one of these disorders, and the rest from apparently unaffected family members (Table 1, Supplementary Table 1A). ASD, ADHD, OCD, and SCZ cases were ascertained using criteria explained previously (Supplementary Information).\textsuperscript{8,10,13,23,24} Childhood onset OCD is not classically considered an NDD in the International Classification of Disease (ICD-11) or Diagnostic and Statistical Manual of Mental Disorders (DSM-5), but in view of its early onset, male preponderance and association with imaging findings, we included OCD in this study as others often do. Similarly, since SCZ has both neural and genetic correlates, including some evidence of overlap or sharing genetic risk with NDDs, we considered SCZ as well. The majority of samples (68.1%) were ascertained for ASD, with the others distributed approximately equally among the other disorders. The male:female sex ratio was almost 4:1 for ASD and ADHD, 2:1 for SCZ, and ∼1:1 in the pediatric OCD cases (Table 1). Different family structures were sampled in the four sub-groups: some ASD and SCZ families included multiple affected individuals; some ODM samples were in trios (i.e., affected proband and both parents; details in Table 1, Supplementary Table 1A). We did not genotype parents of cases for ADHD (Table 1, Supplementary Table 1A). We defined a high confidence set of CNVs (Supplementary Table 1B) as those identified using two different detection algorithms, as previously described.\textsuperscript{25} Rare variants were those with a frequency of less than 0.1% in a 10,851 subject population control sample genotyped in multiple microarray platforms, including the Affymetrix CytoScan HD.\textsuperscript{22} We also analyzed prioritized CNVs from our previous publications on ADHD\textsuperscript{10} and SCZ.\textsuperscript{11,24} We had information with respect to intellectual disability (ID) for some cases. ID including borderline intellect and non-verbal learning disability was comorbid with SCZ in 31/204 cases (16.9%)\textsuperscript{11} with ASD in 149/599 cases (24.9%), and with ADHD in 3/427 cases (0.7%). No OCD case had ID.

Clinically relevant CNVs
Clinically relevant CNVs included five categories: aneuploidies, large CNVs (\textgtr\textless 3 Mb), CNVs consistent with known recent genomic disorders, those impacting genes previously established to be associated with NDDs, and all de novo CNVs (i.e., not found in either parent) (details in Table 2). We found 306 clinically relevant CNVs in 284 of 2,691 NDD cases (10.5%) (Tables 2, 3, Supplementary Table 1C). Of these CNVs, 115 found in 111/2,691 cases (4.3%) were "clinically significant" or "likely clinically significant" variants, as evaluated by expert clinical geneticists according to American College of Medical Genetics and Genomics guidelines.\textsuperscript{26} We did not find evidence for uniparental disomy. The clinically relevant autosomal deletions, and chromosome X deletions in females, were all single copy.

The first category included aneuploidies: trisomy 21, 47,XYY, 47,XXY, and 45,X, found in 172/2,691 cases (0.63%), only among ADHD or ASD cases (category A in Tables 2, 3; Fig. 1). The second CNV category contained variants larger than 3 Mb, but excluding those

| Table 1. Stratification of 4,460 samples in the cross-disorder CNV analysis |
| Disorder | Samples\textsuperscript{1} | Cases\textsuperscript{2} | Adult \textsuperscript{3} | Sex ratio (male/female) | Only probands\textsuperscript{4} | Only one parent\textsuperscript{5} | Trios\textsuperscript{6} | Quartets\textsuperscript{7} | Others\textsuperscript{8} |
|----------|----------------|--------------|----------------|------------------------|----------------|----------------|----------------|----------------|----------------|
| ASD      | 3,034          | 1,838        | 4.0%           | 4.0 (1,474/366)        | 1,138          | 89             | 305            | 126            | 17             |
| ADHD     | 427            | 427          | 0              | 3.7 (337/90)           | 427            | 0              | 0              | 0              | 0              |
| SCZ      | 435\textsuperscript{9} | 204\textsuperscript{10} | 100%           | 2.0 (137/67)           | 155            | 36             | 13             | 0              | 0              |
| OCD      | 564            | 222          | 0              | 0.9 (107/115)          | 49             | 6              | 167            | 0              | 0              |
| All      | 4,460          | 2,691        | 10.3%          | 3.2 (2,055/638)        | 1,769          | 131            | 485            | 126            | 17             |

\textsuperscript{1}Count of all samples that passed our stringent QC, including assessing genotype quality, removing duplicated samples or those with non-mendelian segregation implying alternative parental relationships. This includes parents, extended family members, and siblings. \textsuperscript{2}Includes probands and affected sibling (s). Affected parents are not considered. \textsuperscript{3}Proportion of cases at least 18 years of age at the time of diagnosis (for schizophrenia) or time of DNA submission for the microarray genotyping (for ASD \( n=73 \)). \textsuperscript{4}Includes affected individuals for whom parents were not sampled or failed QC. No parents were sampled for ADHD. \textsuperscript{5}Affected individuals with one parent sampled or passed QC. \textsuperscript{6}Proband and both parents. \textsuperscript{7}Proband, one affected sibling and both parents. \textsuperscript{8}Families with three or more affected individuals (15 families with three, one family with four, and one with five affected children) for ASD. \textsuperscript{9}Includes 231 samples of extended family members or parents, including affected individuals. \textsuperscript{10}Of SCZ cases, 31 had intellectual disability.
associated with recurrent genomic disorders and aneuploidies (category B in Tables 2, 3). We found these large variants in 23/2,691 NDD cases (0.85%), with none among the OCD cases. The third category was the variants associated with known recurrent genomic disorders, found in 115/2,691 cases (4.3%); category C in Tables 2, 3; Fig. 1). The most frequent were 16p13.11 duplications (17 cases), 15q11.2 deletions (breakpoint (BP1-BP2) (16 cases), and 15q11.2 duplications (BP1-BP2) (15 cases). Three distal duplications of 16p11.2 were found in ADHD, ASD, and SCZ cases. We found 15q11-q13 duplications in six cases diagnosed with ADHD, ASD, ODC, or SCZ. The high prevalence of 15q11.2 duplications (BP1-BP2) and 16p11.13 duplications likely reflects the relatively mild expression and reduced penetrance of these genotypes.

Fourth, we identified de novo CNVs impacting genes in 35 cases: 31/448 (6.9%) for ASD and 3/167 (1.8%) for ODC (category D in Tables 2, 3). Of 13 SCZ trios with data, we identified one de novo genic CNV. Parents of ADHD cases were not analyzed, thus no data on de novo CNVs were available for this disorder.

The last class of variants included inherited clinically relevant CNVs that did not belong to any of the first three categories. These included CNVs impacting genes previously implicated for NDDs but either inherited or with unknown inheritance. We found these variants in 117/2,691 cases (4.4%) (category E in Tables 2, 3).

### Cases with more than one clinically relevant CNV

We found more than one clinically relevant CNV in 20/2,691 (0.74%) of NDD cases: 12/1,838 (0.65%) of ASD cases, 4/427 (0.95%) of ADHD cases; 3/222 (1.35%) of those with ODC, and 1/204 (0.5%) of individuals with SCZ.

In ASD, there were eleven cases with two CNVs and one case with three relevant variants. The latter, case 4-0040-003, had a medulloblastoma mutation CTNN2, a medulloblastoma 54 kb duplication impacting CTNN2, and a de novo 547 kb duplication affecting YES1, ADCYAP1, and six other genes. Examples with two clinically relevant CNVs each included: (i) a male (7-0293-003) with a 611 kb deletion of 16p11.2 and a 513 kb deletion of 15q11.2, of unknown inheritance; (ii) a female (2-1525-003) with a maternally inherited 1.5 Mb deletion consistent with the 16p13.11 recurrent microdeletion (neurocognitive disorder susceptibility locus), and a de novo 326 kb duplication impacting USP7, PMM2, C16orf72, and CARHSP1; and (iii) a male (2-0305-004) with a maternally inherited 432 kb deletion in the 1q21.1 locus associated with thrombocytopenia-absent radius syndrome and a de novo 1.09 Mb duplication impacting GPD2 and NR4A2. The four cases with ADHD with more than one CNV were all male; each had two clinically relevant CNVs, all of unknown inheritance. Examples included: (i) 213050 with Klinefelter syndrome (47,XXY) and an autosomal 181 kb duplication affecting MCPH1; (ii) 206760 with a 191 kb deletion impacting DLGAP2 and a 154 kb duplication involving DRD4 and eight more genes; and (iii) 235983S with Klinefelter syndrome (47,XXY) and an autosomal 123 kb duplication impacting DPP6.

Of the three OCD cases carrying more than one CNV, there was one male (OCD146-JS-1254_188613) with three clinically significant CNVs: a maternally inherited 143 kb deletion impacting NLGN1, a second maternally inherited 115 kb deletion impacting DPP9, and a paternally inherited 67 kb duplication impacting PTPRN2. Another male (OCD125-896993) had a de novo 165 kb deletion affecting ADRA2C and a paternally inherited 1.7 Mb duplication impacting CNTN4 and CNTN6. A female (OCD109-1648) had a maternally inherited 1.4 Mb deletion of 17p12 and a 190 kb deletion impacting PTPRT. This case also had a de novo frameshift deletion in LRCH2 (c.2190_2193delinsC).7

The only individual with SCZ having two clinically relevant CNVs (Supplementary Table 1C) was a female (222720) with a 10.2 Mb deletion of chromosome 5p congruent with the cri-du-chat syndrome region, and a 7.3 Mb duplication of 6q26-27. Karyotyping confirmed this to be the result of an unbalanced translocation.11

Cases with clinically relevant CNVs identified by microarray also had SNVs or CNVs detected by WGS or WES.

No WGS data were available for the ADHD and SCZ samples. However, we have previously published WGS data for 106/209 (50.7%) ASD cases with clinically relevant CNVs,9,27,28 Of these, 15 had a clinically relevant LoF mutation and one had a 4.5 kb deletion (Supplementary Table 1C). The latter case was a male (2-1086-004) with a maternally inherited 80.1 kb duplication impacting CTNND2, identified using array data. He also had paternally inherited 4.5 kb deletion in ANO3, known to be associated with autosomal dominant dystonia (OMIM: 6101110). This deletion was missed by microarray for lack of probes in this region, due to its size (Supplementary Table 1C). Examples of LoF SNVs were: (i) a female autism case (7-0133-003) with a 2.5 Mb de novo duplication in 10q11.22-11.23 and a de novo nonsense mutation in SOX5 (c.C313T:p.R105X). Mutations of SOX5 cause autosomal dominant Lamb-Shaffer syndrome, characterized by global developmental delay and intellectual disability (OMIM:604975), and (ii) a male autism case (7-0123-003) with a 2.9 Mb duplication involving

| Category | ASD (%) | ADHD (%) | OCD (%) | SCZ (%) | All cases (%) |
|----------|---------|----------|---------|---------|---------------|
| A: aneuploidies | 11 (0.6) | 6 (1.4) | 0 | 0 | 17 (0.63) |
| B: large CNVs (>3 Mb) | 16 (0.9) | 4 (0.9) | 0 | 3 (1.5) | 23 (0.85) |
| C: genomic disorder loci | 80 (4.3) | 18 (4.2) | 4 (1.8) | 13 (6.4) | 115 (4.3) |
| D: de novo | 31 (6.9) | NA | 3 (1.8) | 1 (NA) | 35 (5.6) |
| E: others | 89 (4.9) | 14 (3.3) | 8 (3.6) | 6 (2.9) | 117 (4.4) |
| Sum of unique samples across all categories | 209 (11.4) | 40 (9.4) | 13 (5.6) | 22 (10.8) | 284 (10.5) |

1This category included variants larger than 3 Mb but not aneuploidies or variants of known recurrent genomic disorder loci larger than 3 Mb, i.e., 15q11-13 duplication (Table 3, Supplementary Table 1C). 2Prevalence of known recurrent genomic syndromes in the general population is 0.8–1.0%. 3This category includes de novo variants from aneuploidies, large CNVs and genomic disorder loci. The rate of de novo CNVs in the general population is 0.9–1.4%23. This category included all other NDD relevant CNVs. Indicates that we did not sample parents of ADHD cases to establish inheritance pattern of variants. We sampled only 13 trios; most SCZ samples were unrelated. Sum of the counts in each column is higher in some cases than the total number of cases with relevant CNVs in the corresponding column, due to the fact that some subjects might carry multiple relevant variants, and (ii) de novo case included CNVs from aneuploidies, large CNVs, and the genomic disorder loci. 5Number of prioritized CNVs for ASD, ADHD, OCD, and SCZ were 223, 43, 17, and 23, respectively. 6Total number of prioritized CNVs across all cases was 306. The coordinates for clinically relevant CNVs are in Supplementary Table 1C.
### Table 3. Copy number variations of clinical significance among four neurodevelopmental disorders

| Category | Type | Size (kb) | Sex | Disorder Class | Sex | Origin |
|----------|------|-----------|-----|----------------|-----|--------|
| A: aneuploidies (n = 17) | 45,X | 155,271 | F | ASD, ADHD, SCZ | CS | D, U |
| | 47,XX,+ 21/47,XY,+ 21 | 155,271 | M | ASD(2), ADHD(3), SCZ | CS | D, U(5) |
| | 47,XXY | 155,271 | M | ASD(2), ADHD(2), SCZ | CS | D(2), U(2) |
| | 47,XYY | 59,374 | M | ASD(2), ADHD(2), SCZ | CS | D(2), U(2) |

B: large CNVs (> 3 Mb; n = 23) excluding aneuploidies (category A) and CNVs associated with known recurrent genomic disorders (category C)

| Type | Size (kb) | Sex | Disorder Class | Sex | Origin |
|------|-----------|-----|----------------|-----|--------|
| 46,XX,dup(2)(q23.3q24.1) | 3,071 | F | ASD | VUS | U |
| 46,XY,del(3)(p14.1p13) | 5,301 | M | ASD | LCS | D |
| 46,XY,dup(4)(q22.1q22.2) | 3,060 | M | ASD | LCS | M |
| 46,XY,del(4)(q32.3) | 3,209 | M | ASD | VUS | U |
| 46,XX,dup(4)(q34.3q35.1) | 3,886 | F | SCZ, SCZ | VUS | U |
| 46,XY,del(6)(q15q16.1) | 4,739 | M | ASD | LCS | M |
| 46,XX,dup,(6)(p26q27) | 7,302 | F | SCZ | LCS | U |
| 46,XY,del(7)(q31.1q31.31) | 11,051 | M | ASD, SCZ | A | CS | D |
| 46,XY,del(8)(p23.3p22) | 14,917 | M | ADHD | CS | U |
| 46,XY,del(9)(p22.1p21.2) | 8,005 | M | ASD | CS | U |
| 46,XY,del(13)(q14.13q14.3) | 5,350 | M | SCZ | LCS | U |
| 46,XY,dup(15)(q25.3q26.1) | 6,519 | M | ASD | LCS | U |
| 46,XY,del(18)(p11.32p11.21) | 15,264 | M | ASD | CS | U |
| 46,XY,del(18)(q22.1q23) | 13,469 | M | ASD | CS | U |
| 46,XY,del(18)(q24.21q24.22) | 4,945 | M | ADHD | CS | U |
| 46,XY,del(22)(q11.21q12.3) | 12,518 | F | ADHD | CS | U |
| 46,XX,dup(17)(q11.1q12) | 3,140 | M | ASD | CS | U |
| 1q21.1 Proximal Deletion (TAR Syndrome) | 1q21.1 DEL | 410–513 | M | ASD | VUS | M(2), U |
| 1q21.1 Distal Duplication | 1q21.1-q21.2 DUP | 1,817–2,034 | F(2), M(2) | ADHD, ASD(3), SZC(2) | CS | M, U(3) |
| 2q37.3 Deletion | 2q37.3 DEL | 5,275 | F | ADHD | CS | U |
| 3q29 Deletion | 3q29 DEL | 2,603 | M | ASD | CS | U |
| 4p16.3 Duplication | 4p16.3 DUP | 846–849 | M | ASD | VUS | P |
| 5p Deletion (Cri-du-chat) | 5p15.33-p15.2 DEL | 10,191 | F | SCZ, SCZ | VUS | D |
| 7q11.23 Deletion (WBS; includes ELN) | 1q21.23 DEL | 1,429 | M | ASD | CS | D |
| 9q34.3 Duplication | 9q34.3 DUP | 3,140 | M | ADHD | CS | U |
| 10q11.22-q11.23 Duplication | 10q11.22-q11.23 DUP | 2,650–2,661 | F | SCZ | VUS | D |
| 15q11.1-q13.1 Duplication | 15q11.1-q13.1 DUP | 4,918–6,158 | F(2), M(4) | ADHD, ASD(3), OCD, SCZ, SCZ | CS | U |
| 15q11.2 Deletion (BP1-BP2) | 15q11.2 DEL | 312–521 | F, M(15) | ADHD(3), ASD(12), OCD | VUS | M(4), U(12) |
| 15q11.2 Duplication (BP1-BP2) | 15q11.2 DUP | 311–850 | F(7), M(8) | ADHD(3), ASD(12) | VUS | M(2), P(2), U(11) |
| 15q11.3 Deletion | 15q11.3-q13.3 DEL | 1,533–2,072 | F(2), M(2) | ASD, SCZ | CS | P, U(3) |
| 15q25 Distal Deletion | 15q25.2-q25.3 DEL | 959 | F | ASD | VUS | U |
| 16p11.2 Distal Duplication | 16p11.2 DUP | 204 | M | ASD | VUS | U |
| 16p11.21 Deletion | 16p11.21 DEL | 1,496–3,352 | F, M(2) | ADHD, ASD(2) | CS, VUS(2) | D, M, U |
| 16p11.21 Duplication | 16p11.21 DUP | 783–2,950 | F(3), M(14) | ADHD, ASD(12), OCD, SCZ(3), SCZ | VUS | M(3), P(3), U(11) |
| 16p12.1 Deletion | 16p12.1 DEL | 670–680 | F, M(2) | ASD | VUS | M(2), U |
| 16p12.1 Duplication | 16p12.1 DUP | 613–655 | M | ADHD(2), ASD(2), SCZ | VUS | U |
| 16p11.2 Distal Duplication | 16p11.2 DUP | 273–362 | F, M(2) | ADHD, ASD, SCZ | CS | U |
| 16p11.2 Distal Deletion | 16p11.2 DEL | 227–243 | F, M | ASD | CS | D, U |
| 16p11.2 Proximal Duplication | 16p11.2 DUP | 625 | M | ADHD, ASD(2), SCZ | CS | U |
| 16p11.2 Proximal Deletion | 16p11.2 DEL | 598–746 | F, M(3) | ADHD, ASD(2), SCZ | CS | U |
| 17p12 Deletion | 17p12 DEL | 1,404 | F | OCD | CS | M |
| 17p12 Duplication | 17p12 DUP | 1,858–1,970 | M | ADHD | CS | U |
| 22q11.21 Deletion | 22q11.21 DEL | 1,396–3,154 | F(3), M(4) | ASD, SCZ(6), SCZ | CS | U |
| 22q11.21 Duplication | 22q11.21 DUP | 2,546–3,271 | M | ASD | CS | D, P, U(5) |
| Category       | subtype                                                                 | # | Cytoband   | Type     | Size (kb) | Sex | Disorder | Class       | Origin |
|----------------|--------------------------------------------------------------------------|---|------------|----------|-----------|-----|----------|-------------|--------|
| 22q11.2        | Distal Duplication (LCR22-F to LCR22-H)                                  | 1 | 22q11.22-q11.23 | DUP      | 2,062     | F   | ASD      | LCS         | U      |
| 22q13 Deletion |                                                                          | 1 | 22q13.33    | DEL      | 507       | F   | ASD      | CS          | U      |
| Xp22.3 Deletion|                                                                          | 1 | Xp22.3      | DEL      | 1,681     | M   | ADHD     | CS          | U      |
| D: de novo    |                                                                          |   |            |          |           |     |          |             |        |
| 1q21.3 Deletion| PSMD4 + 3 genes                                                          | 1 | 1q21.3     | DEL      | 98        | M   | ASD      | VUS         | D      |
| 2q23.1 Deletion| MBDS + ORC4                                                              | 1 | 2q23.1     | DUP      | 251       | M   | ASD      | CS          | D      |
| 2q24.1 Duplication| (NRRA2, GP2D2)                                                          | 1 | 2q24.1     | DEL      | 1,090     | M   | ASD      | VUS         | D      |
| 2q32.1-q32.2 Deletion| (GULP1)                                                               | 1 | 2q32.1-q32.2 | DEL      | 328       | F   | ASD      | VUS         | D      |
| 4p16.3 Deletion| ADRA2C                                                                  | 1 | 4p16.3     | DEL      | 165       | M   | OCD      | VUS         | D      |
| 7p22.1 Deletion| (AP5Z1, FOXK1)                                                           | 1 | 7p22.1     | DEL      | 44        | M   | ASD      | VUS         | D      |
| 7p22.1 Deletion| (FBX18 + TNRC18)                                                        | 1 | 7p22.1     | DEL      | 130       | M   | ASD      | VUS         | D      |
| 7q11.22 Deletion| AUTS2                                                                   | 1 | 7q11.22    | DEL      | 428       | M   | ASD      | LCS         | D      |
| 7q36.3 Deletion| EZH2                                                                    | 1 | 7q36.3     | DUP      | 341       | F   | ASD      | VUS         | D      |
| 1q21.3 Deletion| (PSMD4 + 3 genes)                                                       | 1 | 1q21.3     | DEL      | 98        | M   | ASD      | VUS         | D      |
| 2q23.1 Deletion| (NRRA2, GP2D2)                                                           | 1 | 2q23.1     | DUP      | 251       | M   | ASD      | CS          | D      |
| 2q24.1 Duplication| (NRRA2, GP2D2)                                                          | 1 | 2q24.1     | DEL      | 1,090     | M   | ASD      | VUS         | D      |
| 2q32.1-q32.2 Deletion| (GULP1)                                                               | 1 | 2q32.1-q32.2 | DEL      | 328       | F   | ASD      | VUS         | D      |
| 4p16.3 Deletion| ADRA2C                                                                  | 1 | 4p16.3     | DEL      | 165       | M   | OCD      | VUS         | D      |
| 7p22.1 Deletion| (AP5Z1, FOXK1)                                                           | 1 | 7p22.1     | DEL      | 44        | M   | ASD      | VUS         | D      |
| 7p22.1 Deletion| (FBX18 + TNRC18)                                                        | 1 | 7p22.1     | DEL      | 130       | M   | ASD      | VUS         | D      |
| 7q11.22 Deletion| AUTS2                                                                   | 1 | 7q11.22    | DEL      | 428       | M   | ASD      | LCS         | D      |
| 7q36.3 Deletion| EZH2                                                                    | 1 | 7q36.3     | DUP      | 341       | F   | ASD      | VUS         | D      |
| 1q21.3 Deletion| (PSMD4 + 3 genes)                                                       | 1 | 1q21.3     | DEL      | 98        | M   | ASD      | VUS         | D      |
| 2q23.1 Deletion| (NRRA2, GP2D2)                                                           | 1 | 2q23.1     | DUP      | 251       | M   | ASD      | CS          | D      |
| 2q24.1 Duplication| (NRRA2, GP2D2)                                                          | 1 | 2q24.1     | DEL      | 1,090     | M   | ASD      | VUS         | D      |
| 2q32.1-q32.2 Deletion| (GULP1)                                                               | 1 | 2q32.1-q32.2 | DEL      | 328       | F   | ASD      | VUS         | D      |
| 4p16.3 Deletion| ADRA2C                                                                  | 1 | 4p16.3     | DEL      | 165       | M   | OCD      | VUS         | D      |
| 7p22.1 Deletion| (AP5Z1, FOXK1)                                                           | 1 | 7p22.1     | DEL      | 44        | M   | ASD      | VUS         | D      |
| 7p22.1 Deletion| (FBX18 + TNRC18)                                                        | 1 | 7p22.1     | DEL      | 130       | M   | ASD      | VUS         | D      |
| 7q11.22 Deletion| AUTS2                                                                   | 1 | 7q11.22    | DEL      | 428       | M   | ASD      | LCS         | D      |
| 7q36.3 Deletion| EZH2                                                                    | 1 | 7q36.3     | DUP      | 341       | F   | ASD      | VUS         | D      |
| 1q21.3 Deletion| (PSMD4 + 3 genes)                                                       | 1 | 1q21.3     | DEL      | 98        | M   | ASD      | VUS         | D      |
| 2q23.1 Deletion| (NRRA2, GP2D2)                                                           | 1 | 2q23.1     | DUP      | 251       | M   | ASD      | CS          | D      |
| 2q24.1 Duplication| (NRRA2, GP2D2)                                                          | 1 | 2q24.1     | DEL      | 1,090     | M   | ASD      | VUS         | D      |
| 2q32.1-q32.2 Deletion| (GULP1)                                                               | 1 | 2q32.1-q32.2 | DEL      | 328       | F   | ASD      | VUS         | D      |
| 4p16.3 Deletion| ADRA2C                                                                  | 1 | 4p16.3     | DEL      | 165       | M   | OCD      | VUS         | D      |

E: CNVs not in categories A-D (n = 127)

1p36.33-p36.32 Deletion (SKI + 61 genes) 1

1q21.1 Duplication (HFE2 + 19 genes) 1

1q31.3 Duplication (GRN2D, KNC, PKNP + 107 genes) 1

21q2.2 Duplication (PDE4A + 4 genes) 1

2q11.2 Duplication (UPB1) 1

Xp22.3 Duplication (ANOS1, VCX3B) 1

Xp11.22 Duplication (SMC1A) 1

E: CNVs not in categories A-D (n = 127)
### Table 3 continued

| Category | n | Cytoband | Type | Size (kb) | Sex | Disorder | Class | Origin |
|----------|---|----------|------|-----------|-----|----------|-------|--------|
| KMT2C    | 1 | 7q36.1   | DUP  | 440       | F   | ASD      | LB    | U      |
| QPP6, PAXIP1, HTRSA | 3 | 7q36.2   | DUP  | 123–1,800 | M   | ADHD, ASD(2) | VUS   | M, U(2) |
| QPP6     | 3 | 7q36.2   | DEL  | 52–396    | M   | ADHD, ASD, OCD, SCZ<sup>a</sup> | VUS   | M, P, U |
| PTTPR2, ESYT2, NCAFPG2 | 3 | 7q36.3   | DUP  | 50–430    | M   | ADHD, ADHD<sup>a</sup>, ASD, OCD | LB    | M, P, U |
| DLGAP2, CLN8, ARHGEF10 | 6 | 8p23.3   | DEL  | 58–191    | M   | ADHD, ASD | VUS   | U      |
| PTPRD    | 2 | 8p23.3   | DEL  | 181–272   | M   | ADHD, OCD | LB    | M, U    |
| MCHF1    | 2 | 8p23.2–8p23.1 | DUP  | 220–312   | M   | ASD      | VUS   | P      |
| ASTN2    | 1 | 9p24.1   | DEL  | 58–291    | M   | ASD      | VUS   | U      |
| PTPRD    | 1 | 9q24.1   | DEL  | 81        | F   | ASD      | VUS   | U      |
| ASTN2, TRIMP32, PAPPA | 3 | 9q33.1   | DEL  | 25–523    | F, M(2) | ADHD<sup>a</sup>, ASD(2) | VUS   | U      |
| PCDH15   | 2 | 10q21.1  | DEL  | 53–306    | F(2), M(5) | ASD | B | M(3), P, U(3) |
| CTNNN3   | 7 | 10q21.3  | DEL  | 224–254   | F, M(2) | ASD | LCS | P(2), U |
| 1qo24.32 Duplication (POLL, BTRC, DPDCD) | 1 | 1qo24.32 | DUP  | 226       | M   | ASD      | LCS   | P      |
| DRD4 + 8 genes | 1 | 11p15.5  | DUP  | 154       | F   | ADHD     | VUS   | U      |
| PAX6, ELPA4 | 1 | 11p13    | DUP  | 593       | M   | ADHD     | VUS   | U      |
| SHANK2   | 1 | 11q13.4  | DEL  | 132       | M   | ASD      | LCS   | M      |
| DLG2     | 1 | 11q14.1  | DEL  | 210       | F   | ASD      | VUS   | U      |
| CTNN5    | 1 | 11q22.1  | DEL  | 25        | M   | ASD      | VUS   | U      |
| KMT2A + 6 genes | 1 | 11q23.3  | DUP  | 265       | F   | ASD      | VUS   | M      |
| CACNA1C  | 1 | 12p13.33 | DUP  | 56        | M   | ASD      | VUS   | M      |
| PCDH9    | 1 | 13q21.32 | DUP  | 1,006     | M   | ASD      | VUS   | M      |
| CHD8     | 1 | 14q11.2  | DUP  | 23        | M   | ASD      | VUS   | P      |
| NRXN3    | 3<sup>a</sup> | 14q31.1 | DEL  | 224–254   | F, M(2) | ASD | LCS | P(2), U |
| 15q13.1-q13.2 Duplication (APBA2 + 4 genes) | 1 | 15q13.1-q13.2 | DUP  | 1,376     | M   | ASD      | VUS   | M      |

<sup>ǂ</sup>Williams-Beuren Syndrome, <sup>ǂ</sup>F female, <sup>ǂ</sup>M male, <sup>ǂ</sup>DEL deletion, <sup>ǂ</sup>DUP duplication, <sup>ǂ</sup>ASD autism spectrum disorder, <sup>ǂ</sup>ADHD attention deficit hyperactivity disorder, <sup>ǂ</sup>SCZ schizophrenia, <sup>ǂ</sup>OCD obsessive compulsive disorder, <sup>ǂ</sup>VUS variants of unknown significance, <sup>ǂ</sup>B benign, <sup>ǂ</sup>LB likely benign, <sup>ǂ</sup>CS clinically significant, <sup>ǂ</sup>LCS likely clinically significant, <sup>ǂ</sup>M maternally inherited, <sup>ǂ</sup>P paternally inherited, <sup>ǂ</sup>U unknown inheritance, <sup>ǂ</sup>D de novo. <sup>ǂ</sup>Number of subjects carrying CNVs in the indicated cytoband. <sup>ǂ</sup>Clinical impact of variants was evaluated according to accepted clinical guidelines. <sup>ǂ</sup>Mosaic CNV. <sup>ǂ</sup>These two duplications are in two brothers. <sup>ǂ</sup>qPCR showed that the father was also carrying these duplications on chromosome X, which implies that there might be translocation between X and an unknown autosome, which then was transmitted to these two boys. We did not attempt identify the location of this duplication on autosomes. <sup>ǂ</sup>AIndicates the presence of a similar CNV detected in a subject published previously as part of two large-scale schizophrenia CNV studies (Costain et al. 24 and Lowther et al. 11) and one ADHD study (Lionel et al. 10). <sup>ǂ</sup>Sex, size of CNVs, inheritance, and the clinical classifications of these CNVs were not provided. <sup>ǂ</sup>Only genes whose coding sequences are impacted by CNVs are shown. However, CNVs impacting exons (UTRs) in MBDs and PTCHD-AS were indicated. <sup>ǂ</sup>Monogyzotic pair of twins from two different families; <sup>ǂ</sup>Siblings; <sup>ǂ</sup>two duplications from one case. <sup>ǂ</sup>Two of these deletions are paternally inherited in siblings diagnosed with ASD (one male, one female). <sup>ǂ</sup>The coordinates for clinically relevant CNVs are in Supplementary Table 1C.
the 16p13.11 recurrent microduplication who also had a de novo splice-site variant impacting SHANK3 (c.2223 + 1 G > A), which is a gene strongly associated with NDDs. Of 13 OCD cases with clinically relevant CNVs, three also had clinically relevant LoF mutations previously identified using WES (Supplementary Table 1A). One example is a male OCD case (OCD146-JS-1254_188613) with three clinically relevant CNVs: a maternally inherited 142 kb deletion in 3q26.31, a maternally inherited 115 kb deletion in 7q36.2, and a paternally inherited 67 kb duplication in 7q36.3. He also had three clinically relevant SNVs found by WES: (i) a maternally inherited frameshift deletion in AFF2 (c.2976_2988del:p.992_996del), which is an X-linked recessive variant associated with mental retardation (OMIM:300806), (ii) a maternally inherited frameshift deletion in DRD4 (c.233_245del:p.A78fs), an autosomal dominant variant associated with autonomic nervous system dysfunction and ADHD (OMIM:126452), and (iii) a maternally inherited frameshift in MBD4 (c.939_940ins:p.E314fs), a gene involved in DNA methylation (OMIM:603574).

Complex phenotypes
Because we had clinical information on NDD phenotypes beyond the primary diagnoses for some cases, we investigated the pleiotropy of CNVs shared among different NDDs (Supplementary Table 1A, DEL deletion, DUP duplication, TAR Thrombocytopenia-Absent Radius syndrome locus, STS includes STS, BP breakpoint, LCR low-copy repeat, Prox proximal, Dist distal).

Fig. 1 Distribution of a) aneuploidies and b) known recurrent genomic disorder CNVs found in cases diagnosed with autism spectrum disorder (ASD), attention deficit hyperactivity disorder (ADHD), schizophrenia (SCZ), or obsessive compulsive disorder (OCD). Details of the copy number variants, sex, and variant sizes are in Table 1, Supplementary Table 1A. DEL deletion, DUP duplication, TAR Thrombocytopenia-Absent Radius syndrome locus, STS includes STS, BP breakpoint, LCR low-copy repeat, Prox proximal, Dist distal.
Table 1). We defined complex as having multiple different NDDs. Of ADHD cases with an NDD relevant CNV, a few were noteworthy and highlight the clinical pleiotropy associated with many of these variants. A male ascertained for ADHD (176004), who also had a learning disability but no ASD, carried a duplication of 16p11.2, which is known to be associated with ASD.8 We found deletions at this locus in a female with SCZ, a male with ASD, and another male with ADHD, but none in our OCD cases. A male diagnosed with ADHD (206773), who carried a duplication of chromosome X (Klinefelter syndrome), also had ASD, learning disability, language delay, general anxiety disorder, and enuresis, all known features of Klinefelter syndrome.29 Male case 181220 with ADHD, with 15q11.2 duplication (BP1-BP2), also had ASD. Of OCD cases, a male (OCD75-SB-1213) with a paternally inherited 62 kb duplication of $\text{DLGAP2}$ also had separation anxiety disorder, a Tourette disorder with tic, oppositional defiant disorder, and panic disorders and agoraphobia. Another male with OCD (OCD146-JS-1254_188613), ADHD (inattentive subtype), and a Tourette disorder with tic, had three different CNVs, impacting $\text{NLGN1}$, $\text{DPP6}$, and $\text{PTPRN2}$. Of SCZ cases, a male (153030) with a 1.6 Mb duplication of 16p13.11 also had a learning disorder but no ID (details in Supplementary Table 1C). A female (213684) with SCZ and a 549 kb deletion of $\text{NRXN1}$ also had moderate intellectual disability. A male with SCZ (166808) with a 15q11-q13 duplication had mild intellectual disability.

Cross-disorder gene discovery and genes in multiple cases in a single disorder
We searched for genes, excluding those from regions of known recurrent genomic disorders, that were affected in multiple cases by CNVs. We first restricted analysis to brain-expressed genes that are at least moderately constrained for LoF variants, (pLI > 0.45; Fig. 2; Supplementary Table 1D). We searched these for genes impacted by CNVs in at least two cases each, and found 20 genes impacted by deletions. Notably, $\text{NRXN1}$ was impacted in 10 subjects (three SCZ, six ASD, and one ADHD); deletions of 18p11.21 impacting novel candidate genes ($\text{GNAL}$, $\text{LDLRAD4}$, and $\text{SEH1L}$) were in four cases (three ASD, one ADHD). Genes $\text{MKRN1}$ and
MYH9 involved CNVs in ASD and SCZ cases. Eight genes — NRXN1, GNL1, LDLR4AD4, SEH1L, DLPAP2, DCTN2, GRID2, and KIF5A — involved CNVs in ASD and ADHD cases. Only CNVs containing ABR were shared between OCD and ASD, and no gene-containing variants were shared between OCD and ADHD or OCD and SCZ.

Genes impacted by deletions in multiple ASD cases (only) were: ASTN2, NRXN3, ANKIS1B, GALNT13, DCL1, LAMP1, METRN1, PTRPK, and SPOCK1 (Fig. 2; Supplementary Table 1D). Although ASTN2 deletions were previously reported in ADHD,36 in this study we made no such observations. We saw no genes impacted by deletions in multiple cases of ADHD, OCD, or SCZ, other than those in the known genomic syndromes (Fig. 2; Supplementary Table 1E).

We identified 53 brain-expressed genes impacted in multiple cases by duplications of the entire longest transcript (Fig. 2; Supplementary Table 1E). Examples included DNAJC15, GNGT3, CARRHSP1, PCMTD2, RPS3A, and TMEM158 (Supplementary Table 1E). Most whole-gene duplications were from ASD cases, probably due to the latter’s disproportionate representation. Examples of genes with such variants in multiple disorders were: PCMTD2 in ASD, OCD and SCZ; KNDC1 in ASD and SCZ; CARRHSP1, PCMTD2, SYNM, and EXOC3 in both ASD and OCD; and GNG13, MRP5S3, RPS3A, FAM69C, and CD24 in ASD and ADHD.

We found 38 genes duplicated in multiple ASD cases (only). Examples included CDH15, UBF3, DUT, HYPK, ATXN7L3, and GLOD4. Duplication of NDUFV1 and RIMKLB were each observed in two ADHD cases (Fig. 2; Supplementary Table 1E). We found no repeated full gene duplications in OCD or SCZ cases in this collection.

Increased burden of rare CNVs impacting brain-expressed protein coding genes and brain-expressed long non-coding RNA (IncRNA) We sought rare CNVs impacting exons of IncRNAs and found these in 1,130/2,691 cases (42%). Restricting to brain-expressed IncRNAs, only 234/2,691 cases (8.7%) carried such rare CNVs. We tested for the extent to which the protein coding genes and IncRNAs were impacted by rare CNVs in cases compared with parents of cases. We found a nearly significant excess in cases over controls of deletions in protein-coding genes (p = 0.08; false discovery rate (FDR) = 0.21), but not for IncRNAs (p > 0.1). We found no global excess burden of duplications for protein-coding genes and IncRNAs (p > 0.1). However, when focused on brain-expressed elements, we observed a modest increase of rare deletions impacting both protein-coding genes (p = 0.03; FDR = 0.19) and IncRNAs (p = 0.06; FDR = 0.21). We then performed a multivariate analysis to test whether the burden signal was from protein-coding, IncRNAs, or both. This analysis showed a statistically significant signal (p = 0.02) for deletions impacting protein-coding genes, suggesting an overlap between the protein-coding and IncRNA burden signal.

Given the increasing association of IncRNAs in disease, we highlight two example of such genes identified in multiple unrelated individuals. (i) AK127244: three subjects with ASD (1-0045-004, 7-0103-003, and 1-0629-003) harbored 2p16.3 deletions that directly disrupted the exonic sequence AK127244 (L0CT730100). This is a 1.38 Mb non-coding RNA of unknown function adjacent to NRXN1 and transcribed in the opposite direction. Rare, inherited deletions intragenic to AK127244 have been identified in five individuals with ASD. Such deletions have been proposed as candidate factors for a broad range of neuropsychiatric disorders including SCZ and autistic disorder.6,31,52 We identified seven additional subjects here, five with ASD and two with SCZ, with coding deletions of NRXN1 that extend and disrupt the transcription start site and exonic sequence of AK127244. (ii) PTCHD1-AS: we found five males with ASD and deletions impacting exons of this gene (Table 3, Supplementary Table 1C). Disruption of PTCHD1-AS has been linked to ASD.33,34

Increased burden of rare CNVs impacting NDD genes in cases carrying CNVs that impact genomic instability genes and fragile sites We hypothesized that CNVs affecting genes involved in genome stability might lead to a higher incidence of additional variants. Such subsequent variants could then add to the phenotypic complexity, by impacting genes involved in the development and functions of the nervous system. We therefore tested if individuals carrying a CNV that impacts “genomic instability genes” (GIG-CNV) have a higher burden of rare CNVs (measured as the number and the cumulative length of rare CNVs per individual) than do individuals not carrying such CNVs.35 We compiled a set of 958 protein coding “genomic instability genes” from the AmiGO database.36 The “genomic instability genes” were not disproportionately impacted by rare CNVs in cases compared with controls (parents or unaffected individuals for this analysis) (p > 0.1). In individuals who had a GIG-CNV, we found an increase in mean number of CNVs (3.3 vs 2.1; p = 2.11 × 10−05) and cumulative length of rare CNVs (4.4 Mb vs 315 kb; p = 2.11 × 10−05) compared to individuals without a GIG-CNV. We observed a similar trend excluding CNVs impacting the “genomic instability genes” from the burden analysis (mean number of rare CNVs: 2.8 vs 2.1; p = 0.003; cumulative length of rare CNVs: 564 kb vs 315 kb; p = 0.024). This difference was even higher when considering only cases (i.e. not controls)(mean number of rare CNVs: 2.9 vs 2.1; p = 0.013; cumulative length of rare CNVs: 686 kb vs 358 kb; p = 0.051). We also found a 2.77-fold increase in the number of cases with rare CNVs impacting NDD-associated genes (NDD-CNV)(n = 1,160; Supplementary Information) and a GIG-CNV (Fischer’s exact test, p = 4.7 × 10−05, odds ratio: 2.77(C1:1.66–4.54)), compared with cases with NDD-CNVs but without a GIG-CNV. We then excluded individuals with aneuploidy or a CNV impacting both “genomic instability genes” and NDD genes. We still observed the excess number of CNVs with NDD-CNVs and GIG-CNVs over those with NDD-CNVs only (odds ratio of 2.52(C1:1.50–4.16) (Fisher’s exact test, p = 2 × 10−05)).

We then tested whether there was an increase CNV burden among cases whose parents had a GIG-CNV. Such CNVs could have been generated de novo anywhere previously in the pedigree and had not necessarily arisen de novo in the affected individual. We excluded the cases carrying a de novo GIG-CNV that was not found in the parents. We found a higher average number of rare CNVs in cases whose parents had GIG-CNVs compared to cases whose parents did not (2.66 vs 2.10; p = 0.02). We observed a similar trend for this global burden when excluding the GIG-CNVs (2.49 vs 2.10; p = 0.08). However, we did not observe an increased global burden in the cumulative length of rare CNVs.

We then further investigated cases with de novo CNVs for whom WGS data were available (n = 16, all from the ASD cohort). We found no over-representation (p > 0.1) of cases with de novo CNVs from families in which at least one parent carried a LoF variant impacting a “genomic instability gene” (n = 10) compared to other families (n = 6). Again, this was a small sample size. There were notable examples of cases with de novo CNVs whose parents had LoF variant(s) on “genomic instability genes” (i) Case 1-0627-007 had a paternally inherited frameshift deletion impacting PALB2 (c.509_510del:p.R170fs). He also had a 1.9 Mb de novo deletion in 16q23.3-q241 (Table 3, Supplementary Table 1C). Disruption of PTCHD1-AS has been linked to ASD.33,34

Published in partnership with CEGMR, King Abdulaziz University npj Genomic Medicine (2019) 26

M. Zarrei et al.
EXO1 functions in DNA replication, repair, and recombination (OMIM:600603). (iv) The father of ASD case 2-1693-003 had a RAD1 variant (c.168_172delp.A56fs) - a gene required for DNA replication and repair (OMIM: 603153). She carried a de novo 24 kb deletion at Xp11.22.

We also studied the extent of de novo CNVs overlapping genomic fragile sites. Of 33 de novo CNVs (excluding aneuploidies), 25 (75.6%) overlapped fragile site regions (Supplementary Table 1F). In addition, eleven of the de novo CNVs overlapped long genes (>300 kb), a feature associated with fragile sites and neuronal genes.36 Notably, five of these genes – MBDS, FAM19A1, FOXP2, AUTS2, and DLGAP2 – are involved in neuron formation and differentiation.

**DISCUSSION**

We generated a biosource to investigate the contribution of rare CNVs to the etiology of four NDDs – ASD, ADH, OCD, and SCZ – among 2,691 diagnosed cases. We found that 10.5% of these cases carried CNVs with potential clinical relevance to NDDs. Of all cases, 4.1% carried CNVs that were formally classified as clinically significant or likely clinically significant, when evaluated according to ACMG guidelines.26 We also found variant genes/regions that were shared across some or all of the NDDs. Evidence included recurrent or non-recurrent CNVs impacting the same genes in cases with different NDDs, and in patients diagnosed with multiple comorbid NDDs.

Of the four NDDs, ASD had the highest proportion of cases with a clinically relevant CNV (11.4%). ODD cases had the lowest proportion with identified CNVs (5.6%). Deletions of 22q11.21 were found in 6/204 SCZ cases (2.9%) - three with mild intellectual disability - contributing to the relatively high proportion of SCZ cases deemed to have a clinically relevant CNV (10.8%). 22q11.2 deletions are expected to be identified in about one in every 100-200 individuals with SCZ and about one in 10 with dual diagnosis of SCZ and ID.11 The enrichment of the SCZ cohort studied for ID likely contributed more to the prevalence observed of 22q11.2 deletions. Of ADHD cases, 9.4% carried clinically relevant CNVs, which is slightly higher than 8.9% previously reported using a different microarray (Affymetrix SNP 6.0).13 We also found multiple clinically relevant CNVs in 20/2,691 (0.74%) of NDD cases.

We found 17 aneuploidies (45,X; 47,XXY and 47,YYY) in cases diagnosed with either ASD or ADH (Table 2). The prevalence of Turner syndrome (45,X) (n = 2) was 1/1,300 among our cases, which is similar to previous reports.39 One had ADHD and the other ASD, similar to other reports.39,40 Cases with 47,XXY (n = 5) or 47,YYY (n = 4) had diagnoses of either ASD or ADHD, similar to previous reports.29 We found trisomy 21 (n = 6) only among ASD patients.40 Large CNVs other than aneuploidies were found in 23/2,691 (0.85%) cases, mainly in gene-rich regions of the genome. Although we did not find aneuploidies in SCZ cases, they have been reported in association with this phenotype previously.11,24

We found CNVs associated with known recurrent genomic disorders in 4.3% of cases (Table 2). This signified an increase of this type of CNVs among NDDs compared with that from a community population (1.1% (52/4,817); unpublished data), but similar to that of subjects with neurocognitive deficits in the UK Biobank (3.8%).41

Known recurrent genomic disorders were distributed differently among the four NDDs (Table 2; Fig. 1). We observed ASD in 80/115 (70%) of subjects with known recurrent genomic disorders, e.g., 7q11.23 deletion, 16p11.2 distal deletion, and 22q11.21 duplications. The 5p deletion was unique to SCZ. Deletions of 2q37.3, 16p11.2 proximal duplications, and Xp22.3 deletions were found only in ADHD, whereas 17q12 deletion was only found in ODD. Duplications of 15q11.1-q13, deletions of 15q11.2 (BP1-BP2), and 16p13.11 duplications were observed among cases of all four disorders (Fig. 1). Proximal duplications of 16p11.2 are found in up to 1% of individuals with SCZ.11,24,42

When parents were available to determine origin, we found 5.6% of this subset of cases to have a de novo CNV (Table 2). The highest de novo rate was for ASD (6.9%), consistent with previous reports from 4.7 to 7.1%.43,44 For OCD, 1.8% had de novo CNVs, which is higher than the rate found in the general population (0.9–1.4%), but lower than 2.3% for OCD previously reported from a larger sample size.13

We observed deletions and duplications, other than those associated with known recurrent genomic disorders, in the same genes in different NDDs, including some in multiple cases (Fig. 2; Table 2, Supplementary Table 1D). Deletions impacted NRXN1 (in 9 males, 1 female) among ASD, ADH, and SCZ cases (Fig. 2). Similarly, we found deletions impacting GNAL, LDLRAD4, SEH1L, DLGAP2, DCTN2, GRID2, and Kif5A among both ASD and ADH cases (Fig. 2).10

Disruptive variants in gene-sets involved in multiple intracellular signaling pathways and DNA instability have been observed previously in ASD.23 Variants in gene pathways associated with DNA”genomic instability” are increased in both ASD and SCZ.20,47 Consistent with these studies, we observed a 2.77-fold higher proportion of cases with NDD-CNVs among those with GIG-CNVs, than among those without.

This study had certain limitations. (i) Most cases, with the exception of SCZ, were recruited as children or adolescents on the basis of a specific diagnosis. ASD and ADH have early onset, and many participants would not have reached the age for adolescent or adult-onset disorders, including ODD and SCZ. It is possible that individuals with early onset conditions will develop additional later-onset comorbidities. All SCZ cases in the current study were adults. (ii) Recruitment was by clinicians who focus on a single disorder. It is possible that some cases may have had other NDDs, which were not reported. For example, we had data on intellectual disability/QI for the SCZ cohort and for some cases with other NDDs. We searched the genotype data for possible multiple ascertainment of any case and found no examples of subjects that were recruited through multiple disorders. We examined for non-primary phenotypes for specific cases with variants in NDD-relevant genes. (iii) Due to limitations of the technology, we studied CNVs only of a certain size (>20 kb) for the majority of samples where we did not have sequence data. Smaller CNVs and single nucleotide polymorphisms also contribute to the etiology of NDDs,8,35 but these would have been missed. A more sensitive technology such as genome sequencing would allow more comprehensive detection of all relevant variants.46,47 The dataset also needs to be analyzed iteratively as more data and better analysis tools become available.45

In summary, we highlighted clinically relevant CNVs found through microarray data for ASD, ADH, ODD, and SCZ. We also demonstrated that identical CNVs or genes could potentially contribute to the etiology of multiple NDDs, consistent with previous reports,10,20,46,47 and providing a valuable resource for comparison in other studies.

**METHODS**

**Samples**

This project was a part of a multilateral collaborative project to investigate genetic etiology across four neurodevelopmental disorders: ADHD, ADH, ODD, and SCZ. This study was approved by the Research Ethics Board at The Hospital for Sick Children. A written informed consent was obtained from all participating or substitute decision makers. CNVs were detected on the same high-resolution microarray platform. The criteria for meeting a diagnosis of ASD, ADH, ODD, or SCZ were detailed in our previous publications10,13,23,24 with a few modifications for ADH (see Supplementary Information). Data from all ODD individuals and 139/435 (32%) of the SCZ cohort had been...
Genotyping and detection of rare variants

We extracted genomic DNA from saliva or blood and genotyped samples on the Affymetrix CytoScan HD platform. Quality control and ancestry assessment procedures were as discussed previously. Using PLINK v1.90b2, we found 1,995 (74.1%) of cases to be of European ancestry (Supplementary Table 1A). CNVs were identified as previously described. Briefly, four different algorithms were used to call high-confidence CNVs. These included the Affymetrix Chromosome Analysis Suite, iPattern, BioDiscovery Nexus, and Partek Genomics Suite. We defined a stringent set of variants of at least 20 kb wherein each was identified by at least two algorithms and spanned by at least five consecutive probes (Supplementary Table 1B). We defined rare CNVs as those present at no more than 0.1% frequency among 10,851 controls samples (detailed in Zarrei et al.). We further restricted our list to those with more than 75% overlap with copy-number stable regions, according to our stringent CNV map of the human genome. We confirmed clinically relevant CNVs (Tables 1, 2, Supplementary Table 1C) (as defined below) using a SYBR Green-based real-time quantitative PCR assays, TaqMan copy number assays or whole genome sequencing data (if available). The genomic coordinates used are based on Human Genome Build GRCh37/hg19.

Prioritizing variants relevant to NDDs and the NDD gene list

To focus on CNVs relevant to NDDs, we first selected those variants coinciding with known recurrent genomic disorders, aneuploidies, and large (³3 Mb) deletions and duplications. We also analyzed whether rare CNVs in our cases were similar to those in clinically relevant CNV databases at The Department of Paediatric Laboratory Medicine, The Hospital for Sick Children, comprising over 20,000 cases. We classified variants for their clinical impact according to American College of Medical Genetics guidelines. Our prioritized variants also included those impacting the coding sequences of genes with sufficient evidence for being clinically relevant to NDDs (Supplementary Information).

Cross-disorder gene discovery and genes in more than one case in a single disorder

We searched for genes that were impacted by CNVs of 20 kb to 3 Mb in more than one case. Of these, we analyzed brain-expressed genes that were implicated by rare deletions and that are moderately to strongly expressed genes (Supplementary Information). We considered whether cases with perturbed genomic instability genes had a different burden of rare CNVs in NDD genes compared to that in cases with intact instability genes.

Global burden test for protein-coding genes and IncRNAs

We performed a univariate analysis to test the global burden of variants impacting coding sequences of protein-coding genes and all exons of IncRNAs using a logistic regression model. We further tested a burden for brain-expressed protein-coding genes (n = 3,666) and IncRNAs (n = 1,070) to compare with those not expressed in the brain. Chromatin states from the Roadmap Epigenomics Consortium were used to identify brain-expressed genes (Supplementary Information). We defined controls as parents of cases in the regression analysis. We used sex and the first three principal components of population stratification calculated using PLINK as covariates. The model was also corrected for the total length of CNVs. Finally, we performed a multivariate analysis to investigate whether the burden signals were from the same sets of CNVs as in the univariate analysis (details in Supplementary Information). We considered p < 0.05 as statistically significant. We also reported 0.05 < p < 0.1 as nearly significant.

Genomic instability and fragile sites

Replication stress can lead to CNV formation, and fragile sites. A recent study using genome-wide CNVs demonstrated a link between DNA/genomic integrity and ASD and SCZ. However, using a larger sample size than the current study (1,108 ASD and 2,458 SCZ), they were unable to find pathways enriched in ASD versus SCZ and vice versa. Given smaller cohorts, we performed our analyses in a combined set of all four NDDs to achieve an acceptable statistical power. We first investigated CNVs impacting the coding sequences of genomic instability genes, looking for change in the proportion of cases with rare CNVs in these genes, compared with that of controls. The genomic instability genes comprised 95 genes, including coding genes identified from the AmiGO database by searching for the following terms: DNA repair, DNA replication, genome maintenance, DNA damage, and DNA integrity. We also tested for the overall number of rare CNVs and total length of rare CNVs. We then considered whether cases with perturbed genomic instability genes had a different burden of rare CNVs in NDD genes compared to that in cases with intact instability genes.

Reporting summary

Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

DATA AVAILABILITY

Relevant microarray data are deposited in the database of Genotypes and Phenotypes (https://www.ncbi.nlm.nih.gov/gap/; IDPsh01881.v1.p1). The relevant CNVs are available in dbVar (IDStrd173).

Received: 6 May 2019; Accepted: 5 September 2019; Published online: 07 October 2019

REFERENCES

Received: 6 May 2019; Accepted: 5 September 2019; Published online: 07 October 2019

M. Zarrei et al.
19. Gonzalez-Martilla, A. J., Moreno-De-Luca, A., Ledbetter, D. H. & Martin, C. L. A cross-disorder method to identify novel candidate genes for developmental brain disorders. *JAMA Psychiatry* 73, 275–283 (2016).
20. Kushima, I. et al. Comparative analyses of copy-number variation in autism spectrum disorder and schizophrenia reveal etiological overlap and biological insights. *Cell Rep.* 24, 2838–2856 (2018).
21. Nishioch, M. et al. Psychiatric disorders in children with 16p11.2 deletion and duplication. *Transl. Psychiatry* 9, 8 (2019).
22. Uddin, M. et al. A high-resolution copy-number variation resource for clinical and population genetics. *Genet. Med.* 17, 747–752 (2015).
23. Prasad, A. et al. A discovery resource of rare copy number variations in individuals with autism spectrum disorder. *G3 (Bethesda)* 2, 1665–1685 (2012).
24. Costain, G. et al. Pathogenic rare copy number variants in community-based schizophrenia suggest a potential role for clinical microarrays. *Hum. Mol. Genet.* 22, 4485–4501 (2013).
25. Zarrei, M. et al. De novo and rare inherited copy-number variations in the hemiplegic form of cerebral palsy. *Genet. Med.* 20, 172–180 (2018).
26. Kearney, H. M. et al. American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. *Genet. Med.* 13, 680–685 (2011).
27. Yuen, R. K. et al. Whole-genome sequencing of quartet families with autism spectrum disorder. *Nat. Med.* 21, 185–191 (2015).
28. Jiang, Y. H. et al. Detection of clinically relevant genetic variants in autism spectrum disorder by whole-genome sequencing. *Am. J. Hum. Genet.* 93, 249–263 (2013).
29. van Rijn, S. et al. The nature of social cognitive deficits in children and adults with Klinefelter syndrome (47,XXY). *Genes Brain Behav.* 17, e12465 (2018).
30. Lionel, A. C. et al. Disruption of the AST2N/TRMM32 locus at Xq23.3 is a risk factor in males for autism spectrum disorders, ADHD and other neurodevelopmental phenotypes. *Hum. Mol. Genet.* 23, 2752–2768 (2014).
31. Duong, L. T. et al. Two rare deletions upstream of the NRXN1 gene (2p16.3) affecting the non-coding mRNA AK127244 segregate with diverse psycho-pathological phenotypes in a family. *Eur. J. Med. Genet.* 58, 650–653 (2015).
32. Rizzo, A. et al. The noncoding RNA AK127244 in 2p16.3 locus: a new susceptibility region for neuropsychiatric disorders. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 177, 557–562 (2018).
33. Nguyen, T. & Di Giovanni, S. NFAT signaling in neural development and axon growth. *Int. J. Dev. Neurosci.* 26, 141–145 (2008).
34. Noor, A. et al. Disruption at the PTCHD1 Locus on Xp22.11 in Autism spectrum disorders in males. *Cell Rep.* 6, 285–289 (2019).
35. Du, F. et al. Periodic reanalysis of whole-genome sequencing data enhances the discovery of clinical CNVs. *NPJ Genomic Medicine* (2019) 26 Published in partnership with CEGMR, King Abdulaziz University
36. Molloy, C. A. et al. Differences in the clinical presentation of Trisomy 21 with and without autism. *J. Intell. Disabil. Res.* 53, 143–151 (2009).
37. Kendall, K. M. et al. Cognitive performance among carriers of pathogenic copy number variant numbers: analysis of 152,000 UK biobank subjects. *Biol. Psychiatry* 82, 103–110 (2017).
38. Kovarikova, C., Costain, G., Baribeau, D. A. & Basset, A. S. Genomic disorders in psychiatry-what does the clinician need to know? *Curr. Psychiatry Rep.* 19, 82 (2017).
39. Oskouei, M. et al. Clinically relevant copy number variations detected in cerebral palsy. *Nat. Commun.* 6, 7949 (2015).
40. Mauger, C. et al. Executive functions in children and adolescents with Turner syndrome: a systematic review and meta-analysis. *Neuropsychol. Rev.* 28, 188–215 (2018).
41. Hardegen, A. et al. Identification of submicroscopic CNVs using high-resolution array-CGH in a cohort of 3,446 patients with developmental delay and behavioral issues. *Am. J. Med. Genet. A Hum. Genet.* 151A, 187–198 (2011).
42. Molloy, C. A. et al. Periodic reanalysis of whole-genome sequencing data enhances the diagnostic advantage over standard clinical genetic testing. *Eur. J. Hum. Genet.* 26, 740–744 (2018).
43. Diaz-Beltran, L. et al. Cross-disorder comparative analysis of comorbidity conditions reveals novel autism candidate genes. *BMJ Genomics* 18, 315 (2017).
44. Zeng, L. et al. Cross-disorder comparison of four neuropsychiatric CNV loci. *Curr. Genet. Med. Rep.* 2, 151–161 (2014).
45. Lukk, M. et al. A global map of human gene expression. *Nat. Biotechnol.* 28, 322–324 (2010).
46. Leik, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536, 285–291 (2016).
were obtained from dbGaP at https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000169.v1.p1 through dbGaP accession number phs000169.v1.p1.

**AUTHOR CONTRIBUTIONS**

M.Z. and S.W.S. conceived and designed the experiments. M.Z., C.L.B., W.E., E.J.Y., E.J.H., J.W., J.R.M., K.R., G.P., M.F., A.J.C., B.D.M., S.W., B.T., M.W., and D.M processed and analysed the microarray data. S.L. designed and performed experiments for variant characterization and validation. C.L., T.W., K.S., and X.W. performed the CytoScan HD microarray laboratory experiments. B.K. and J.L.W. helped perform different components of analysis and sample collection. M.Z., E.J.Y., E.J.H., G.C., M.R., R.K.C.U., J.A.B., J.A.S.V., R.V.P., C.R.M., R.F.W., and D.J.S. helped perform different components of analysis and data interpretations. M.W., C.C., L.Z., M.E., J.F., B.A.F., M.T.C., P.Z., X.L., R.N., G.L.H., D.R.R., S.G., R.W., T.G., M.S., I.D., M.W.S., P.D.A., A.S.B., J.C., R.S., and E.A. diagnosed, examined and recruited the participants. M.Z. and S.W.S. wrote the manuscript.

**COMPETING INTERESTS**

S.W.S. serves on the Scientific Advisory Committees of Population Bio and Deep Genomics; intellectual property originating from his research and held at the Hospital for Sick Children is licensed to Lineagen, and separately Athena Diagnostics. D.M. is a full-time employee of Deep Genomics and is entitled to a stock option. R.J.S., P.D.A., and J.C. consult for Highland Therapeutics. Intellectual property from ADHD research at the Hospital for Sick Children is licensed to Ehave and the National Research Council of Canada. Other authors declare no competing interests for the data and interpretation presented in this study. R.J.S., P.D.A., and J.C. consult for Highland Therapeutics. Intellectual property from their research at the Hospital for Sick Children is licensed to Ehave and the National Research Council. D.M. is a full-time employee of Deep Genomics and is entitled to stock options. S.W.S. is on the Scientific Advisory Committees of Population Bio and Deep Genomics; intellectual property from his research held at the Hospital for Sick Children is licensed to Athena Diagnostics, and separately to Lineagen. These relationships did not influence data interpretation or presentation during this study, but are disclosed for potential future consideration.

**ADDITIONAL INFORMATION**

Supplementary information is available for this paper at https://doi.org/10.1038/s41525-019-0098-3.

Correspondence and requests for materials should be addressed to S.W.S.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2019