The Penetrance of Copy Number Variations for Schizophrenia and Developmental Delay

George Kirov, Elliott Rees, James T.R. Walters, Valentina Escott-Price, Lyudmila Georgieva, Alexander L. Richards, Kimberly D. Chambert, Gerwyn Davies, Sophie E. Legge, Jennifer L. Moran, Steven A. McCarroll, Michael C. O'Donovan, and Michael J. Owen

Background: Several recurrent copy number variants (CNVs) have been shown to increase the risk of developing schizophrenia (SCZ), developmental delay (DD), autism spectrum disorders (ASD), and various congenital malformations (CM). Their penetrance for SCZ has been estimated to be modest. However, comparisons between their penetrance for SCZ or DD/ASD/CM, or estimates of the total penetrance for any of these disorders have not yet been made.

Methods: We use data from the largest available studies on SCZ and DD/ASD/CM, including a new sample of 6882 cases and 6316 controls, to estimate the frequencies of 70 implicated CNVs in carriers with these disorders, healthy control subjects, and the general population. On the basis of these frequencies, we estimate their penetrance. We also estimate the strength of the selection pressure against CNVs and correlate this against their overall penetrance.

Results: The rates of nearly all CNVs are higher in DD/ASD/CM compared with SCZ. The penetrance of CNVs is at least several times higher for the development of a disorder from the group of DD/ASD/CM. The overall penetrance of SCZ-associated CNVs for developing any disorder is high, ranging between 10.6% and 100%.

Conclusions: CNVs associated with SCZ have high pathogenicity. The majority of the increased risk conferred by CNVs is toward the development of an earlier-onset disorder, such as DD/ASD/CM, rather than SCZ. The penetrance of CNVs correlates strongly with their selection coefficients. The improved estimates of penetrance will provide crucial information for genetic counselling.

Key Words: Autism spectrum disorder, developmental delay, CNV, penetrance, schizophrenia, selection

A number of rare genomic rearrangements, called copy number variants (CNVs) have been shown to increase the risk of developing early-onset neurodevelopmental disorders. These were first identified in patients with characteristic and recognizable syndromic features (e.g., Williams-Beuren syndrome, Smith-Magenis syndrome, Sotos syndrome, DiGeorge/velo-cardio-facial syndrome [VCFS]). Over the past few years with the introduction of high-throughput microarray technologies, more CNVs of smaller size and incomplete penetrance have also been identified. Some of these have been shown to also increase the risk of developing SCZ, ASD, and other neuropsychiatric disorders. For example, in 2008–2009, a deletion at 15q13.3 was shown to increase the risk of developing developmental delay (DD) (1), schizophrenia (SCZ) (2,3), epilepsy (4), and autism (5). Similar findings of increased risk for developing SCZ, DD, and autism spectrum disorders (ASD) were made for deletions at 1q21.1 and 15q11.2 and duplications at 16p11.2 and 16p13.11 (reviewed by Malhotra and Sebat) (6). A number of CNVs have now been consistently associated with SCZ, and each of them also increases the risk for the group of DD/ASD/congenital malformations (CM) (6–9).

The number of CNVs known to increase the risk of developing a disorder from the group of DD/ASD/CM is higher than those implicated in SCZ. Thus, Girirajan *et al.* (7) tested 32,587 samples from children who had DD or ASD with or without CM for 72 CNV regions (39 deletions and 33 reciprocal duplications) that had previously been implicated in neurodevelopmental phenotypes or genomic disorders, including nine of uncertain pathogenic significance. When compared with a set of 8329 healthy control subjects, 38 (25 deletions and 13 duplications) were nominally statistically associated with the disorders (at $p < .05$), and several more showed trends that might also represent true associations if tested in larger samples. Similar results were reported by Kaminsky *et al.* (8) on 15,749 individuals who presented for diagnostic array testing with abnormal clinical phenotypes including DD, intellectual deficit, ASD, and/or multiple CM. These authors reported that 21 CNV regions (14 deletions and 7 duplications) were significantly associated with one or more of these disorders.

It is clinically important to know the risk to carriers of these CNVs for developing each of the possible associated disorders (i.e., their penetrance). Vassos *et al.* (9) were the first to estimate the penetrance for SCZ for seven CNVs that had been shown to increase risk for this disorder. They found rather modest rates of 2% to 7.4% except for the VCFS deletion on 22q11.2, which had a much higher penetrance of 55%, although with broad confidence intervals because no CNV was observed in control subjects. The authors concluded that these CNVs were neither necessary nor sufficient to cause the disorder and that the level of penetrance was not sufficient for them to be considered as useful clinical tools in genetic counselling, diagnosis, and testing. However, they pointed out that the overall penetrance for any neuropsychiatric disorder was likely to be much higher. The penetrance of 12 CNVs for DD/ASD/CM was estimated by Rosenfeld *et al.* (10). Estimates

From the Medical Research Council Centre for Neuropsychiatric Genetics and Genomics (GK, ER, JTRW, VE-P, LG, ALR, GD, SEL, MCO, MJO), Institute of Psychological Medicine and Clinical Neurosciences, Cardiff University, Cardiff, United Kingdom, and Stanley Center for Psychiatric Research (KDC, JLM, SAM), The Broad Institute of MIT and Harvard, Cambridge, Massachusetts.

Address correspondence to George Kirov, M.R.C.Psych., Ph.D., MRC Centre for Neuropsychiatric Genetics and Genomics, Institute of Psychological Medicine and Clinical Neurosciences, Cardiff University, Hadyn Ellis Building, Cardiff, CF24 4HQ, United Kingdom; E-mail: kirov@cardiff.ac.uk.

Received May 13, 2013; revised Jul 3, 2013; accepted Jul 17, 2013.

http://dx.doi.org/10.1016/j.biopsych.2013.07.022

of the risk for an abnormal phenotype ranged from 10.4% for 15q11.2 deletions to 62.4% for distal 16p11.2 deletions. These values are much higher than those for SCZ. The most highly penetrant CNVs were not tested because their absence in control subjects prevented accurate estimates.

Here we estimate the penetrance of all CNVs listed in the Girirajan *et al.* (7) article. Nearly all SCZ-associated CNVs are on this list as well, and we added only exonic deletions at *NRXN1*, a gene consistently implicated in SCZ (11–13). We performed estimates for both SCZ and the group of early-onset developmental disorders: DD/ASD/CM. This joint analysis allowed us to provide estimates for each CNV, even for those that are never found in SCZ or in healthy control subjects. We use a large new sample of cases and controls and add to this the data from the two largest previous studies on SCZ or from previous meta-analyzes to derive more reliable estimates.

Methods and Materials

Choice of CNVs

We analyzed CNVs previously associated with SCZ or severe neurodevelopmental phenotypes. These were taken from the list of CNV regions proposed by Girirajan *et al.* (7): 37 deletions and 32 reciprocal duplications, after excluding an overlapping segment at 17p13.3 and a CNV on the X chromosome because the X chromosome was not analyzed in our samples. Most of the strongly implicated SCZ-associated CNVs are on this list, but we added exonic deletions at *NRXN1*. For some of the analyzes, we focus on CNVs that we regard as associated with SCZ (Table 1), based on the review by Malhotra and Sebat (6), with the addition of *NRXN1* and newly implicated loci (Table S2 in Supplement 1). We did not analyze other SCZ-implicated loci (e.g., *VIPR2* duplications) because they have not been tested in sufficiently large samples of DD/ASD/CM or have not received consistent support.

Estimating the Rate of CNVs in Different Disorders

We only included data from the largest studies/samples available to simplify the presentation. The numbers that follow are those after exclusion of poorly performing arrays and duplicate samples. A CNV was called as covering a CNV locus if it spanned more than 50% of the commonly affected region (Table S2 in Supplement 1). In the case of loci that include only single genes (*NRXN1*, *SIM1*, *YWHAE*, *PAFAH1B1*, and *NF1*), we accepted CNVs that intersected at least one exon of the gene.

The rates of CNVs in DD/ASD/CM are taken from the largest study on these phenotypes: 32,587 patients referred for genetic testing to one laboratory (Signature Genomics) described by Girirajan *et al.* (7). For some of the loci, the reported sample numbers are smaller (23,380), for others they are larger (33,226) because the same team subsequently published data on several CNVs in an enlarged data set (10).

For SCZ cases, we analyzed three large data sets where we had access to the raw data, for a total of at least 13,465 cases (and more for the SCZ-associated loci, see below): 1) 6882 patients from our new Clozapine UK and Cardiff University Cognition in Schizophrenia samples (Supplement 1), 2) 3391 cases from the International Schizophrenia Consortium study (ISC, 2008) (2), and 3) 3192 cases from the Molecular Genetics of Schizophrenia (MGS) study (dbGAP accession numbers phs000167.v1.p1 and phs000021.v3.p2).

For controls, we analyzed samples from four publicly available data sets genotyped with high-resolution Illumina arrays (San Diego, California), similar to our new SCZ sample and analyzed by us with the same methods. These include individuals who took part in a study on smoking cessation in the United States (n = 1488); a study on melanoma in the United States (n = 2971); a study on refractive error Kooperative Gesundheitsforschung in der Region Augsburg (KORA) study from Germany (n = 1857), and the Wellcome Trust Case Control Consortium (WTCCC2) in the United Kingdom (n = 4939). To those we added 3181 control subjects from the ISC and 3437 control subjects from the MGS studies listed above, for a total of 17,873 control subjects.

For the SCZ-associated loci, we added data from previous studies, as reviewed by Malhotra and Sebat (6) or presented in the relevant articles that implicated them (Table S2 in Supplement 1). For these loci we excluded our WTCCC2 control subjects because they are completely or partially included in the previous reviews.

Estimation of the penetrance was performed with an adaptation of the method proposed by Vassos *et al.* (9). These authors estimated the penetrance as the probability of developing the disease (D) for individuals carrying the CNV (G) with the following formula:

$$P(D|G) = \frac{P(G|D)P(D)}{P(G|D)P(D) + P(G|\overline{D})P(\overline{D})}$$

Table 1. Frequencies and Penetrance for the Schizophrenia (SCZ)-associated Copy Number Variants Only

Locus	Selection Coefficient	Frequency %				Penetrance % (95% CI)		
		Controls	SCZ	DD/ASD/CM	General Population	SCZ	DD/CM/ASD	Total
1q21.1 del	.26	.021	.17	.29	.033	5.2 (2.5–11)	35 (18–67)	40 (20–78)
1q21.1 dup	.23	.038	.13	.2	.045	2.9 (1.3–6.3)	18 (10–33)	21 (11–39)
NRXN1 del	.23	.02	.18	.18	.028	6.4 (2.5–8.3)	26 (16–80)	33 (18–88)
3q29 del	.83	.0014	.082	.061	.0046	18 (4.7–67)	53 (15–100)	71 (20–100)
WBS dup	.61	.0058	.066	.12	.011	6.0 (1.4–20)	44 (13–100)	50 (14–100)
15q11.2 del	.09	.28	.59	.81	.3	2.0 (1.4–2.7)	11 (8.2–14)	13 (9.6–17)
Prader-Willi/Angelman dup	.5	.0083	.079	.25	.019	4.2 (1.4–12)	54 (25–100)	58 (26–100)
15q13.3 del	.31	.019	.14	.26	.03	4.7 (2.2–9.9)	35 (19–62)	40 (21–72)
16p13.11 dup	.13	.13	.31	.3	.14	2.2 (1.3–3.7)	8.4 (5.7–13)	10.6 (7–17)
16p11.2 distal del	.29	.018	.063	.14	.024	2.6 (.8–9.2)	23 (8.4–63)	26 (9.2–72)
16p11.2 dup	.33	.03	.35	.28	.043	8.0 (4.3–14)	26 (18–43)	34 (22–57)
17q12 del	.68	.0054	.036	.087	.009	4.0 (.8–18)	39 (13–100)	43 (14–100)
DiGeorge/VCFS del	.8	0	.29	.54	.024	12 (6.5–18)	88 (53–100)	100 (60–100)

The full list of CNVs is presented in Table S4 of Supplement 1.

ASD, autism spectrum disorders; CI, confidence interval; DD, developmental delay; del, deletion; dup, duplicate; CM, various congenital malformations; VCFS, velo-cardio-facial syndrome; WBS, Williams-Beuren syndrome.