Archival Report

Biological Psychiatry

Cognitive Performance Among Carriers of Pathogenic Copy Number Variants: Analysis of 152,000 UK Biobank Subjects

Kimberley M. Kendall, Elliott Rees, Valentina Escott-Price, Mark Einon, Rhys Thomas, Jonathan Hewitt, Michael C. O'Donovan, Michael J. Owen, James T.R. Walters, and George Kirov

ABSTRACT

BACKGROUND: The UK Biobank is a unique resource for biomedical research, with extensive phenotypic and genetic data on half a million adults from the general population. We aimed to examine the effect of neurodevelopmental copy number variants (CNVs) on the cognitive performance of participants.

METHODS: We used Affymetrix Power Tools and PennCNV-Affy software to analyze Affymetrix microarrays of the first 152,728 genotyped individuals. We annotated a list of 93 CNVs and compared their frequencies with control datasets. We analyzed the performance on seven cognitive tests of carriers of 12 CNVs associated with schizophrenia ($n = 1087$) and of carriers of another 41 neurodevelopmental CNVs ($n = 484$).

RESULTS: The frequencies of the 93 CNVs in the Biobank subjects were remarkably similar to those among 26,628 control subjects from other datasets. Carriers of schizophrenia-associated CNVs and of the group of 41 other neurodevelopmental CNVs had impaired performance on the cognitive tests, with nine of 14 comparisons remaining statistically significant after correction for multiple testing. They also had lower educational and occupational attainment (p values between 10^{-7} and 10^{-18}). The deficits in cognitive performance were modest (Z score reductions between 0.01 and 0.51), compared with individuals with schizophrenia in the Biobank (Z score reductions between 0.35 and 0.90).

CONCLUSIONS: This is the largest study on the cognitive phenotypes of CNVs to date. Adult carriers of neurodevelopmental CNVs from the general population have significant cognitive deficits. The UK Biobank will allow unprecedented opportunities for analysis of further phenotypic consequences of CNVs.

Keywords: Affymetrix, CNV, Cognition, Neurodevelopmental, Schizophrenia, UK Biobank

http://dx.doi.org/10.1016/j.biopsych.2016.08.014

Copy number variants (CNVs) are >1000 base pair DNA segments that are present at a variable copy number in comparison with a reference genome (1,2). CNVs may be recurrent and have similar breakpoints when they are formed by nonallelic homologous recombination between sites of low copy repeats, or nonrecurrent, with variable breakpoints, when formed as a result of defects in DNA replication or repair (3). There is an increased rate of CNVs in neurodevelopmental spectrum disorders, including intellectual disability (ID), autism spectrum disorder (ASD) (4,5), epilepsy (6), and schizophrenia (7). To date, 12 CNVs have been robustly associated with risk of schizophrenia, and they also increase risk of ASD and ID (7,8). Many more CNVs are implicated in ID, ASD, and cases with congenital anomalies (4,5) but they have not been implicated in schizophrenia, although this could be due to insufficient statistical power (8).

The phenotypes of many highly penetrant CNVs, such as those implicated in Prader-Willi syndrome/Angelman syndrome or DiGeorge syndrome, are well established. Many others have incomplete penetrance (9), and their phenotypic spectrum is not fully established (e.g., 1q21.1 duplication, 15q11.2 deletion). There are many adult carriers of incompletely penetrant CNVs in the general population, who have escaped the development of early-onset developmental disorders and are apparently healthy. However, they might still have an increased burden of cognitive or physical impairments.

Limited data, based on relatively small sample sizes, are available on the cognitive phenotypes of CNVs in adults. An Icelandic study of 144 carriers of 11 pathogenic CNVs found that healthy CNV carriers had impaired cognition and performed intermediately between CNV noncarriers and individuals with schizophrenia (10). An Estonian study of 56 CNV carriers found an association between rare CNVs and lower educational attainment (11).

The UK Biobank provides a great opportunity to study the effects of pathogenic CNVs on physical and mental health characteristics, especially the incompletely penetrant ones. The half a million individuals recruited by the Biobank have provided extensive demographic, health, and cognitive data; will be followed up prospectively; and are being genotyped on

ISSN: 0006-3223

Affymetrix microarrays. We aimed to identify pathogenic CNVs in the first 152,728 individuals in the UK Biobank for whom genotype data have been released so far and to analyze the cognitive consequences of neurodevelopmental CNVs.

METHODS AND MATERIALS

Participants

The UK Biobank (www.ukbiobank.ac.uk) recruited half a million participants in the United Kingdom between 2006 and 2010, aged 40–69 years, and 53% are women. All subjects provided informed consent to participate in UK Biobank projects and agreed to have their health followed over many years. Ethical approval for the study was granted by the North West Multi-Centre Ethics Committee. The Biobank used 22 assessment centers in England, Scotland, and Wales. Participants were recruited from National Health Service patient registers, with no exclusion criteria, provided they lived within a reasonable proximity to an assessment center. Participants spent approximately 3 hours at the assessment centers, providing detailed demographic, socioeconomic, and health-related data via a touch screen questionnaire. After this, a trained nurse performed an interview to clarify any questions arising from the touch screen questionnaire. Participants underwent several physical assessment measures and provided blood, urine, and saliva samples. Data were released to Cardiff University after application to the UK Biobank.

Genotyping

Samples were genotyped at Affymetrix Research Services Laboratory, Santa Clara, CA. Approximately 100,000 samples were genotyped on the UK Biobank Axiom Array (820,967 probes), and approximately 50,000 samples were genotyped on the UK BiLEVE Array (807,411 probes) (12). There is 95% common content between the two arrays (https://biobank.ctsu.ox.ac.uk/crystal/docs/genotyping_sample_workflow.pdf). Sample processing at UK Biobank is described in Supplement 1. Here, we present data from 152,728 individuals genotyped in the first phase of genotyping, of which 151,659 passed our quality control (QC) filters (Table S1 in Supplement 1).

CNV Calling

In our previous work on CNVs, we analyzed datasets of up to 20,000 samples at a time (13–16). To call CNVs in the UK Biobank, we had to substantially increase the speed of processing by analyzing batches in parallel; standardizing QC processes across all batches; and omitting Z score analysis of CNVs, reclustering after removal of poorly performing samples, and performing manual inspection of CNV traces. Anonymized genotype data were downloaded as raw (CEL) files from the UK Biobank website, stored on a secure Linux server, and analyzed with UNIX-based commands (detailed in Supplement 1). Briefly, we used the apt-probeset-genotype command with Affymetrix Power Tools software (www.affymetrix.com/estore/partners_programs/programs/developer/tools/powertools.affx) to generate normalized signal intensity data, genotype calls, and confidences. We analyzed only the approximately 750,000 biallelic markers because these can be used in PennCNV-Affy to generate cluster plots. Each predefined batch of approximately 4600 CEL files was analyzed separately to reduce potential batch effects. The genotype calls, confidences, and summary files were processed with PennCNV-Affy software (17). We generated canonical genotype clusters, log-R ratios, and B allele frequencies and completed the subsequent steps recommended in PennCNV for the generation of CNV calls (17). After CNV detection, adjacent CNVs were joined if separated by $<25\%$ of their combined length. Individual samples were excluded if they had ≥ 30 CNVs, had a waviness factor >0.03 or <-0.03, or had a call rate $<96\%$. Individual CNVs were excluded if they were covered by <10 probes or had a density coverage of <1 probe per 20,000 base pairs. A total of 1069 samples (0.7%) were excluded during QC (Table S1 in Supplement 1).

CNV Annotation

We compiled a list of 93 CNVs proposed to be pathogenic in two widely accepted sources (4,5). The full list with the corresponding critical region coordinates is presented in Table S2 in Supplement 2. The breakpoints of the initially called CNVs were inspected to confirm that they met our CNV calling criteria (Table S4 in Supplement 1). Briefly, we required a CNV to cover more than half of the critical interval and to include the key genes in the region (if known), or in the case of single gene CNVs (e.g., *NRXN1*) we required deletions to intersect at least one exon and duplications to cover the whole gene. For comparison of CNV frequencies with previous results, we used data from 26,628 control individuals from other large datasets, for which we have access to the raw CNV data and used the same CNV calling methods (Table S3 in Supplement 1) (7,18,19). Because not all 93 of these CNVs are known to affect cognition (or such a link has not been statistically confirmed), for our analyses we selected 53 of these CNVs that have been statistically associated with neurodevelopmental phenotypes (4) (after removing the common duplications at 15q11.2, because they would account for more than half of the other neurodevelopmental CNVs, thus skewing all analyses). The 53 CNVs were further subdivided into 12 CNVs associated with schizophrenia (7,8), and the remaining 41 were denoted as other neurodevelopmental CNVs (Table S2 in Supplement 2).

Cognitive Tests

Participants completed a battery of cognitive tests organized by the UK Biobank. We chose to analyze tests done by at least 10% of participants and restricted the analysis to individuals who self-reported as being of white British or Irish descent. Different numbers of participants were asked by the Biobank to complete the various tests, ranging from nearly the complete sample (Pairs Matching, Reaction Time) to only parts of the sample (the remaining tests). The scores were first normalized, if not normally distributed, and then converted to Z scores (Supplement 1). We analyzed data on the following tests.

Pairs Matching Test. This tested episodic memory and was completed at the assessment centers at the first visit by 136,292 individuals with available genotypes. Participants

were shown 6 pairs of cards for 3 seconds, which were then turned over. Participants were asked to identify the matching pairs. We used the total number of errors made during this task and restricted our analyses to individuals who finished the test. We applied a log +1 transformation to these data.

Reaction Time Test. This tested simple processing speed and was completed at the assessment centers at the first visit by 138,603 individuals with available genotypes. Participants were asked to play 12 rounds of a computerized Snap game in which they had to click a button as quickly as possible when shown two matching cards. We used the mean reaction time from their attempts. The data were log-transformed.

Fluid Intelligence Test. This tested reasoning and problem solving and was completed at the assessment centers at the first visit by 44,575 individuals with available genotypes; that is, this test was not given to everybody. Participants were presented with 13 verbal and numerical reasoning questions and had to answer as many as they could within 2 minutes. We used the total number of correct answers for our analyses. Data were normally distributed and did not require transformation.

Digit Span Test. This tested numeric working memory and was completed at the assessment centers at the first visit by 14,495 individuals with available genotypes. Participants were presented with progressively longer numbers (maximum 12) and asked to enter them back once the number had disappeared. We used the maximum number of digits remembered for our analyses. Data were normally distributed and did not require transformation.

Symbol Digit Substitution Test. This tested complex processing speed and was completed at follow-up on home computers by 33,057 individuals with available genotypes. This test involves matching numbers to a set of symbols. We used the number of correct substitutions for our analyses. There were small numbers of outliers, that suggested technical issues, so we excluded results of <3 and >36 substitutions. The remaining scores were normally distributed and did not require transformation.

Trail Making Test A and B. This tested visual attention and was completed at follow-up on home computers by 29,251 individuals with available genotypes. Participants were asked to connect scattered circles according to numbers (trail A) and to alternating numbers and letters (trail B). We used the time taken to complete these tests for our analyses, and these data were log-transformed before the generation of *Z* scores.

For comparison, we also analyzed data obtained from 507 individuals from the Biobank, who self-reported a diagnosis of schizophrenia at the interviews at the assessment centers, because schizophrenia is associated with impaired cognitive performance (10). Only 169 of the 507 have been genotyped so far, and just three of them carried a CNV from our list of 53 loci.

We calculated the effect size reductions on cognitive tests of the carriers of the two CNV groups, and of individuals with schizophrenia, in linear regression analyses, corrected for age and sex (Table 1). We also compared the carriers of schizophrenia CNVs with carriers with other neurodevelopmental CNVs using linear regression analyses, corrected for age and sex (Table 2).

Educational attainment and occupational level are highly correlated with cognitive performance (20,21). We analyzed these variables using ordinal regression with CNV carrier status and sex as factors and age as a covariate.

RESULTS

We found that 151,659 samples (99.3%) passed QC; 790,761 CNVs were retained after QC (without filtering for frequency or size), an average of 5.2 per person (range, 0–29; Figure S1 in Supplement 1). The frequencies of individual CNVs were

Table 1. Results on Cognitive Tests in the Two Groups of CNV Carriers and Individuals With Schizophrenia, Compared With CNV Noncarriers

	CNV Noncarriers	Carriers of Schizophrenia CNVs			Carriers of Other Neurodevelopmental CNVs			Individuals With Schizophrenia		
Cognitive Test	*n*	*n*	B (SEM)	*p*	*n*	B (SEM)	*p*	*n*	B (SEM)	*p*
Pairs Matching, Number of Incorrect Matches	134,781	1048	0.09 (0.03)	.003	463	0.17 (0.05)	3.2×10^{-4}	436	0.35 (0.05)	7.82×10^{-14}
Reaction Time, Mean Time to Correctly Identify Matches	137,053	1073	0.21 (0.03)	1.52×10^{-13}	477	0.35 (0.04)	9.46×10^{-16}	480	0.69 (0.04)	1.05×10^{-57}
Fluid Intelligence, Score	44,107	321	−0.48 (0.05)	5.29×10^{-19}	147	−0.51 (0.08)	3.97×10^{-10}	168	−0.61 (0.08)	2.51×10^{-15}
Digit Span, Number of Digits Remembered	14,343	102	−0.42 (0.09)	1.8×10^{-5}	50	−0.01 (0.14)	.931	39	−0.78 (0.16)	8.7×10^{-7}
Symbol Digit Substitution, Number of Correct Matches	32,770	204	−0.29 (0.06)	3.0×10^{-6}	83	−0.22 (0.09)	.024	42	−0.90 (0.14)	8.33×10^{-11}
Trail Making Test A, Time to Complete	28,988	185	0.19 (0.07)	.008	78	0.14 (0.11)	.199	36	0.83 (0.16)	1.51×10^{-7}
Trail Making Test B, Time to Complete	28,988	185	0.38 (0.07)	2.07×10^{-8}	78	0.28 (0.10)	.008	36	0.72 (0.15)	3.00×10^{-6}

B indicates the differences from CNV noncarriers, expressed as *Z* scores (SEM) in linear regression analysis, corrected for age and sex. CNV noncarriers, by definition, have *Z* scores of 0.

CNV, copy number variant.

Table 2. Comparison of Cognitive Tests in Carriers of 12 Schizophrenia-Associated CNVs and Individuals With Other Neurodevelopmental CNVs

Cognitive Test	Carriers of Schizophrenia CNVs, *n*	Carriers of Other Neurodevelopmental CNVs, *n*	B (SEM)	*p*
Pairs Matching	1048	463	−0.02 (0.05)	.74
Reaction Time	1073	477	−0.13 (0.06)	.026
Fluid Intelligence Score	321	147	0.046 (0.09)	.63
Digit Span	102	50	−0.39 (0.18)	.031
Symbol Digit Substitution	204	83	−0.08 (0.12)	.52
Trail Making Test A	185	78	0.06 (0.13)	.68
Trail Making Test B	185	78	0.11 (0.13)	.41

B indicates the differences between the two groups, expressed as *Z* score (SEM) as in Table 1. The *p* values are based on linear regression analysis, as above.

CNV, copy number variant.

consistent between batches, indicating a lack of significant batch effects (Table S12 in Supplement 2). Overall, the CNVs from the UK Biobank sample occurred at rates strikingly similar to those among the control datasets genotyped on different arrays and called by us with the same methods (Table S2 in Supplement 2). Only two CNVs reached nominally significant difference between the two datasets, but neither survived correction for multiple testing of 93 loci. We found that 3.8% of people in the Biobank carry a CNV from the list of 93 that were annotated. (The list of CNVs will be made available for download from the UK Biobank.) Of those CNVs, 54 have been shown to be significantly associated with neurodevelopmental disorders (4), including all 12 schizophrenia CNVs. In the Biobank, 1.12% of participants carried one of these neurodevelopmental CNVs (after excluding the common 15q11.2 duplication, found in 0.5% of subjects).

Cognitive Test Results

Carriers of both the schizophrenia CNVs and the other neurodevelopmental CNVs had impaired performance on the seven cognitive tests, compared with CNV noncarriers, with 9 of the 14 comparisons reaching statistically significant differences that survive a conservative Bonferroni correction for multiple testing for 14 tests. Table 1 presents these differences, expressed as unstandardized B coefficients (*Z* scores), corrected for age and sex in linear regression analyses. Most differences were modest in magnitude (*Z* scores between 0.01 and 0.51 below the CNV noncarriers). Individuals with schizophrenia performed worse than either group of CNV carriers (*Z* scores between 0.35 and 0.90 below the CNV noncarriers), and all differences were highly significant.

We then examined the differences in cognitive performance between carriers of the 12 schizophrenia CNVs ($n = 1087$) and carriers of the remaining 41 neurodevelopmental CNVs ($n = 484$), generated again from linear regression analysis, corrected for age and sex. Their performance tended to be similar (Figure 1, Table 2). Although two of the tests reached nominal levels of statistical significance, these do not survive correction for multiple testing for seven tests and were in opposite directions (Table 2).

Educational and Occupational Attainment

We compared the educational and occupational attainment of neurodevelopmental CNV carriers against the CNV noncarriers. Both groups of CNV carriers attained lower educational qualifications; for example, a smaller proportion obtained a university/college degree or achieved A/AS levels at school (postcompulsory education qualifications taken at 16–18 years of age; Figure 2). We performed ordinal regression analysis, with qualifications as the dependent variable, CNV status and sex as factors, and age as a covariate. This indicated lower odds (0.61) for carriers of schizophrenia CNVs to finish in a higher qualifications group (95% confidence interval [CI] 0.55 to 0.68, Wald 76.3, $p = 2.4 \times 10^{-18}$). Similar results were found for carriers of the other neurodevelopmental CNVs (lower odds 0.54, 95% CI 0.46 to 0.64, Wald 52.5, $p = 4.4 \times 10^{-13}$). CNV carriers also tended to have occupations that require less training or academic skills (Figure 3). Ordinal regression analysis, with major job group as the dependent variable, CNV status and sex as factors, and age as a covariate indicated lower odds (0.64) for carriers of schizophrenia CNVs to have a job in an occupational group that requires higher skills and longer training, as defined by Office of National Statistics (22) (95% CI 0.56 to 0.73, Wald 43.7, $p = 3.7 \times 10^{-11}$). Similar results were found for the carriers of other neurodevelopmental CNVs (lower odds 0.58, 95% CI 0.47 to 0.71, Wald 28.4, $p = 1.0 \times 10^{-7}$).

DISCUSSION

CNVs are a rare but important cause of serious neurodevelopmental disorders such as ID, ASD, schizophrenia, and a variety of congenital malformations (4,5,23). The UK Biobank sample, with half a million participants, is a unique resource for establishing the effects of CNVs on phenotypic outcomes. The data quality was high as indexed by the low fraction of samples that failed our QC (0.7%). Identical QC steps allowed us to call reliably the selected set of pathogenic CNVs, but researchers wanting to analyze smaller or common CNVs might have to use different filtering criteria.

Frequencies of Pathogenic CNVs

We established the frequencies of a set of 93 CNVs that have been proposed to be pathogenic (4,5) in approximately 150,000 participants in the UK Biobank genotyped so far (Table S2 in Supplement 2). We compared these with a large control dataset comprising 26,628 people, where we had access to raw CNV data or had ourselves called the CNVs

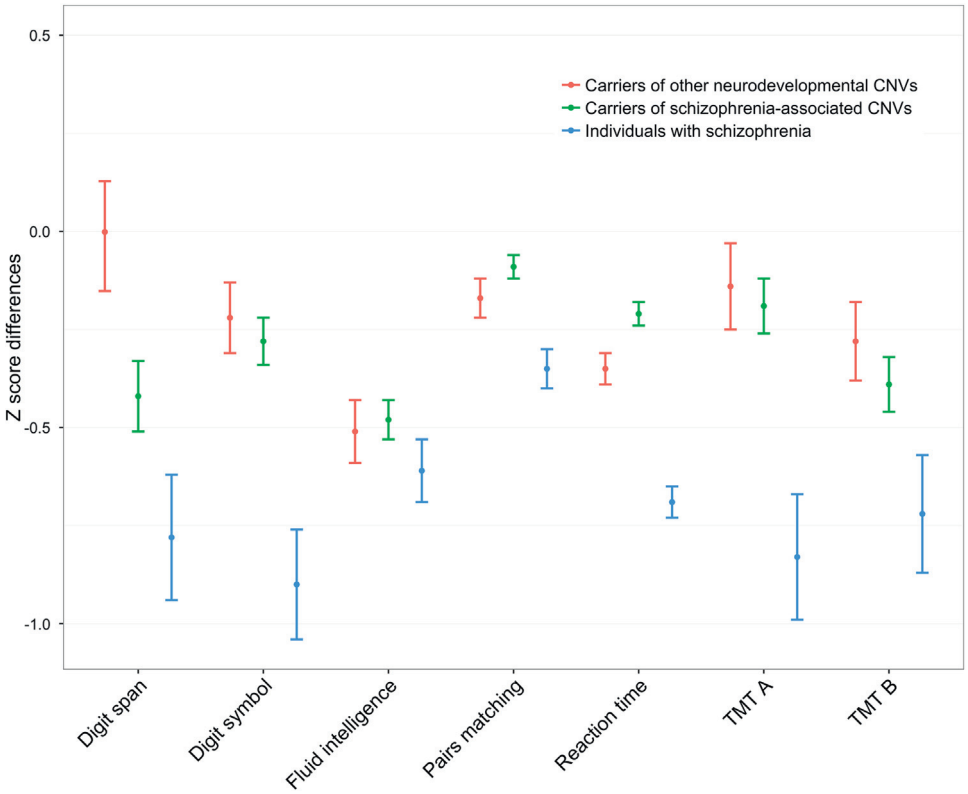

Figure 1. *Z* score differences in cognitive performance. Shown are the *Z* score differences for seven cognitive tests in the different groups of individuals, after correction for age and sex in a linear regression analysis. The bars represent the *Z* score means and SEMs. In blue are the scores among individuals with schizophrenia (including those that have not been genotyped), in green are the scores of carriers of schizophrenia-associated copy number variants (CNVs), and in red are those of carriers of other neurodevelopmental CNVs. A minus sign on the x axis indicates a worse score for all tests (e.g., a lower score or a longer time to complete a test). TMT A, Trail Making Test with numbers; TMT B, Trail Making Test with alternating numbers and letters.

from raw microarray files (Illumina or other versions of Affymetrix arrays). In the absence of an opportunity to perform technical replication on different arrays, we reasoned that finding similar CNV frequencies was the best validation of our CNV calling. These were indeed remarkably similar, with just two reaching nominally significant differences that would not survive correction for the multiple testing involved (Table S2 in Supplement 2).

Carriers of CNVs Implicated in Neurodevelopmental Phenotypes Have Reduced Cognitive Performance

To assess the cognitive performance of CNV carriers, we first selected, from the 93 annotated CNVs, a list of 54 CNVs that have been statistically associated with neurodevelopmental phenotypes, such as ID and ASD (4). This was done to exclude CNVs that are not confirmed to be associated with cognitive impairment (although they could be pathogenic for other medical conditions). We excluded from this list the common 15q11.2 duplication, found in 0.5% of the sample, because it would disproportionately affect the results.

Each CNV is likely to have its own set of phenotypic characteristics and cognitive signature. It is premature to analyze each one separately, because the study will be better powered for such analysis after all participants are genotyped. To provide an initial subgroup analysis, we divided the CNVs into a set of 12 that have been confirmed as associated with schizophrenia (7,8), and the 41 other neurodevelopmental CNVs. This division is unlikely to represent an actual dichotomy, because all 12 schizophrenia loci are also neurodevelopmental ones. We recently proposed (8) that many of the other neurodevelopmental CNVs increase risk of schizophrenia, but this has escaped statistical confirmation due to their rarity. Our finding of similar cognitive deficit among carriers of the two groups provides another argument that this distinction is somewhat arbitrary.

A study on the Icelandic population found reduced cognitive performance in 144 healthy carriers of 11 pathogenic CNVs: 1q21.1dup, *NRXN1*del, 13q31.3dup, 15q11.2del, 16p12.1del, 16p11.2del+dup, 16p13.11dup, 17p12del+dup, and 22q11.21dup (10). Most of these (except the 13q31.3dup and 17p12del+dup) are on our list of 53 neurodevelopmental loci. We analyzed the participants' performance on seven cognitive tests. To allow comparison with previous studies (10), we also show results for 507 individuals recruited in the Biobank, who self-reported to have schizophrenia, because they are expected to perform even worse on these tests. Carriers of neurodevelopmental CNVs (of both groups) performed intermediately between noncarriers and individuals with schizophrenia, with reductions between 0.01 and 0.51 SDs (*Z* scores), compared with CNV noncarriers. The results reached significance on most tests, even if corrected conservatively for multiple testing for 14 tests (Figure 1, Table 1). The two CNV groups were similar to each other, and no *p* value would survive correction for multiple testing (Table 2). Individuals with schizophrenia performed worse than any other group on all tests, at 0.35 to 0.90 SDs (*Z* score) reductions from CNV noncarriers.

It is well established that cognitive performance predicts achievement at school and employment (20,21,24). Here, we show that adult carriers of neurodevelopmental CNVs also had

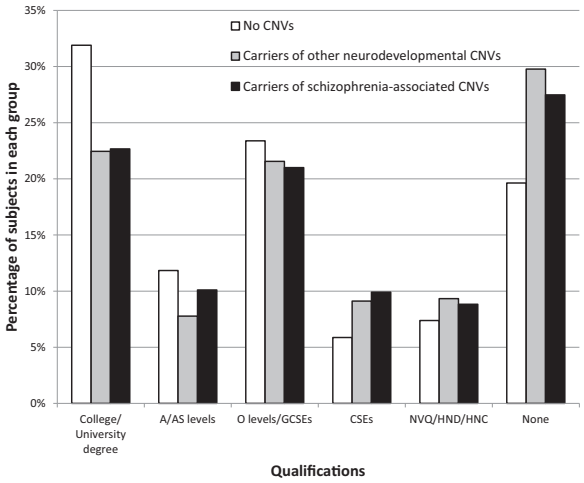

Figure 2. Distribution of the two groups of copy number variant (CNV) carriers and CNV noncarriers in each educational qualification group. The British qualifications are grouped as follows: college/university degree; advanced (A)/AS levels or equivalent: qualifications taken at 16–18 years of age, postcompulsory education; Ordinary (O) levels/General Certificates of Secondary Education (GCSEs) or equivalent: qualifications taken at 14–16 years of age, at the end of compulsory education; Certificates of Secondary Education (CSEs) or equivalent: a predecessor to GCSEs, including vocational subjects; National Vocational Qualification (NVQ) or Higher National Diploma (HND) or Higher National Certificate (HNC) or equivalent: vocational qualifications. Black bars represent schizophrenia-related CNV carriers; gray bars, other neurodevelopmental CNV carriers; white bars, CNV noncarriers.

lower educational attainment and tended to have occupations requiring less time in training, with all comparisons being highly significant. We wanted to address the question of whether the effect on school/occupational attainment is entirely explained by a reduction in cognitive performance among CNV carriers. We tested in a logistic regression analysis the effect of CNV carrier status on educational and occupational attainment, with and without the fluid intelligence score as a covariate (this test showed the highest effect size). Most of the effect of the CNV status on education/occupation was explained via the effect of the fluid intelligence score (Table S7 in Supplement 1), indicating that while cognitive impairment has a major effect on educational and occupational attainment, other phenotypic consequences of having a pathogenic CNV also play a role (Supplement 1). We note that 30.9% of carriers of neurodevelopmental CNV hold managerial or professional occupations and that the distribution of their cognitive tests performance overlaps with that of CNV noncarriers, with only a modest shift (see Figure S7C in Supplement 1 as an example for the Fluid Intelligence Test scores). This suggests that significantly impaired performance in the presence of a pathogenic CNV is not inevitable, at least for many CNV loci. It is possible that these highly functioning individuals may have performed even better had they not carried a CNV. It has been suggested that pathogenic CNVs produce a consistent degree of cognitive impairment (25), in the context of the individual's genetic background (e.g., by 2 SDs in the case of 22q11.2 deletions). Our results for the Fluid Intelligence Test are consistent with this observation (Figure S7C in Supplement 1), albeit with a more modest difference.

Most previous studies on pathogenic CNVs have recruited individuals from health services for ID, ASD, congenital anomalies, or schizophrenia. Much less is known about the effects of these CNVs in adults from the general population (9,10). Carriers of neurodevelopmental CNVs had reduced cognitive performance and educational and occupational attainment, with highly significant differences compared with CNV noncarriers. The effect size however is modest, <0.5 SD for all tests, with large overlap between the groups (Figure S7C in Supplement 1). We suggest that this may partly be

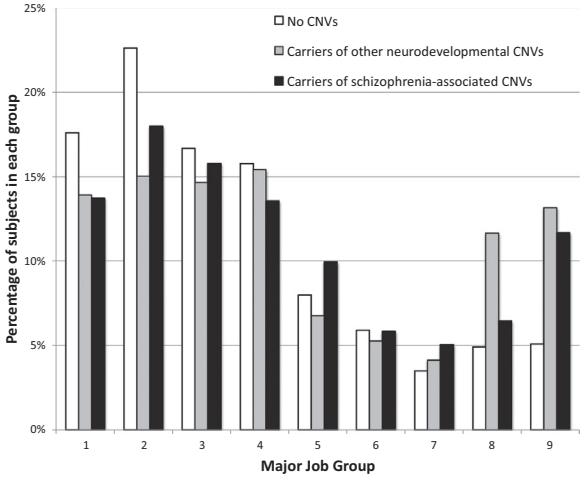

Figure 3. Distribution of the two groups of copy number variant (CNV) carriers and CNV noncarriers in each major job group as defined by the Office of National Statistics (22). X axis coding: 1, managers and senior officials; 2, professional occupations; 3, associate professionals and technical occupations; 4, administrative and secretarial occupations; 5, skilled trades occupations; 6, personal service occupations; 7, sale and customer service occupations; 8, process, plant, and machine operatives; 9, elementary occupations. Black bars represent schizophrenia-related CNV carriers; gray bars, other neurodevelopmental CNV carriers; white bars, CNV noncarriers.

explained by many severely affected CNV carriers, such as adults with ID, having not taken part in UK Biobank. This would result in highly functioning individuals being overrepresented. This may be in part due to the recruitment strategy and in part because some people with complex disabilities may have died before the recruitment age of 40–69 years. For example, there were only 5 individuals with 22q11.2 deletions, while we would expect about 37 carriers in a population of this size (the rate of this deletion among newborns is approximately 1:4000) (9). This is a severe disorder, with a decreased IQ of approximately 30 points, multiple congenital malformations, and reduced life expectancy. Similarly, there were no cases consistent with Prader-Willi/Angelman syndromes or Down syndrome, and there was only a single case with deletion at the Smith-Magenis syndrome region. Individuals with schizophrenia are also underrepresented, at only 0.12%, despite the disorder's lifetime risk of 0.4%–0.5% (26). An even smaller proportion of such individuals took part in the follow-up cognitive tests (e.g., only 7% of individuals with schizophrenia completed the Trail Making Tests at follow-up, compared with 21% of the remaining participants). Therefore, it is reasonable to assume that, although the UK Biobank attempted to recruit a sample that reflects the general population, it probably underrepresents seriously affected individuals. By analogy, carriers of pathogenic CNVs who have taken part in the UK Biobank might be among the higher functioning CNV carriers. The UK Biobank is recognized to be a generalizable sample, rather than one that is representative of the general population (27). However, it has been suggested that for data sources of this magnitude, data generated can still be applied to the population as a whole (28).

To further assess the potential of the Biobank to discover phenotypes caused by pathogenic CNVs, we checked whether we can detect the clearly defined phenotypes of certain common and incompletely penetrant CNVs. We examined CNVs at 16p11.2 (44 deletions and 42 duplications) and 17p12 (84 deletions and 45 duplications). Deletions at 16p11.2 have been associated with obesity, while duplications at the same locus have been associated with reduced weight (29). Carriers of deletions and duplications at this locus are expected to have reduced cognitive performance (30). In contrast, deletions and duplications at 17p12 cause peripheral neuropathies: hereditary neuropathy with liability to pressure palsies for deletions (31), or Charcot-Marie-Tooth disease type 1A, for duplications (32), but do not have an associated cognitive phenotype. These phenotypes were detected with high levels of statistical significance (Supplement 1). Thus, there were highly increased rates of peripheral neuropathy but normal cognitive performance in 17p12 CNV carriers, while 16p11.2 carriers had reduced cognitive performance and increased or reduced body mass index for deletions and duplications, respectively.

The full Biobank dataset will allow detailed analysis on the health consequences of many more individual CNV loci, and we will provide our list of CNVs to the Biobank, to assist researchers.

Limitations

There are several limitations to this study. While the UK Biobank made attempts to make their sample as representative as possible, the low proportion of individuals with severe disorders such as schizophrenia means that the sample cannot be considered perfectly representative of the general population (discussed above). There was also a considerable variation in the number of people who were approached to perform each cognitive test, and it is possible that some tests are more likely to have been performed by higher functioning individuals. This results in large variability in power between the tests and limits inferences that may be made for the tests with a lower completion rate.

ACKNOWLEDGMENTS AND DISCLOSURES

This work used the UK Biobank health resource. The work at Cardiff University was funded by the Medical Research Council Centre Grant No. MR/L010305/1 (to MJO and MCO) and Program Grant No. G0800509 (to MJO, MCO, and GK).

We thank Geoff Scopes (Affymetrix) and Kai Wang (author of PennCNV) for their advice for the analysis of Axiom arrays and Louise Wain and Martin Tobin for allowing us access to the library files on BiLEVE arrays.

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

ARTICLE INFORMATION

From the MRC Centre for Neuropsychiatric Genetics & Genomics (KMK, ER, VE-P, ME, RT, MCO, MJO, JTRW, GK), Division of Psychological Medicine and Clinical Neuroscience, Cardiff University School of Medicine, Cardiff; and Department of Geriatric Medicine (JH), Division of Population Medicine, Llandough Hospital, Penarth, United Kingdom.

Address correspondence to George Kirov, Ph.D., Cardiff University, Department of Psychological Medicine and Neurology, Maindy Road, Hadyn Ellis Building, Cardiff, CF24 4HQ, United Kingdom; E-mail: kirov@cardiff.ac.uk.

Received May 18, 2016; revised Aug 5, 2016; accepted Aug 7, 2016.

Supplementary material cited in this article is available online at http://dx.doi.org/10.1016/j.biopsych.2016.08.014.

REFERENCES

1. Lee C, Scherer SW (2010): The clinical context of copy number variation in the human genome. Expert Rev Mol Med 12:e8.
2. Feuk L, Carson AR, Scherer SW (2006): Structural variation in the human genome. Nat Rev Genet 7:85–97.
3. Shaffer LG, Lupski JR (2000): Molecular mechanisms for constitutional chromosomal rearrangements in humans. Annu Rev Genet 34: 297–329.
4. Coe BP, Witherspoon K, Rosenfeld JA, van Bon BW, Vulto-van Silfhout AT, Bosco P, *et al*. (2014): Refining analyses of copy number variation identifies specific genes associated with developmental delay. Nat Genet 46:1063–1071.
5. Dittwald P, Gambin T, Szafranski P, Li J, Amato S, Divon MY, *et al*. (2013): NAHR-mediated copy-number variants in a clinical population: Mechanistic insights into both genomic disorders and Mendelizing traits. Genome Res 23:1395–1409.
6. Lal D, Ruppert AK, Trucks H, Schulz H, de Kovel CG, Kasteleijn-Nolst Trenité D, *et al*. (2015): Burden analysis of rare microdeletions suggests a strong impact of neurodevelopmental genes in genetic generalised epilepsies. PLoS Genet 11:e1005226.
7. Rees E, Walters JT, Georgieva L, Isles AR, Chambert KD, Richards AL, *et al*. (2014): Analysis of copy number variations at 15 schizophrenia-associated loci. Br J Psychiatry 204:108–114.
8. Rees E, Kendall K, Pardinas A, Legge S, Pocklington A, Escott-Price A, *et al*. (2016): Analysis of intellectual disability copy number variants for association with schizophrenia [published online ahead of print Aug 17]. JAMA Psychiatry.

9. Kirov G, Rees E, Walters JT, Escott-Price V, Georgieva L, Richards AL, *et al*. (2014): The penetrance of copy number variations for schizophrenia and developmental delay. Biol Psychiatry 75:378–385.
10. Stefansson H, Meyer-Lindenberg A, Steinberg S, Magnusdottir B, Morgen K, Arnarsdottir S, *et al*. (2014): CNVs conferring risk of autism or schizophrenia affect cognition in controls. Nature 505:361–366.
11. Männik K, Mägi R, Macé A, Cole B, Guyatt AL, Shihab HA, *et al*. (2015): Copy number variations and cognitive phenotypes in unselected populations. JAMA 313:2044–2054.
12. Wain LV, Shrine N, Miller S, Jackson VE, Ntalla I, Soler Artigas M, *et al*. (2015): Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): A genetic association study in UK Biobank. Lancet Respir Med 3: 769–781.
13. Rees E, Walters JT, Chambert KD, O'Dushlaine C, Szatkiewicz J, Richards AL, *et al*. (2014): CNV analysis in a large schizophrenia sample implicates deletions at 16p12.1 and SLC1A1 and duplications at 1p36.33 and CGNL1. Hum Mol Genet 23:1669–1676.
14. Kirov G, Pocklington AJ, Holmans P, Ivanov D, Ikeda M, Ruderfer D, *et al*. (2012): De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. Mol Psychiatry 17:142–153.
15. Ikeda M, Aleksic B, Kirov G, Kinoshita Y, Yamanouchi Y, Kitajima T, *et al*. (2010): Copy number variation in schizophrenia in the Japanese population. Biol Psychiatry 67:283–286.
16. Grozeva D, Kirov G, Conrad DF, Barnes CP, Hurles M, Owen MJ, *et al*. (2013): Reduced burden of very large and rare CNVs in bipolar affective disorder. Bipolar Disord 15:893–898.
17. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF, *et al*. (2007): PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res 17:1665–1674.
18. Levinson DF, Duan J, Oh S, Wang K, Sanders AR, Shi J, *et al*. (2011): Copy number variants in schizophrenia: Confirmation of five previous findings and new evidence for 3q29 microdeletions and VIPR2 duplications. Am J Psychiatry 168:302–316.
19. International Schizophrenia Consortium (2008): Rare chromosomal deletions and duplications increase risk of schizophrenia. Nature 455:237–241.
20. Wagner R (1997): Intelligence, training and employment. Am Psychologist 52:1059–1069.
21. Schmidt F (2002): The role of general cognitive ability and job performance: Why there cannot be a debate. Hum Performance 15:187–210.
22. Office for National Statistics (2000): Standard Occupational Classification 2000. London: The Stationery Office.
23. Kirov G (2015): CNVs in neuropsychiatric disorders. Hum Mol Genet 24:R45–R49.
24. Strenze T (2007): Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. Intelligence 35:401–426.
25. Moreno-De-Luca A, Myers SM, Challman TD, Moreno-De-Luca D, Evans DW, Ledbetter DH (2013): Developmental brain dysfunction: Revival and expansion of old concepts based on new genetic evidence. Lancet Neurol 12:406–414.
26. McGrath J, Saha S, Chant D, Welham J (2008): Schizophrenia: A concise overview of incidence, prevalence, and mortality. Epidemiol Rev 30:67–76.
27. Collins R (2012): What makes UK Biobank special? Lancet 379: 1173–1174.
28. Manolio TA, Collins R (2010): Enhancing the feasibility of large cohort studies. JAMA 304:2290–2291.
29. McCarthy SE, Makarov V, Kirov G, Addington AM, McClellan J, Yoon S, *et al*. (2009): Microduplications of 16p11.2 are associated with schizophrenia. Nat Genet 41:1223–1227.
30. Jacquemont S, Reymond A, Zufferey F, Harewood L, Walters RG, Kutalik Z, *et al*. (2011): Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus. Nature 478:97–102.
31. Chance PF, Abbas N, Lensch MW, Pentao L, Roa BB, Patel PI, *et al*. (1994): Two autosomal dominant neuropathies result from reciprocal DNA duplication/deletion of a region on chromosome 17. Hum Mol Genet 3:223–228.
32. Lupski JR, Wise CA, Kuwano A, Pentao L, Parke JT, Glaze DG, *et al*. (1992): Gene dosage is a mechanism for Charcot-Marie-Tooth disease type 1A. Nat Genet 1:29–33.