Published in final edited form as:
JAMA Psychiatry. 2016 September ; 73(9): 963–969. doi:10.1001/jamapsychiatry.2016.1831.

Analysis of Intellectual Disability Copy Number Variants for Association With Schizophrenia

Elliott Rees, PhD,
Medical Research Council Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, Wales

Kimberley Kendall, MBBCh,
Medical Research Council Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, Wales

Antonio F. Pardiñas, PhD,
Medical Research Council Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, Wales

Sophie E. Legge, PhD,
Medical Research Council Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, Wales

Andrew Pocklington, PhD,
Medical Research Council Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, Wales

Valentina Escott-Price, PhD,
Medical Research Council Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, Wales

James H. MacCabe, PhD, MRCPsych,
Department of Psychosis Studies, Institute of Psychiatry Psychology and Neuroscience, King's College London, London, England

Corresponding Author: George Kirov, PhD, MRCPsych, Medical Research Council Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Hadyn Ellis Building, Maindy Road, Cardiff CF24 4HQ, Wales (kirov@cardiff.ac.uk).

Author Contributions: Dr Rees and Prof Kirov had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.
Concept and design: Rees, MacCabe, Collier, O'Donovan, Owen, Kirov.
Acquisition, analysis, or interpretation of data: Rees, Kendall, Pardiñas, Legge, Pocklington, Escott-Price, Collier, Holmans, O'Donovan, Walters, Kirov.
Drafting of the manuscript: Rees, Kendall, O'Donovan, Owen, Walters, Kirov.
Critical revision of the manuscript for important intellectual content: Rees, Pardiñas, Legge, Pocklington, Escott-Price, MacCabe, Collier, Holmans, O'Donovan, Walters, Kirov.
Statistical analysis: Rees, Escott-Price, Holmans, O'Donovan, Kirov.
Obtaining funding: MacCabe, Collier, O'Donovan, Owen, Walters.
Administrative, technical, or material support: Pardiñas, O'Donovan.
Study supervision: Collier, O'Donovan, Owen, Walters, Kirov.
No additional contributions: Kendall, Legge, Pocklington.

Conflict of Interest Disclosures: Prof Collier is a full-time employee and stockholder of Eli Lilly and Company Ltd. Prof O'Donovan has received a consultancy fee from Roche for participation in a discussion about using genetics to identify drug targets. No other disclosures were reported.

Europe PMC Funders Author Manuscripts

David A. Collier, PhD,
SGDP Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, England; Discovery Neuroscience Research, Eli Lilly and Company Ltd, Lilly Research Laboratories, Surrey, England

Peter Holmans, PhD,
Medical Research Council Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, Wales

Michael C. O'Donovan, PhD, FRCPsych,
Medical Research Council Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, Wales

Michael J. Owen, PhD, FRCPsych,
Medical Research Council Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, Wales; Division of Psychological Medicine and Clinical Neurosciences, Cardiff University, Cardiff, Wales

James T. R. Walters, MRCPsych, PhD[#], and
Medical Research Council Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, Wales

George Kirov, PhD, MRCPsych[#]
Medical Research Council Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, Wales

[#] These authors contributed equally to this work.

Abstract

Importance—At least 11 rare copy number variants (CNVs) have been shown to be major risk factors for schizophrenia (SZ). These CNVs also increase the risk for other neurodevelopmental disorders, such as intellectual disability. It is possible that additional intellectual disability–associated CNVs increase the risk for SZ but have not yet been implicated in SZ because of previous studies being underpowered.

Objective—To examine whether additional CNVs implicated in intellectual disability represent novel SZ risk loci.

Design, Setting, and Participants—We used single-nucleotide polymorphism (SNP) array data to evaluate a set of 51 CNVs implicated in intellectual disability (excluding the known SZ loci) in a large data set of patients with SZ and healthy persons serving as controls recruited in a variety of settings. We analyzed a new sample of 6934 individuals with SZ and 8751 controls and combined those data with previously published large data sets for a total of 20 403 cases of SZ and 26 628 controls.

Main Outcomes and Measures—Burden analysis of CNVs implicated in intellectual disability (excluding known SZ CNVs) for association with SZ. Association of individual intellectual disability CNV loci with SZ.

Results—Of data on the 20 403 cases (6151 [30.15%] female) and 26 628 controls (14 252 [53.52%] female), 51 intellectual disability CNVs were analyzed. Collectively, intellectual disability CNVs were significantly enriched for SZ ($P = 1.0 \times 10^{-6}$; odds ratio [OR], 1.9 [95% CI, 1.46-2.49]). Of the 51 CNVs tested, 19 (37%) were more common in SZ cases; only 4 (8%) were more common in controls (no observations were made for the remaining 28 [55%] loci). One novel locus, deletion at 16p12.1, was significantly associated with SZ after correction for multiple testing (rate in SZ, 33 [0.16%]; rate in controls, 12 [0.05%]; corrected $P = .017$; OR, 3.3; 95% CI, 1.61-7.05), and 2 loci reached nominal levels of significance (deletions at 2q11.2: 6 [0.03%] vs 1 [0.004%]; OR, 9.3; 95% CI, 1.03-447.76; corrected $P > .99$; and duplications at 10q11.21q11.23: 5 [0.2%] vs 0 [0.03%]; OR, infinity; 95% CI, 1.26-infinity; corrected $P = .71$). Our new data set also provided independent support for the 11 SZ risk loci previously reported to be associated with the disorder and for the protective effect of 22q11.2 duplication.

Conclusions and Relevance—A large proportion of CNV loci implicated in intellectual disability are risk factors for SZ, but the available sample size precludes statistical confirmation for additional individual loci.

The risk for developing schizophrenia (SZ) is increased by both rare and common alleles distributed across the genome.1 Although many common alleles are associated with very small increases in risk (odds ratios [ORs], <1.2),2 a few copy number variants (CNVs) are associated with substantial increases in risk, with ORs of 1.5 to higher than 50.3,4 Only 11 specific CNVs have currently been robustly identified as SZ risk factors.4 These SZ-associated CNVs are very rare, being found in 1 in 200 to 1 in several thousand people with SZ, and have required very large sample sizes to confidently implicate them.3,4 The genome-wide burden of CNVs greater than 500 kilobase (kb) has been shown5 to be significantly increased in patients with SZ compared with the burden in controls after excluding known SZ risk loci, suggesting the existence of additional SZ-risk CNVs.

All 11 known SZ-associated CNV risk loci have been implicated in other neurodevelopmental disorders, such as intellectual disability (ID) and autism spectrum disorder,3,6–9 usually with similar or higher ORs than for SZ.10 Studies involving tens of thousands of patients with autism spectrum disorder, ID, and congenital malformations referred to clinical genetics clinics for chromosomal microarray analysis have suggested that more than 90 loci could be enriched for CNVs in these disorders.6–9,11 Most of these CNVs are large, recurrent, and formed through nonallelic homologous recombination between directly orientated, paralogous low-copy repeats.12

We hypothesized that, beyond those already identified as risk factors for SZ, CNVs that are implicated in other neurodevelopmental disorders also increase the risk for SZ but have not been discovered owing to the limited power of existing studies of SZ. To test this hypothesis, we selected 51 CNVs that are significantly associated with ID6 and tested them for association with SZ in a new sample of 6934 patients with SZ and 8751 individuals serving as controls. We added those data to a synthesis of the largest studies on SZ for which we had access to the raw CNV calls, for a total of 20 403 cases and 26 628 controls.

Methods

Samples

The new data set (which we call CLOZUK2) is fully independent of any samples used for earlier CNV studies of SZ and comprises (after quality control) 6934 SZ cases and 8751 controls. In CLOZUK2, 6680 new cases were ascertained on the basis of patients receiving clozapine and having a clinical diagnosis of treatment-resistant SZ. This new sample comes from our ongoing anonymized collection (CLOZUK) and was recruited as part of the European Union Seventh Framework Programme (EU-FP7) study, CRESTAR, in collaboration with Leyden Delta, a company that is contracted in large parts of the United Kingdom to supply clozapine and provide blood monitoring in patients receiving this drug. The CLOZUK samples were collected anonymously across the United Kingdom (thus, without express patient consent), consistent with the UK Human Tissue Act and with the approval of the UK National Research Ethics Committee for use in genetic studies. The remaining 254 new cases were obtained from the Cardiff Cognition in Schizophrenia (CardiffCOGS) study and were recruited from community, in-patient, and voluntary sector mental health services in the United Kingdom. That study was also approved by the UK National Research Ethics Committee, and the participants provided written informed consent. Further details about the CLOZUK and CardiffCOGS samples have been published4 and are available in eAppendix 1 in the Supplement. All case samples were genotyped on HumanOmniExpress-12v1-1_B arrays (Illumina) at DeCode Genetics. For controls, we used data sets from dbGaP that did not involve individuals specifically selected for neurodevelopmental phenotypes (eAppendix 1 and eTable 1 in the Supplement). All control samples were genotyped on Illumina Omni arrays for compatibility with the present data set. These cohorts were from studies on chronic obstructive pulmonary disease (n = 992), melanoma (n = 2416), breast cancer (n = 954), and corneal dystrophy (n = 3529). We also used data from 860 psychiatrically unscreened blood donors recruited by our department.

For our primary analysis of ID CNVs, the new case-control sample was combined with 3 large, previously published data sets4,13,14 for which we had access to the raw CNV calls, allowing us to control for possible bias from differential calling of CNVs in different studies and for possible sex differences in CNV burden.15–17 These data sets included our previous CLOZUK1/CardiffCOGS sample,4 the International Schizophrenia Consortium sample (ISC),13 and the Molecular Genetics of Schizophrenia (MGS) sample.14 In total, we analyzed ID CNVs in 20 403 SZ cases (6151 [30.15%] female) and 26 628 controls (14 252 [53.52%] female).

CNV Analysis

Full details of the CNV calling and quality control are provided in eAppendix 2 in the Supplement. Briefly, raw intensity data were processed using Illumina Genome Studio software, version 2011.1. Log R ratios and B allele frequencies were used to call CNVs using PennCNV software, version 1.0.3.18 We called CNVs in the new sample using 666 868 probes that are common to all Illumina arrays used in either cases or controls. The CNVs were joined if the distance between them was less than 50% of their combined length

and excluded if they were called with fewer than 10 probes, were less than 10 kb in size, overlapped low copy repeats by more than 50% of their length, or had a probe density of less than 1 probe/20 kb. The CNV loci with a frequency greater than 1% in the new sample were excluded using PLINK.19 The CNVs from the remaining data sets (ISC, MGS, and CLOZUK1) were analyzed using the methods reported in Rees et al.5

Selection of ID CNVs

We defined ID CNVs as those that were significantly associated with ID ($P < .05$) in a large study6 that involved patients referred to clinical genetics clinics. We excluded ID CNVs with a population frequency of more than 0.5% given our focus on rare CNV loci. We chose a relaxed cutoff level of $P < .05$ to not exclude loci that are likely to be true if tested in larger samples (eTable 2 in the Supplement). The full list involves 63 CNVs, 12 of which are already strongly associated with SZ (11 risk and 1 protective locus, listed in eTable3 in the Supplement). Given that these 12 CNVs provided the basis for our hypothesis, they were excluded from the primary analysis to allow an independent test of our hypothesis. However, as a reference, we provide a full list of all 63 CNVs (eTable 2 in the Supplement); their rates in patients with SZ as well as those with ID, autism spectrum disorder, or congenital malformations; healthy controls; and the general population. The relative risk (RR) for SZ is also given.

Statistical Analysis

For the primary test of whether, en masse, ID CNVs increase the risk for SZ, we analyzed 20 403 cases and 26 628 controls using a 2-sided Fisher exact test. We used the same sample to test the association of individual ID CNVs with SZ using a Cochran-Mantel-Haenszel exact test, stratified by sex and study. For analysis of individual loci, we adjusted *P* values using Bonferroni correction for the 51 ID CNVs tested in our analysis of novel SZ loci. All statistical analyses were performed using R, version 3.1.2 (R Foundation).

For a separate analysis of known SZ CNVs in the new data set alone (6934 SZ cases and 8751 controls), we evaluated 11 risk loci and the protective 22q11.2 duplication using a 1-tailed Fisher exact test (loci tested listed in eTable 3 in the Supplement). In addition, we individually tested an additional set of 11 loci that received suggestive support for association with SZ in a previous publication5 using a 1-tailed Fisher exact test and performed a combined analysis with the published data5 using a 2-tailed Cochran-Mantel-Haenszel test stratified by study.

Power Calculation

Our power estimates for each CNV locus (Figure and eFigure in the Supplement) are based on an α level of .05, 20 403 cases and 26 628 controls, the RR of CNV carriers to have SZ, the CNV frequencies observed in the present samples, and assuming a dominant model (ie, the RR of heterozygotes = RR of homozygotes). Power calculations were performed using the online Genetic Power Calculator (http://pngu.mgh.harvard.edu/~purcell/gpc/).20 Estimates of RR for SZ were generated by comparing the CNV rates in patients with SZ with those in the general population (instead of those in the controls). We used this comparison because more than half of the CNVs have zero observations in controls, thus

precluding the estimation of RR (division by 0). Further details on these estimates are presented in eAppendix 3 in the Supplement. Clearly, the true frequencies and RRs for each locus will differ, but the purpose of this analysis is to give an approximation of the overall distribution of RRs and frequencies for all loci.

Results

Analysis of Novel SZ CNVs

Table 1 presents the results for potential novel SZ-associated CNVs, limited to the 23 ID CNV loci in which at least 1 CNV was observed in the combined sample (20 403 cases or 26 628 controls). For completeness, the frequencies for all 63 ID CNVs in each data set are presented in eTable 2 in the Supplement. Collectively, CNVs at the 51 ID loci (excluding known SZ loci) were found 141 times in cases (0.69%) and 97 times in controls (0.36%) (2-sided Fisher exact test, $P = 1.0 \times 10^{-6}$; OR, 1.9; 95% CI, 1.46-2.49). Of the 23 informative loci, 19 had higher frequencies in SZ and only 4 in controls.

Several loci reported in Table 1 have been previously implicated4 in SZ but did not meet our predefined criteria for multiple testing correction. We now found 16p12.1 deletions to be significantly associated with SZ after correction for studywide multiple testing (SZ rate, 0.16%; control rate, 0.045%; corrected $P = .017$; OR, 3.3; 95% CI, 1.61-7.05). This association would also survive a more stringent, predefined Bonferroni correction for testing of 120 genomic loci prone to nonallelic homologous recombination8 ($P = .041$, corrected for 120 loci). The excess burden of all ID CNVs in SZ remains significant after excluding the 16p12.1 deletion (50 CNV loci tested, $P = 4.6 \times 10^{-4}$; OR, 1.66; 95% CI, 1.24-2.24), suggesting that additional ID loci drive this collective excess, as reported in Table 1. Two additional loci were nominally significantly associated with SZ risk: 10q11.21q11.23 duplication (SZ rate, 0.025%, control rate, 0%; uncorrected $P = .014$; OR, infinity; 95% CI, 1.26-infinity); 2q11.2 deletion (SZ rate, 0.029%; control rate, 0.004%; uncorrected $P = .037$; OR, 9.3; 95% CI, 1.03-447.76). The CNVs at 7 ID loci were found only in cases and for 2 loci only in controls.

We estimate that our study (20 403 cases and 26 628 controls) has 80% power at an α level of .05 to detect an association with most of the previously identified SZ CNV loci, but few of the remaining 51 ID CNV loci (Figure). Under the simplifying assumption that the observed RR for SZ and the population frequencies in the present study are accurate, we estimate that even a sample of 100 000 cases and 100 000 controls would identify only 5 additional novel SZ CNV associations (Figure) among the set of ID CNVs at nominal significance.

Analysis of Previously Implicated SZ CNVs in the New Independent Data set

In the new independent CLOZUK2 data set (6934 cases, 8751 controls), each of the 11 known SZ risk loci had higher rates in the cases than in the controls (Table 2). For 6 of these loci, the differences were nominally significant at $P < .05$. Collectively, these 11 risk CNVs were observed 147 times in the new 6934 cases (2.1%) and 55 times in the new 8751 controls (0.6%) ($P = 2.1 \times 10^{-16}$; OR, 3.42; 95% CI, 2.49-4.77). We also found lower rates

of the 22q11.2 duplication in the SZ cases compared with controls, supporting work that implicated it as a protective factor.21

Previous work5 reported a different set of 12 CNV loci (mostly nonrecurrent) with suggestive evidence for association with SZ (ie, nominally significant but not after correction for multiple testing). One of these loci was the recurrent 16p12.1 deletion that is now supported in our primary analysis of ID CNV loci (Table 1). Of the remaining 11 previously suggestive loci (all of which are nonrecurrent), 3 were nominally significantly associated ($P < .05$) in our new independent case-control sample (eTable 4 in the Supplement). Our analysis of these 11 nonrecurrent suggestive CNV loci, using our combined sample (new CLOZUK2 data and the data published in the previous study5), increased the strength of association for 5 CNVs, some by several orders of magnitude (eTable 4 in the Supplement). However, none of these nonrecurrent associations withstood correction for multiple testing (herein we applied a more stringent correction for 20 000 genes given the potential for nonrecurrent CNVs to disrupt any gene).

Discussion

Previous work5 reported that 11 rare CNVs are robust risk factors for developing SZ. Because of the low frequencies of these CNVs (some are found in <1 per 1000 patients), very large samples were required to obtain the statistical power needed for their discovery. Copy number variant burden analyses that excluded known SZ loci have provided evidence that additional SZ-associated CNVs exist.5 It has not been possible to confidently identify such CNVs because they are too rare, have smaller effect sizes for the development of SZ, or both.

Given strong evidence for an overlap between CNVs that are known to confer risk for SZ and those that confer risk for neurodevelopmental disorders, including ID,3 we hypothesized that additional CNVs that have shown6 significant evidence for association with ID also increase the risk for SZ. That hypothesis was strongly supported with a collective enrichment in SZ cases for CNVs at 51 loci associated with ID ($P = 1.0 \times 10^{-6}$; OR, 1.9; 95% CI, 1.46-2.49). Given general support for the hypothesis, we tested individual loci within this set of CNVs and obtained significant evidence for 16p12.1 deletion as a novel SZ risk factor. However, even after excluding this locus, cases were still enriched for ID-associated CNVs ($P = 4.6 \times 10^{-4}$; OR, 1.66; 95% CI, 1.24-2.24), indicating the presence of additional risk loci among this set. Overall, 30 of 63 of the CNVs (47.6%) known to be associated with ID disorders (ie, including the known SZ CNVs) have higher frequencies in SZ, and only 5 CNVs have greater frequencies in controls (no observations were made for the remaining 28 loci).

Our power analysis (Figure) demonstrates that there may be numerous additional CNVs that have very high RRs for SZ but cannot be identified with the available sample sizes owing to their rarity. Data in the Figure assume accurate estimates of RR for SZ and CNV population frequencies; thus, each observation has very wide 95% CIs, but the power analysis gives a good representation of the overall distribution. Therefore, it appears that the known SZ CNV associations represent the low-hanging fruit, with their frequencies and RRs for SZ placing

them to the right side of the power curve (solid line in the Figure), allowing them to be identified in sample sizes typical for research conducted in this area. We conclude that many more ID CNVs with an RR for SZ greater than 1 (Figure and Table 1) are likely to be risk factors for SZ, but much larger samples are required for robust associations to be made at these loci. In fact, even 100 000 cases and 100 000 controls would not implicate many loci, even if their RRs are as high as observed in the present analysis (dashed line in the Figure).

We report that, on its own, this new sample supported 11 CNVs previously shown4 to be highly significantly associated with SZ in that CNV rates were greater in cases compared with controls, with 6 of the CNV loci reaching a nominal level of significance: deletions at 1q21.1, 15q11.2, and 22q11.2, and duplications at the Prader-Willi syndrome/Angelman syndrome critical region, 16p13.11, and 16p11.2 (Table 2). Duplications of 22q11.2, a protective factor, were found at a higher rate in controls, as expected (0.05% vs 0.01%; $P = .27$, 1-tailed Fisher exact test; OR, 0.32; 95% CI, 0.006-3.19). Most of these risk loci and the protective locus have recently been supported in a large Chinese SZ sample,22 providing further evidence that the associations are robust. Finally, we used the new data to evaluate 11 nonrecurrent loci identified as potential novel SZ risk factors in a previous genome-wide CNV study.5 In the present study, we found support for 5 of these loci: deletions of *IRGM/ZNF300/SMIM3* and *SLC1A1*, and duplications of *FAM149A/CYP4V2/FLJ38576*, *PHACTR2*, and *GALR1* (according to UCSC Genome Browser [GRCh37/hg19 Assembly]). Nonrecurrent CNV associations require more-stringent correction for multiple testing, since they can potentially affect any genomic region. We4 have suggested that such CNV associations should be corrected for multiple testing of 20 000 genes ($P < 2.5 \times 10^{-6}$). Despite support in our new data for several of these previously identified risk CNVs, the significance of their association still falls short of that threshold. We are not aware of any CNVs that convincingly increase the risk for SZ and not for ID, but expect that such CNVs exist and will be found after larger data sets are analyzed.

We acknowledge limitations in our study, especially the potential problems of analyzing cases and controls genotyped in different laboratories and on different arrays. We have tried to minimize these problems by using only overlapping probes (in the CLOZUK samples), independently producing log R ratios and B-allele frequencies in each data set separately to avoid batch effects, correcting results for study in Cochran-Mantel-Haenszel tests, and checking for differential missingness of all CNV loci, including ensuring adequate probe coverage of each CNV in each data set (further data available in eAppendix 2 in the Supplement).

Conclusions

We provide evidence that a large proportion of the ID loci are likely to be risk factors for SZ. Significant association was noted between 16p12.1 deletions and SZ after correcting for multiple testing; in addition, the results indicate that larger samples will identify additional ID CNVs that are true SZ risk factors. These findings strengthen the evidence for an etiologic overlap between neurodevelopmental disorders. Finding a known ID-associated CNV in a patient with SZ should raise the suspicion that it is relevant for the psychiatric disorder in that individual.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding/Support: This project has received funding from the European Union's Seventh Framework Programme for research, technological development, and demonstration under grant agreement 279227. The work at Cardiff University was funded by Medical Research Council (MRC) Centre grant MR/L010305/1 and program grants G0800509 and the European Community's Seventh Framework Programme HEALTH-F2-2010-241909 (Project EU-GEI). Funding was also provided by the MRC and Wellcome Trust, UK (Drs O'Donovan, Owen, and Kirov) and was supported by a clinical research fellowship from the MRC/Welsh Assembly Government and the Margaret Temple Award from the British Medical Association (Dr Walters).

Role of the Funder/Sponsor: The funding organizations had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Additional Information: Data were contributed by the Database of Genotypes and Phenotypes (dbGaP). *Genetic Epidemiology of COPD* (dbGaP study accession, phs000179.v3.p2). Funded by the National Heart, Lung, and Blood Institute (NHLBI). Principal investigators were James D. Crapo (National Jewish Health) and Edwin K. Silverman (Brigham and Women's Hospital). The study was run at the NHLBI and funded by the National Institutes of Health (NIH) (U01HL089897 and U01HL089856). *A Genome-Wide Association Study of Fuchs' Endothelial Corneal Dystrophy* (dbGaP study accession, phs000421.v1.p). Principal investigators were Natalie Afshari (Duke University), John Gottsch (The Johns Hopkins University), Sudha K. Iyengar (Case Western Reserve University), Nicholas Katsanis (The Johns Hopkins University), Gordon Klintworth (Duke University), and Jonathan Lass (Case Western Reserve University). Co-investigators were Simon Gregory (Duke University) and Yi-Ju Li (Duke University). The study was funded by the National Eye Institute, NIH (R01EY016482 [Dr Iyengar], R01EY016514 [Dr Klintworth], and R01EY016835 [Dr Gottsch]). Genotyping was carried out at the Center for Inherited Disease Research (CIDR), The Johns Hopkins University, and was funded by NIH (contracts HHSN268200782096C [high throughput genotyping for studying the genetic contributions to human disease] and HHSN268201100011I [high throughput genotyping for studying the genetic contributions to human disease]). *California Pacific Medical Center Research Breast Health Cohort* (dbGaP study accession, phs000395.v1.p1). Principal investigator was Elad Ziv (University of California, San Francisco). Coinvestigators were Steven Cummings (California Pacific Medical Center Research Institute and University of California, San Francisco), Karla Kerlikowske, and John Shepherd (University of California, San Francisco). The study was run at the National Cancer Institute, NIH, and was funded by the NIH (P01 CA107584 and R01 CA120120). Genotyping was carried out at The Johns Hopkins University CIDR and was funded by the NIH (contracts HHSN268200782096C [high throughput genotyping for studying the genetic contributions to human disease] and HHSN268201100011I [high throughput genotyping for studying the genetic contributions to human disease]). *Study of Melanoma Risk in Australia and the United Kingdom* (dbGaP study accession: phs000519.v1.p1). Principal investigator was Nicholas Hayward (Queensland Institute of Medical Research). The study was funded by the National Cancer Institute of the NIH (R01CA088363). Genotyping was carried out at the CIDR and The Johns Hopkins University and funded by the NIH (HHSN268201100011I). *Genome-Wide Association of Schizophrenia Study*; (GAIN) (dbGaP accession phs000021.v3.p2). Funding was provided by the National Institute of Mental Health of the NIH (R01 MH67257, R01 MH59588, R01 MH59571, R01 MH59565, R01 MH59587, R01 MH60870, R01 MH59566, R01 MH59586, R01 MH61675, R01 MH60879, R01 MH81800, U01 MH46276, U01 MH46289, U01 MH46318, U01 MH79469, and U01 MH79470), and the genotyping of samples was provided through GAIN. Principal investigator was Pablo V. Gejman (Evanston Northwestern Healthcare and Northwestern University). *Genome-Wide Association of Schizophrenia Study* (*MGS_nonGAIN*) (dbGaP accession phs000167.v1.p1). Samples and associated phenotype data for the MGS nonGAIN study were collected from NIMH Schizophrenia Genetics Initiative grants (U01s: MH46276 [C. R. Cloninger], MH46289 [C. Kaufmann], and MH46318 [M. T. Tsuang]; and MGS Part 1 and Part 2 R01s: MH67257 [N. G. Buccola], MH59588 [B. J. Mowry], MH59571 [P. V. Gejman], MH59565 [Robert Freedman], MH59587 [F. Amin], MH60870 [W. F. Byerley], MH59566 [D. W. Black], MH59586 [J. M. Silverman], MH61675 [D. F. Levinson], and MH60879 [C. R. Cloninger]). *Genetic Architecture of Smoking and Smoking Cessation* (dbGaP study accession, phs000404.v1.p1). Funding support for genotyping, which was performed at the CIDR, was provided by the NIH (1 X01 HG005274-01). CIDR is fully funded through a federal contract from the NIH to The Johns Hopkins University (HHSN268200782096C). Assistance with genotype cleaning, as well as with general study coordination, was provided by the Gene Environment Association Studies Coordinating Center (U01 HG004446). Funding support for collection of data sets and samples was provided by the Collaborative Genetic Study of Nicotine Dependence (P01 CA089392) and the University of Wisconsin Transdisciplinary Tobacco Use Research Center (P50 DA019706 and P50 CA084724). *High Density SNP Association Analysis of Melanoma: Case-Control and Outcomes Investigation* (dbGaP study accession, phs000187.v1.p1). Research support to collect data and develop an application to support this project was provided by the NIH (3P50CA093459, 5P50CA097007,

Europe PMC Funders Author Manuscripts

5R01ES011740, and 5R01CA133996). *Genetic Epidemiology of Refractive Error in the KORA (Kooperative Gesundheitsforschung in der Region Augsburg) Study* (dbGaP study accession, phs000303.v1.p1). Principal investigators were Dwight Stambolian (University of Pennsylvania, Philadelphia) and H. Erich Wichmann (Institut für Humangenetik and National Eye Institute, NIH). Funded was provided by the NIH (R01 EY020483). Further data are available in eAppendix 1 in the Supplement.

Additional Contributions: We thank the participants and clinicians who took part in the CardiffCOGS study. Sophie Bishop, BSc, and Amy Lynham, BSc (Cardiff University), participated in recruitment, interviewing, and rating of participants. For the CLOZUK2 sample, Leyden Delta, particularly Marinka Helthius, MD, John Jansen, PhD, and Karel Jollie, PharmD, participated in the sample collection, anonymization, and data preparation. Andy Walker, BSc (Magna Laboratories), and CLOZUK1, Novartis Switzerland, and The Doctor's Laboratory staff provided guidance and cooperation. Kiran Mantripragada, PhD, Lesley Bates, BSc, Catherine Bresner, BSc, and Lucinda Hopkins, BSc (Cardiff University), provided laboratory sample management. DeCode Genetics performed the genotyping on the CLOZUK2/CardiffCOGS samples. None of these individuals were financially compensated. International Schizophrenia Consortium provided us with raw CNV calls from that study.

References

1. Sullivan PF, Daly MJ, O'Donovan M. Genetic architectures of psychiatric disorders: the emerging picture and its implications. Nat Rev Genet. 2012; 13(8):537–551. [PubMed: 22777127]
2. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. Nature. 2014; 511(7510):421–427. [PubMed: 25056061]
3. Malhotra D, Sebat J. CNVs: harbingers of a rare variant revolution in psychiatric genetics. Cell. 2012; 148(6):1223–1241. [PubMed: 22424231]
4. Rees E, Walters JTR, Georgieva L, et al. Analysis of copy number variations at 15 schizophrenia-associated loci. Br J Psychiatry. 2014; 204(2):108–114. [PubMed: 24311552]
5. Rees E, Walters JTR, Chambert KD, et al. Wellcome Trust Case Control Consortium. CNV analysis in a large schizophrenia sample implicates deletions at 16p12.1 and SLC1A1 and duplications at 1p36.33 and CGNL1. Hum Mol Genet. 2014; 23(6):1669–1676. [PubMed: 24163246]
6. Coe BP, Witherspoon K, Rosenfeld JA, et al. Refining analyses of copy number variation identifies specific genes associated with developmental delay. Nat Genet. 2014; 46(10):1063–1071. [PubMed: 25217958]
7. Dittwald P, Gambin T, Szafranski P, et al. NAHR-mediated copy-number variants in a clinical population: mechanistic insights into both genomic disorders and Mendelizing traits. Genome Res. 2013; 23(9):1395–1409. [PubMed: 23657883]
8. Girirajan S, Rosenfeld JA, Coe BP, et al. Phenotypic heterogeneity of genomic disorders and rare copy-number variants. N Engl J Med. 2012; 367(14):1321–1331. [PubMed: 22970919]
9. Kaminsky EB, Kaul V, Paschall J, et al. An evidence-based approach to establish the functional and clinical significance of copy number variants in intellectual and developmental disabilities. Genet Med. 2011; 13(9):777–784. [PubMed: 21844811]
10. Kirov G, Rees E, Walters JT, et al. The penetrance of copy number variations for schizophrenia and developmental delay. Biol Psychiatry. 2014; 75(5):378–385. [PubMed: 23992924]
11. Cooper GM, Coe BP, Girirajan S, et al. A copy number variation morbidity map of developmental delay. Nat Genet. 2011; 43(9):838–846. [PubMed: 21841781]
12. Lupski JR. Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. Trends Genet. 1998; 14(10):417–422. [PubMed: 9820031]
13. International Schizophrenia Consortium. Rare chromosomal deletions and duplications increase risk of schizophrenia. Nature. 2008; 455(7210):237–241. [PubMed: 18668038]
14. Levinson DF, Duan J, Oh S, et al. Copy number variants in schizophrenia: confirmation of five previous findings and new evidence for 3q29 microdeletions and VIPR2 duplications. Am J Psychiatry. 2011; 168(3):302–316. [PubMed: 21285140]
15. Desachy G, Croen LA, Torres AR, et al. Increased female autosomal burden of rare copy number variants in human populations and in autism families. Mol Psychiatry. 2015; 20(2):170–175. [PubMed: 25582617]

16. Jacquemont S, Coe BP, Hersch M, et al. A higher mutational burden in females supports a “female protective model” in neurodevelopmental disorders. Am J Hum Genet. 2014; 94(3):415–425. [PubMed: 24581740]
17. Han J, Walters JTR, Kirov G, et al. Gender differences in CNV burden do not confound schizophrenia CNV associations. Sci Rep. 2016; 6:25986. [PubMed: 27185616]
18. Wang K, Li M, Hadley D, et al. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res. 2007; 17(11):1665–1674. [PubMed: 17921354]
19. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81(3):559–575. [PubMed: 17701901]
20. Purcell S, Cherny SS, Sham PC. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. Bioinformatics. 2003; 19(1):149–150. [PubMed: 12499305]
21. Rees E, Kirov G, Sanders A, et al. Wellcome Trust Case Control Consortium. Evidence that duplications of 22q11.2 protect against schizophrenia. Mol Psychiatry. 2014; 19(1):37–40. [PubMed: 24217254]
22. Li Z, Chen J, Xu Y, et al. Genome-wide analysis of the role of copy number variation in schizophrenia risk in Chinese. Biol Psychiatry. 2015 S0006-3223(15)00988-9.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Key Points

Question Given that all known schizophrenia copy number variant (CNV) loci are also intellectual disability risk factors, are there additional schizophrenia loci among the remaining known intellectual disability CNVs?

Findings In this analysis of single-nucleotide polymorphism array data on 20 403 individuals with schizophrenia, after excluding known schizophrenia CNVs, intellectual disability loci were en masse significantly enriched in patients with schizophrenia compared with controls. For specific loci, deletions at 16p12.1 were significantly associated with schizophrenia after correcting for multiple testing.

Meaning Many intellectual disability CNVs are likely to represent novel schizophrenia risk loci, but larger samples are required for their identification.

Europe PMC Funders Author Manuscripts

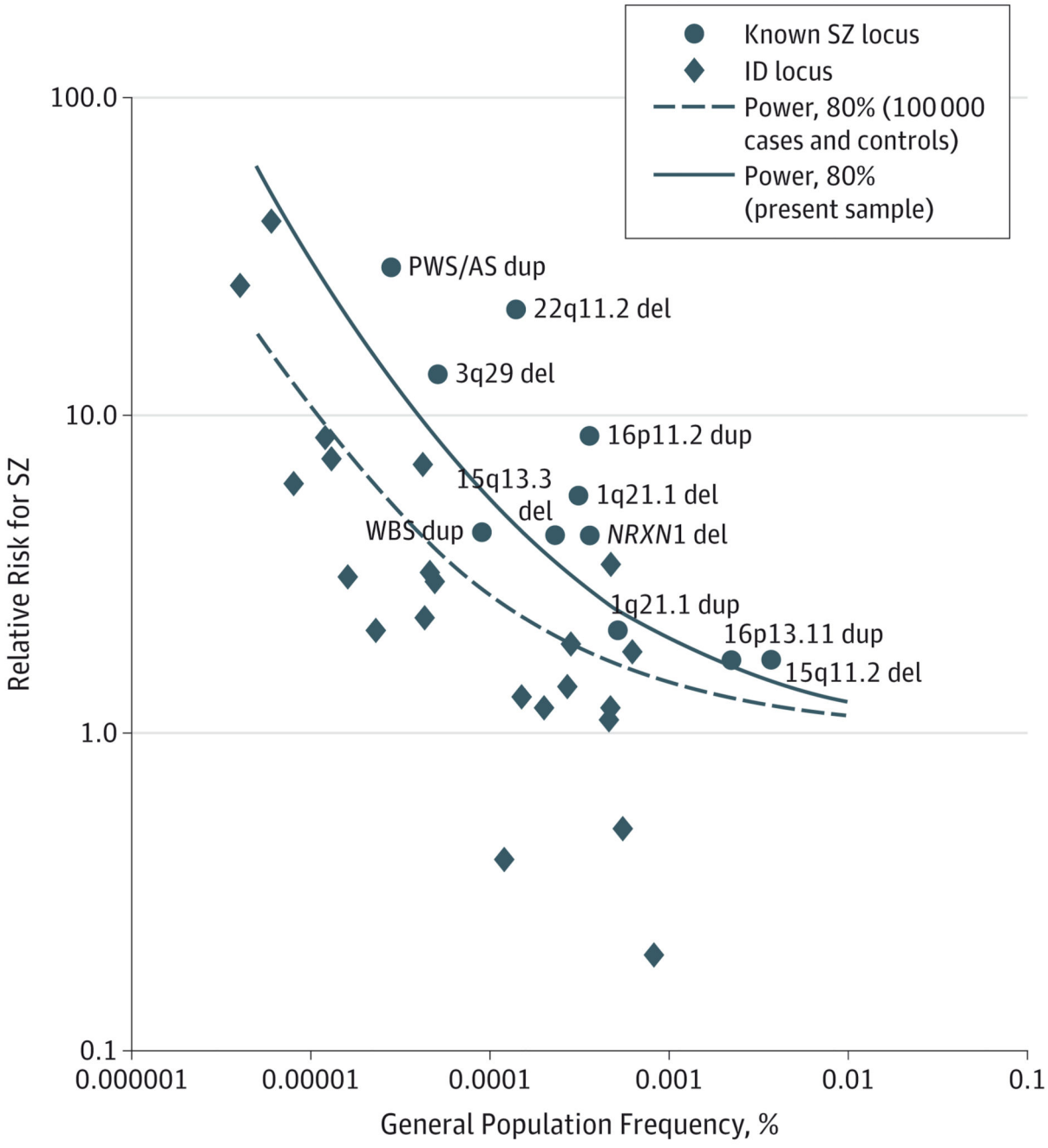

Figure. Analysis of Intellectual Disability Copy Number Variants (CNVs) Power Calculation

Power calculations for CNV loci under a dominant model are based on 20 403 cases and 26 628 controls, an α level of .05, and the relative risks for schizophrenia (SZ) and general population frequencies reported in eTable 2 in the Supplement. The solid line indicates 80% power given a sample size of 20 403 cases and 26 628 controls with an α level of .05. The dashed line indicates 80% power given a sample size of 100 000 cases and 100 000 controls

with an α level of .05. We excluded loci that were not observed in a patient with SZ or a control. del indicates deletion; dup, duplication.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

JAMA Psychiatry. Author manuscript; available in PMC 2017 September 01.

Table 1

Association of ID CNV Loci With Schizophrenia[a]

		CNVs		%						P Value	
CNV	Critical/Unique Sequence Region, hg19, Mb	Controls	SZ	Controls Rate	SZ Rate	Rate in ID/ASD	Rate in General Population	Relative Risk SZ	CMH OR (95% CI)	CMH	Bonferroni Corrected[b]
16p12.1 del	chr16:21.95-22.43	12	33	0.05	0.16	0.124	0.047	3.4	3.3 (1.61-7.05)	.00034	.017
10q11.21q11.23 dup	chr10:49.39-51.06	0	5	0	0.025	0.024	6.0×10^{-4}	40.7	Inf (1.26-Inf)	.014	.71
2q11.2 del	chr2:96.74-97.68	1	6	0.004	0.029	0.018	0.0042	7	9.3 (1.03-447.76)	.037	>.99
TAR dup	chr1:145.39-145.81	16	23	0.06	0.113	0.155	0.062	1.8	1.9 (0.93-3.93)	.066	>.99
Potocki-Lupski syndrome dup	chr17:16.81-20.21	0	2	0	0.01	0.055	0.0012	8.5	Inf (0.5-Inf)	.074	>.99
17q12 del	chr17:34.82-36.22	0	2	0	0.01	0.065	0.0013	7.3	Inf (0.42-Inf)	.095	>.99
17q12 dup	chr17:34.82-36.22	7	11	0.026	0.054	0.089	0.028	1.9	2.2 (0.74-6.76)	.14	>.99
2q37 (*HDAC4*) del	chr2:239.72-243.2	0	1	0	0.005	0.113	0.0023	2.1	Inf (0.067-Inf)	.28	>.99
3q28-29 (*FGF12*) del	chr3:191.86-192.13	1	3	0.004	0.015	0.045	0.0046	3.2	4.1 (0.3-231.36)	.31	>.99
2q13 del	chr2:111.39-112.01	1	3	0.004	0.015	0.057	0.0049	3	3.6 (0.26-206.03)	.34	>.99
2q13 dup	chr2:111.39-112.01	4	4	0.015	0.02	0.022	0.015	1.3	1.7 (0.3-9.98)	.46	>.99
16p11.2 del	chr16:29.65-30.2	13	6	0.049	0.029	0.347	0.055	0.5	0.61 (0.19-1.79)	.47	>.99
16p11.2 distal del	chr16:28.82-29.05	5	5	0.019	0.025	0.094	0.02	1.2	1.7 (0.37-7.6)	.51	>.99
8p23.1 dup	chr8:8.1-11.87	0	2	0	0.01	0.017	0.00038	25.7	Inf (0.13-Inf)	.52	>.99
3p11.2 del	chr3:87.27-87.53	1	2	0.004	0.01	0.031	0.0043	2.3	2.4 (0.11-163.95)	.60	>.99
2p15-16.1 proximal dup	chr2:61.25-61.41	3	1	0.011	0.005	0.031	0.012	0.4	0.37 (0.0066-5.02)	.61	>.99
TAR del	chr1:145.39-145.81	7	8	0.026	0.039	0.076	0.027	1.4	1.2 (0.36-3.9)	.80	>.99
16p11.2 distal dup	chr16:28.82-29.05	12	10	0.045	0.049	0.083	0.046	1.1	1.2 (0.44-3.08)	.82	>.99
8p23.1 del	chr8:8.1-11.87	0	1	0	0.005	0.039	8.0×10^{-4}	6.1	Inf (0.017-Inf)	>.99	>.99
12p13 dup	chr12:6.47-6.83	0	1	0	0.005	0.079	0.0016	3.1	Inf (0.019-Inf)	>.99	>.99
16p13.11 del	chr16:15.51-16.29	12	12	0.045	0.059	0.142	0.047	1.2	1.1 (0.43-2.73)	>.99	>.99
17q11.2 (NF1) del	chr17:29.11-30.27	1	0	0.004	0	0.039	0.0044	0	0 (0-46.05)	>.99	>.99
Distal 22q11.2 del	chr22:21.92-23.65	1	0	0.004	0	0.041	0.0045	0	0 (0-102.32)	>.99	>.99

Abbreviations: ASD, autism spectrum disorder;

CMH, Cochran-Mantel-Haenszel; CNV, copy number variant; del, deletion; dup, duplication; ID, intellectual disability; Inf, infinity; OR, odds ratio; SZ, schizophrenia; TAR, thrombocytopenia-absent radius syndrome.

[a]Results are shown for 23 of the 51 ID CNVs that had at least 1 CNV observed in 20 403 schizophrenia cases or 26 628 controls.

[b] *P* values are Bonferroni corrected for 51 multiple comparisons.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Table 2

CNVs in 11 Schizophrenia Risk Loci and the 22q11.2 Duplication Protective Locus in the New Data Set

CNV	Region (hg19)	CNVs, No. (%)		OR (95% CI)	*P* Value[b]
		Controls (n = 8751)	Cases[a] (n = 6934)		
1q21.1 del	chr1:146,527,987-147,394,444	1 (0.011)	11 (0.16)	13.90 (2.02-596.86)	.0009
1q21.1 dup	chr1:146,527,987-147,394,444	4 (0.046)	6 (0.087)	1.89 (0.45-9.13)	.24
NRXN1 exonic del	chr2:50,145,643-51,259,674	4 (0.046)	5 (0.072)	1.58 (0.34-7.96)	.36
3q29 del	chr3:195,720,167-197,354,826	1 (0.011)	4 (0.058)	5.05 (0.50-248.45)	.12
WBS dup	chr7:72,744,915-74,142,892	0	1 (0.014)	Inf (0.032-Inf)	.44
15q11.2 del	chr15:22,805,313-23,094,530	24 (0.27)	45 (0.65)	2.38 (1.42-4.08)	.00034
PWS/AS dup	chr15:22,805,313-28,390,339	0	6 (0.087)	Inf (1.49-Inf)	.0075
15q13.3 del	chr15:31,080,645-32,462,776	1 (0.011)	3 (0.043)	3.79 (0.30-198.64)	.23
16p13.11 dup	chr16:15,511,655-16,293,689	17 (0.19)	32 (0.46)	2.38 (1.28-4.58)	.0023
16p11.2 dup	chr16:29,650,840-30,200,773	3 (0.034)	20 (0.29)	8.43 (2.50-44.37)	2.8×10^{-5}
22q11.2 del	chr22:19,037,332-21,466,726	0	14 (0.20)	Inf (4.19-Inf)	1.1×10^{-5}
22q11.2 dup	chr22:19,037,332-21,466,726	4 (0.05)	1 (0.01)	0.32 (0.0064-3.19)	.27

Abbreviations: CNV, copy number variant; del, deletion; dup, duplication; Inf, infinity; OR, odds ratio; PWS/AS, Prader-Willi syndrome/Angelman syndrome; WBS, Williams-Beuren syndrome.

[a] All results are in the expected direction of higher/lower rates in cases.

[b] Fisher exact test (1-tailed).