Identifying schizophrenia patients who carry pathogenic genetic copy number variants using standard clinical assessment: retrospective cohort study

Claire Foley, Elizabeth A. Heron, Denise Harold, James Walters, Michael Owen, Michael O'Donovan, Jonathan Sebat, Eric Kelleher, Christina Mooney, Amy Durand, Carlos Pinto, Paul Cormican, Derek Morris, Gary Donohoe, Michael Gill, Louise Gallagher and Aiden Corvin

Background

Copy number variants (CNVs) play a significant role in disease pathogenesis in a small subset of individuals with schizophrenia (~2.5%). Chromosomal microarray testing is a first-tier genetic test for many neurodevelopmental disorders. Similar testing could be useful in schizophrenia.

Aims

To determine whether clinically identifiable phenotypic features could be used to successfully model schizophrenia-associated (SCZ-associated) CNV carrier status in a large schizophrenia cohort.

Method

Logistic regression and receiver operating characteristic (ROC) curves tested the accuracy of readily identifiable phenotypic features in modelling SCZ-associated CNV status in a discovery data-set of 1215 individuals with psychosis. A replication analysis was undertaken in a second psychosis data-set (n = 479).

Results

In the discovery cohort, specific learning disorder (OR = 8.12; 95% CI 1.16–34.88, P = 0.012), developmental delay (OR = 5.19; 95% CI 1.58–14.76, P = 0.003) and comorbid neurodevelopmental disorder (OR = 5.87; 95% CI 1.28–19.69, P = 0.009) were significant independent variables in modelling positive carrier status for a SCZ-associated CNV, with an area under the ROC (AUROC) of 74.2% (95% CI 61.9–86.4%). A model constructed from the discovery cohort including developmental delay and comorbid neurodevelopmental disorder variables resulted in an AUROC of 83% (95% CI 52.0–100.0%) for the replication cohort.

Conclusions

These findings suggest that careful clinical history taking to document specific neurodevelopmental features may be informative in screening for individuals with schizophrenia who are at higher risk of carrying known SCZ-associated CNVs. Identification of genomic disorders in these individuals is likely to have clinical benefits similar to those demonstrated for other neurodevelopmental disorders.

Declaration of interest

None.

Keywords

Genetics; schizophrenia; developmental disorders; autistic spectrum disorders; intellectual disability.

Copyright and usage

© The Authors 2020.

Schizophrenia is a clinically heterogeneous syndrome with substantial heritability. Common small genetic risk factors (polygenic risk) collectively account for some 30% of heritability.[1] Copy number variants (CNVs) (DNA segments of >1 kilobase, present at higher (duplication) or lower (deletion) number than in a reference genome) have a larger role to play in a small subset of cases. There is evidence of genome-wide significant association for at least eight CNV loci in schizophrenia.[2] Individually these events are of moderate penetrance for schizophrenia, with reported odds ratios (ORs) of 2–30, which indicates that they are likely to have a substantial role in disease aetiology, at least for a small group of patients.[3] Almost all of these CNVs are pleiotropic; for carriers of any one of the schizophrenia-associated (SCZ-associated) CNVs, the risk of developing *any* early developmental disorder (e.g. intellectual disability, autism spectrum disorder, developmental delay) is significantly higher than the risk of developing schizophrenia itself.[3] Even in the absence of a psychiatric diagnosis, carriers of CNVs associated with schizophrenia have significant but variable cognitive deficits.[4] These CNVs are therefore potentially pathogenic and clinically significant, but outcomes range from subtle cognitive effects to severe neurodevelopmental disorders. Despite significant progress in our understanding of the genetics of schizophrenia, the process of translating SCZ genetic discovery into clinical impact is in its infancy.

Chromosomal microarray (CMA) testing is recommended as a first-tier genetic test in autism, developmental delay and intellectual disability.[5] Because a smaller proportion of people with schizophrenia (~2.5%) carry known pathogenic CNVs, routine testing is not currently recommended. However, a genetic diagnosis may be empowering for patients and their families, it can inform screening for relevant medical comorbidities and help in reproductive planning.[6] The identification of clinical symptoms or demographic features that differentiate people with schizophrenia who carry SCZ-associated CNVs may be helpful in clarifying who might benefit most from testing. On the basis of the known overlap with other neurodevelopmental disorders and previously reported phenotype studies[3,7–15] we hypothesised that individuals with schizophrenia who carry SCZ-associated CNVs are more likely to have phenotypic features suggestive of pre-existing neurodevelopmental compromise, earlier onset of psychotic symptoms or a positive family history of neurodevelopmental disorder. The objective of this work was to determine whether clinically identifiable phenotypic features could be used to model SCZ-associated CNV carrier status in a large schizophrenia cohort.

Method

Selection of phenotypic variables

A literature review was conducted in PubMed from January 2008 to February 2016 to identify clinical and phenotypic features reported

to be associated with copy number variation in schizophrenia using the search terms 'schizophrenia', 'copy number variant' and 'phenotype'. Publications that specifically described CNV-associated clinical and phenotypic features in schizophrenia were selected to identify neurodevelopmental phenotypic categories. Identified phenotypic domains included early onset of psychosis; premorbid cognitive difficulties; delays in developmental milestones; family history of neurodevelopmental disorder; and syndromal characteristics (dysmorphic features, congenital malformations). Eight specific features falling within these domains were identified through expert clinical consensus that are readily identifiable in a standard clinical evaluation and therefore ultimately of clinical utility and acceptability. Subsequently, 'dysmorphic features' and 'congenital anomalies' were excluded because reliable identification of these features requires additional training or clinical tools.[16] The phenotypic variables selected for analysis are outlined in Table 1.

Clinical sample

The discovery data-set

The discovery data-set consisted of 1215 individuals of Irish ancestry for whom both clinical phenotype and genome-wide SNP array data were available.[17] The individuals were all over 18 years of age and had a diagnosis of schizophrenia or schizoaffective disorder after a structured clinical assessment (as described by First et al[18]). Written informed consent was obtained from all participants. Diagnosis was made on the basis of the consensus lifetime best estimate method using all available information (interview, family or staff report, chart review) with DSM-IV criteria as per the Structured Clinical Interview for DSM-IV-TR Axis I Disorders, research version, patient edition (SCID-I/P). Each referral centre obtained local research ethics committee (REC) approval. There was a preponderance of males in this sample (64%).

Phenotypic data were collected retrospectively from an existing research cohort.[17] The phenotypic data were collected from the SCID-I/P and consisted of interview self-reports. The definitions applied to identify a positive history of the phenotypic variables are outlined in Table 1. Phenotypic data were coded as categorical variables (missing information is described in supplementary Table 1, available at https://doi.org/10.1192/bjp.2019.262).

The replication data-set

The replication data-set was obtained from 19 879 schizophrenia cases published by the Schizophrenia Working Group of the Psychiatric Genomics Consortium (PGC) cohorts (representing 40 cohorts excluding data on Irish individuals).[2] Contributors of the constituent data-sets were approached to request access to additional phenotypic data to replicate the discovery findings. Only one cohort (the Cardiff data-set) was identified with the requisite phenotype data and adequate sample size for replication (many of the well-phenotyped cohorts were small and consequently had no CNV carriers).

The Cardiff data-set (n = 479) consisted of participants from the previously reported Cardiff Cognition in Schizophrenia (CardiffCOGS) study.[19,20] In brief, the sample was recruited with REC approval from community, in-patient and voluntary-sector mental health services in the UK. Written informed consent was obtained from all participants. Participants had a clinical diagnosis of schizophrenia and were interviewed using the Schedules for Clinical Assessment in Neuropsychiatry (SCAN) and case-note review to derive a best-estimate lifetime diagnosis according to DSM-IV criteria. Similar to the discovery set, there was a preponderance of males in the sample (61.2%). The comparable phenotype variables investigated in the Cardiff data-set were: (a) 'history of developmental delay', which was directly comparable to the Irish data-set variable and was defined as 'clinically relevant delays in speech, walking, coordination or diagnosed developmental problem' and (b) a positive history of epilepsy, intellectual disability and/or autism spectrum disorder, which was included as 'comorbid neurodevelopmental diagnosis'. Intellectual disability referred to an IQ <70 and clinical specialist service involvement. The autism spectrum disorder and epilepsy variables were interview self-report of a clinical diagnosis. Missing information is described in supplementary Table 2. The other phenotypic variables selected for in the initial analysis were not collected in this data-set.

Ethical approval

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. All procedures involving human participants were approved by the relevant local research ethics committees in Ireland and the UK, as outlined above.

CNV list

The target CNVs used in the analysis were fifteen CNVs with the strongest evidence of association with schizophrenia (supplementary Table 3) analysed by Rees et al.[20] Twelve of these were also identified in the large PGC CNV meta-analysis[21] and the other three were exon-disrupting deletions at the NRXN1 gene, deletion at distal 16p11.2 and duplications at the Williams–Beuren region identified on the basis of expert consensus or evidence published after the meta-analysis.[20]

Genotyping and CNV calling

The Irish sample was genotyped on the Affymetrix 6.0 array (n = 802) or the Illumina HumanCoreExome chip (n = 413) (full details are available in the literature[17]). The Cardiff sample was

Table 1 Univariate analyses assessing associations[a] between selected phenotypic variables and schizophrenia-associated copy number variant status

Phenotypic variable	Definition	OR (95% CI)	P
Early onset of symptoms[8,9]	Onset of symptoms at <18 years of age	1.55 (0.58–4.27)	0.41
History of learning difficulties[4,11]	Any difficulties with learning reported in school (excluding behavioural difficulties)	3.99 (1.55–10.30)	**0.005**
Specific learning disorder[4,11]	Identified as report of diagnosed dyslexia, dyscalculia or dysgraphia	9.03 (1.38–39.98)	**0.03**
Remedial school support[4,11]	Reported learning support at school, within class, separate classes or special school	1.76 (0.08–10.66)	0.45
Low educational attainment[4,11]	Attained primary school education only	1.84 (0.55–5.74)	0.29
History of developmental delay[3,13]	Delayed milestones: motor, speech, toilet training	5.76 (1.91–16.44)	**0.005**
Comorbid neurodevelopmental diagnosis[3,13]	Diagnosis of ASD, intellectual disability or epilepsy	4.93 (1.18–16.73)	**0.034**
Family history of neurodevelopmental disorder[14,15]	Reported diagnosis of schizophrenia, ASD, intellectual disability or epilepsy in a first- or second-degree family member	1.06 (0.35–3.28)	1

ASD, autism spectrum disorder.
a. Fisher's exact tests were used to assess associations. Significant results are in bold.

Table 2 Multiple logistic regression model to determine whether clinically identifiable phenotypic features could be used to model schizophrenia-associated copy number variant carrier status[a]

Variable	β	s.e.	Wald statistic, Z	P	OR (95% CI)
History of developmental delay	1.65	0.56	2.95	0.003	5.19 (1.58–14.76)
Comorbid neurodevelopmental disorder	1.77	0.67	2.63	0.009	5.87 (1.28–19.69)
Specific learning disorder	2.09	0.83	2.53	0.012	8.12 (1.16–34.88)

a. Predictor coefficients were tested using Wald tests and confidence intervals were obtained using the Wald method. Nagelkerke pseudo $R^2 = 0.196$.

Table 3 Receiver operating characteristic (ROC) curve results[a] for modelling schizophrenia-associated copy number variant status in the discovery (Irish) data-set

Cut-off value	Sensitivity, % (95% CI)	Specificity, % (95% CI)	AUC, % (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)
0.04	58.8 (32.9–81.6)	89.1 (87.1–90.9)	74.2 (61.9–86.4)	7.5 (6.3–20.1)	99.3 (98.0–99.4)

AUC, area under the curve; PPV, positive predictive value; NPV, negative predictive value.
a. Optimal cut-off value, sensitivity, specificity, AUC and predictive values were calculated using three independent variables ('history of developmental delay', 'comorbid neurodevelopmental disorder', 'specific learning disorder').

genotyped using HumanOmniExpress-12v1-1_B arrays (Illumina).[22] To control for platform effects, raw intensity data were provided to the PGC CNV analysis group. This provided a centralised pipeline for systematic CNV calling including multiple CNV callers run in parallel. The final CNV set was defined as those >20 kb in length and including at least 10 probes and <1% minor allele frequency (MAF).[2]

Statistical analyses

Univariate analyses (Fisher's exact tests) were performed first, to assess associations between phenotypic predictors and SCZ-associated CNV status in the Irish cohort. Multiple logistic regression analysis was then carried out to examine the effects of significant phenotypic variables, identified on univariate analysis, in modelling SCZ-associated CNV status. The final independent variables included in the model were those with a significance level of 0.05 following backward elimination steps. Model fit was assessed using Nagelkerke pseudo R^2 index.

Receiver operating characteristic (ROC) curve analysis was used to test the validity, sensitivity and specificity of the logistic regression parameters for modelling SCZ-associated CNV carrier status in the Irish discovery data-set.

The Cardiff replication data-set included data on two of the phenotypic variables of interest. A multiple logistic regression model including these two variables was trained from the Irish discovery data-set and then applied to the Cardiff data-set. ROC curve analysis was used to assess the accuracy of the neurodevelopmental variables in modelling SCZ-associated CNV carrier status in the replication data-set.

The results presented are not corrected for multiple comparisons and all analyses were completed in R version 3.2.3 for Windows.[23]

Results

Discovery data-set

From the total sample of 1215 individuals in the Irish discovery data-set, 19 (1.6%) carried one of the 15 identified SCZ-associated CNVs.[20] No individuals carried more than one SCZ-associated CNV. The details of the CNVs and positions are listed in supplementary Table 4. The proportions of individuals with a positive history of phenotypic variables and SCZ-associated pathogenic CNV status are available in supplementary Table 5.

Univariate analyses identified four phenotypic variables with significant associations with SCZ-associated CNV status: 'history of developmental delay', 'comorbid neurodevelopmental disorder', 'history of learning difficulties' and 'specific learning disorder' (Table 1). A multiple logistic regression model was fitted using these four variables. The variables 'history of learning difficulties' and 'specific learning disorder' were correlated (phi coefficient φ = 0.22) and were likely capturing similar phenotypic information. Backward elimination at this point removed the variable 'history of learning difficulties' from the model. The final independent variables in the model were 'history of developmental delay', 'comorbid neurodevelopmental disorder' and 'specific learning disorder'. These variables had odds ratios of 5.19 (95% CI 1.58–14.76, P = 0.003), 5.87 (95% CI 1.28–19.69, P = 0.009) and 8.12 (95% CI 1.16–34.88, P = 0.012) respectively when included in the logistic regression model (Table 2). Nagelkerke pseudo R^2 for the model was 0.196, indicating that the phenotypic variables accounted for 19.6% of the variance in SCZ-associated CNV status in this sample.

The performance of the three significant independent variables in modelling SCZ-associated CNV carrier status was tested using ROC curve analysis. An area under the ROC (AUROC) curve of 74.2% (95% CI 61.9–86.4%) was achieved, accounting for 58.8% (95% CI 32.9–81.6%) sensitivity and 89.1% (95% CI 87.1–90.9%) specificity in modelling SCZ-associated CNV carrier status (Table 3).

Replication data-set

Eight individuals (1.7%) in the Cardiff replication data-set (n = 479) set carried one of the 15 identified risk CNVs, including one 1q21.2 duplication, one NRXN1 deletion, one Williams–Beuren region duplication, three 15q11.2 deletions and two 22q11.2 deletions. No individual carried more than one of these CNVs.

The Cardiff replication data-set included data on two of the phenotypic variables of interest: 'history of developmental delay' and 'comorbid neurodevelopmental disorder'. The Irish discovery data-set was used to build a multiple logistic regression model using these two variables (supplementary Tables 6 and 7). Applying this model to the Cardiff study population gave an AUROC of 83% (95% CI 52.0–100.0%) in identifying SCZ-associated CNV status. The sensitivity and specificity were 75.0% (95% CI 19.4–99.4%) and 97.6% (95% CI 95.1–99.0%) respectively (Table 4).

Discussion

We investigated whether phenotype information generated by a standard clinical assessment could identify people with schizophrenia at greater risk of carrying pathogenic CNVs. In a discovery cohort of 1215 people with schizophrenia, having a specific learning disorder (OR = 8.12, P = 0.012), developmental delay (OR = 5.19, P = 0.003) or a comorbid neurodevelopmental disorder (OR =

Table 4 Receiver operating characteristic (ROC) curve results[a] for modelling schizophrenia-associated copy number variant status in the replication (Cardiff) data-set

Cut-off value	Sensitivity, % (95% CI)	Specificity, % (95% CI)	AUC, % (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)
0.241	75.0 (19.4–99.4)	97.6 (95.1–99.0)	83.0 (52.0–100)	30.0 (17.0–95.7)	99.6 (95.8–99.9)

AUC, area under the curve; PPV, positive predictive value; NPV, negative predictive value.
a. Optimal cut-off value, sensitivity, specificity, AUC and predictive values were calculated using two independent variables ('history of developmental delay', 'comorbid neurodevelopmental disorder').

5.87, $P = 0.009$) successfully modelled positive carrier status for a SCZ-associated CNV. Other clinical features, such as early onset of psychosis, low educational attainment and a family history of neurodevelopmental disorders, were not associated with SCZ-associated CNV carrier status in this cohort. The three 'neurodevelopmental' variables showed a relatively high specificity (89.1% (95% CI 87.1–90.9%)) but more modest sensitivity (58.8% (95% CI 32.9–81.6%)) in modelling carrier status for a SCZ-associated CNV in the Irish discovery sample. Information on 'specific learning disorders' was not available for the Cardiff replication sample. On the basis of the remaining two variables, 'comorbid neurodevelopmental disorder' and 'history of developmental delay', we applied a model from the original data-set to the Cardiff sample. This too showed relatively high specificity (97.6% (95% CI 95.1–99.0%)) but more modest sensitivity (75.0% (95% CI 19.4–99.4%)) in modelling carrier status for a SCZ-associated CNV.

Recent studies have suggested that identifying people with schizophrenia who have comorbid intellectual disability is likely to be helpful in identifying subsets of individuals with genomic disorders. Thygesen and colleagues reported an approximately three-fold higher rate of pathogenic CNVs in people with psychosis and intellectual disability compared with rates in the general schizophrenia population.[24] Lowther et al examined the genome-wide burden of pathogenic CNVs in a schizophrenia cohort ($n = 546$) and demonstrated a significantly higher burden of pathogenic CNVs (OR = 5.01, $P = 0.0001$) in people with schizophrenia and low IQ (IQ <85) compared with those with average IQ (IQ ≥85). On the basis of their findings, the authors concluded that individuals with schizophrenia and low IQ should be prioritised for clinical microarray testing in clinical and research contexts.[25] We believe that our study provides further support for this recommendation, but that other developmental indices, which could be captured by a clinical neurodevelopmental history, should also be considered in the development of any future guidelines.

Implications

A small subset of people with schizophrenia (~2.5%) carry CNVs that substantially increase the risk for schizophrenia but also for other neurodevelopmental disorders. The clinical benefits of identifying such people been demonstrated for other neurodevelopmental disorders.[5,6] Similar benefits are likely to apply in schizophrenia, but as these events are rare, routine genetic testing for all individuals is probably not indicated. Previous studies suggest that targeting people with schizophrenia and comorbid intellectual disability is likely to be more fruitful in identifying such cases.[24,25] Our findings suggest that careful clinical history taking to document developmental delay, reported learning disorders or a comorbid diagnosis of autism spectrum disorder or epilepsy may also be informative in screening for people with schizophrenia at higher risk of carrying known SCZ-associated CNVs.

These are rare events, but very large cohorts of genotyped people with schizophrenia are available and it is likely that whole genome sequence analysis of >30 000 such individuals will soon be completed. As these data are analysed the subset of people with schizophrenia who carry rare mutations and CNVs of likely

clinical significance will increase, as has been the case for other neurodevelopmental disorders. Regrettably, there is a dearth of phenotype information available from many of the contributory cohorts. We strongly support efforts by the PGC to collect and standardise such phenotype information where it is available. For future cohorts, having detailed phenotype information together with neurodevelopmental and medical history will likely be helpful in refining predictor variables that ultimately may inform guidelines for genetic testing for people with schizophrenia.

Strengths and limitations

The strength of our study lies in the fact that we were able to build a well-characterised phenotype data-set, based on extensive clinical and research data compiled from previous schizophrenia research studies. We were able to test multiple phenotypic features for potential in identifying pathogenic CNV status and identify three variables that are easily clinically identified and that show considerable promise in identifying a high-risk group.

Recurrent SCZ-associated CNVs are rare events (~1:150–1:1000)[3] and individual cohorts are likely to identify only a modest number of known CNVs, as demonstrated in our sample of 1215 people with schizophrenia. The study highlighted the relative limitations of phenotypic information across schizophrenia cohorts and suggested phenotypes derived from a standard clinical interview that could inform future studies. Further analysis of a wider psychosis population and other cross-disorder analyses are also likely to be valuable.

Our discovery and replication cohorts used retrospective phenotypic data from which we identified variables that provided estimates of sensitivity and specificity for modelling SCZ-associated CNV carrier status. Significantly larger, well-characterised phenotypic samples (e.g. prospective cohorts) will be required to provide more refined estimates of sensitivity and specificity to inform genetic screening guidelines. It will be important to consider the patient and family perspective to inform any future guidelines for genetic testing, but that was beyond the scope of the current investigation.

Claire Foley [ID], MB, MRCPsych, Clinical Research Fellow, Neuropsychiatric Genetics Research Group, Department of Psychiatry, School of Medicine, Trinity College Dublin, Ireland; **Elizabeth A. Heron**, PhD, Assistant Professor, Neuropsychiatric Genetics Research Group, Department of Psychiatry, School of Medicine, Trinity College Dublin, Ireland; **Denise Harold**, PhD, Assistant Professor, School of Biotechnology, Dublin City University, Ireland; **James Walters**, MRCPsych, PhD, Professor, MRC Centre for Neuropsychiatric Genetics and Genomics, Division of Psychological Medicine and Clinical Neurosciences, School of Medicine, Cardiff University, UK; **Michael Owen**, FRCPsych, PhD, Director, MRC Centre for Neuropsychiatric Genetics and Genomics, Division of Psychological Medicine and Clinical Neurosciences, School of Medicine, Cardiff University, UK; **Michael O'Donovan**, FRCPsych, PhD, Professor, MRC Centre for Neuropsychiatric Genetics and Genomics, Division of Psychological Medicine and Clinical Neurosciences, School of Medicine, Cardiff University, UK; **Jonathan Sebat**, PhD, Chief, Beyster Center for Genomics of Psychiatric Diseases, Departments of Psychiatry, Cellular and Molecular Medicine and Pediatrics, University of California, San Diego; and Professor, Institute for Genomic Medicine, University of California, San Diego; and Department of Pediatrics, University of California, San Diego, USA; **Eric Kelleher**, MRCPsych, PhD, Honorary Clinical Senior Lecturer, Department of Psychiatry and Neurobehavioural Science, University College Cork; and Visiting Research Fellow, Neuropsychiatric Genetics Research Group, Department of Psychiatry, School of Medicine, Trinity College Dublin, Ireland; **Christina Mooney**, HDip in Mental Health Nursing, Clinical Research Nurse, Neuropsychiatric Genetics Research Group, Department of Psychiatry, School of Medicine, Trinity College Dublin, Ireland; **Amy Durand**, Medical Student, University of Texas Health Science Center at Houston, McGovern Medical School, Texas, USA; and

Research Assistant, Neuropsychiatric Genetics Research Group, Department of Psychiatry, School of Medicine, Trinity College Dublin, Ireland; **Carlos Pinto**, PhD, Research Fellow, Neuropsychiatric Genetics Research Group, Department of Psychiatry, School of Medicine, Trinity College Dublin, Ireland; **Paul Cormican**, PhD, Lecturer, Neuropsychiatric Genetics Research Group, Department of Psychiatry, School of Medicine, Trinity College Dublin, Ireland; **Derek Morris**, PhD, Lecturer, Cognitive Genetics and Cognitive Therapy Group, Neuroimaging, Cognition and Genomics (NICOG) Centre, School of Psychology and Discipline of Biochemistry, National University of Ireland Galway, Ireland; **Gary Donohoe**, PhD, Professor, Cognitive Genetics and Cognitive Therapy Group, Neuroimaging, Cognition and Genomics (NICOG) Centre, School of Psychology and Discipline of Biochemistry, National University of Ireland Galway, Ireland; **Michael Gill**, MD, MRCPsych, FTCD, Professor, Head of School of Medicine, Neuropsychiatric Genetics Research Group, Department of Psychiatry, School of Medicine, Trinity College Dublin, Ireland; **Louise Gallagher**, MB, MRCPsych, PhD, Director of Research, School of Medicine, Trinity College Dublin; and Professor, Neuropsychiatric Genetics Research Group, Department of Psychiatry, School of Medicine, Trinity College Dublin, Ireland; **Aiden Corvin**, MB, MRCPsych, PhD, Professor, Head of Discipline, Neuropsychiatric Genetics Research Group, Department of Psychiatry, School of Medicine, Trinity College Dublin, Ireland.

Correspondence: Professor Aiden Corvin. Email: acorvin@tcd.ie

First received 27 Jan 2019, final revision 18 Oct 2019, accepted 22 Nov 2019

Data availability

Data are available from the authors on reasonable request.

Author contributions

C.F., E.A.H., M.G., L.G. and A.C. were responsible for the study conception and design. C.M., E.K., C.F. and A.D. were responsible for collection and coding of primary phenotypic data from the discovery data-set. C.M., E.K., D.H., D.M., C.P., P.C. and G.D. were responsible for the collection and processing of the genetic data from the discovery data-set. J.W., M.O. and M.O'D. contributed the data for the replication data-set and provided feedback with regard to the data analysis and interpretation. J.S. contributed to the core analysis of the genetic data used in the study. C.F., E.A.H., L.G. and A.C. contributed to the data analysis and interpretation and drafted the manuscript. All authors reviewed and approved the final manuscript.

Supplementary material

Supplementary material is available online at https://doi.org/10.1192/bjp.2019.262.

References

1 International Schizophrenia Consortium, Purcell SM, Wray NR, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 2009; **460**: 748–52.

2 Marshall CR, Howrigan DP, Merico D, Thiruvahindrapuram B, Wu W, Greer DS, et al. Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nat Genet* 2017; **49**: 27–35.

3 Kirov G, Rees E, Walters JT, Escott-Price V, Georgieva L, Richards AL, et al. The penetrance of copy number variations for schizophrenia and developmental delay. *Biol Psychiatry* 2014; **75**: 378–85.

4 Kendall KM, Rees E, Escott-Price V, Einon M, Thomas R, Hewitt J, et al. Cognitive performance among carriers of pathogenic copy number variants: analysis of 152,000 UK biobank subjects. *Biol Psychiatry* 2017; **82**: 103–10.

5 Miller DT, Adam MP, Aradhya S, Biesecker LG, Brothman AR, Carter NP, et al. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *Am J Hum Genet* 2010; **86**: 749–64.

6 Schaefer GB, Mendelsohn NJ. Clinical genetics evaluation in identifying the etiology of autism spectrum disorders: 2013 guideline revisions. *Genet Med* 2013; **15**: 399–407.

7 Philip N, Bassett A. Cognitive, behavioural and psychiatric phenotype in 22q11.2 deletion syndrome. *Behav Genet* 2011; **41**: 403–12.

8 Ahn K, Gotay N, Andersen TM, Anvari AA, Gochman P, Lee Y, et al. High rate of disease-related copy number variations in childhood onset schizophrenia. *Mol Psychiatry* 2014; **19**: 568–72.

9 Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM, et al. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* 2008; **320**: 539–43.

10 Yeo RA, Gangestad SW, Liu J, Ehrlich S, Thoma RJ, Pommy J, et al. The impact of copy number deletions on general cognitive ability and ventricle size in patients with schizophrenia and healthy control subjects. *Biol Psychiatry* 2013; **73**: 540–5.

11 Stefansson H, Meyer-Lindenberg A, Steinberg S, Magnusdottir B, Morgen K, Arnarsdottir S, et al. CNVs conferring risk of autism or schizophrenia affect cognition in controls. *Nature* 2014; **505**: 361–6.

12 Derks EM, Ayub M, Chambert K, Del Favero J, Johnstone M, MacGregor S, et al. A genome wide survey supports the involvement of large copy number variants in schizophrenia with and without intellectual disability. *Am J Med Genet B Neuropsychiatr Genet* 2013; **162B**: 847–54.

13 Sahoo T, Theisen A, Rosenfeld JA, Lamb AN, Ravnan JB, Schultz RA, et al. Copy number variants of schizophrenia susceptibility loci are associated with a spectrum of speech and developmental delays and behavior problems. *Genet Med* 2011; **13**: 868–80.

14 Wilson NK, Lee Y, Long R, Hermetz K, Rudd MK, Miller R, et al. A novel micro-duplication in the neurodevelopmental gene SRGAP3 that segregates with psychotic illness in the family of a COS proband. *Case Rep Genet* 2011; **2011**: 585893.

15 Costain G, Lionel AC, Fu F, Stavropoulos DJ, Gazzellone MJ, Marshall CR, et al. Adult neuropsychiatric expression and familial segregation of 2q13 duplications. *Am J Med Genet B Neuropsychiatr Genet* 2014; **165B**: 337–44.

16 Miles JH, Takahashi TN, Hong J, Munden N, Flournoy N, Braddock SR, et al. Development and validation of a measure of dysmorphology: useful for autism subgroup classification. *Am J Med Genet A* 2008; **146A**: 1101–16.

17 Irish Schizophrenia Genomics Consortium, Wellcome Trust Case Control Consortium 2. Genome-wide association study implicates HLA-C*01:02 as a risk factor at the major histocompatibility complex locus in schizophrenia. *Biol Psychiatry* 2012; **72**: 620–8.

18 First MB, Spitzer RL, Gibbon M, Williams JBW. *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Patient Edition (SCID-I/P)*. Biometrics Research, New York State Psychiatric Institute, 2002.

19 Hamshere ML, Walters JT, Smith R, Richards AL, Green E, Grozeva D, et al. Genome-wide significant associations in schizophrenia to ITIH3/4, CACNA1C and SDCCAG8, and extensive replication of associations reported by the Schizophrenia PGC. *Mol Psychiatry* 2013; **18**: 708–12.

20 Rees E, Walters JTR, Georgieva L, Isles AR, Chambert KD, Richards AL, et al. Analysis of copy number variations at 15 schizophrenia-associated loci. *Br J Psychiatry* 2014; **204**: 108–14.

21 Malhotra D, Sebat J. CNVs: harbingers of a rare variant revolution in psychiatric genetics. *Cell* 2012; **148**(6): 1223–41.

22 Rees E, Kendall K, Pardiñas AF, Legge SE, Pocklington A, Escott-Price V, et al. Analysis of intellectual disability copy number variants for association with schizophrenia. *JAMA Psychiatry* 2016; **73**: 963–9.

23 R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, 2013 (http://www.R-project.org/).

24 Thygesen JH, Wolfe K, McQuillin A, Viñas-Jornet M, Baena N, Brison N, et al. Neurodevelopmental risk copy number variants in adults with intellectual disabilities and comorbid psychiatric disorders. *Br J Psychiatry* 2018; **212**: 287–94.

25 Lowther C, Merico D, Costain G, Waserman J, Boyd K, Noor A, et al. Impact of IQ on the diagnostic yield of chromosomal microarray in a community sample of adults with schizophrenia. *Genome Med* 2017; **9**(1): 105.

EXTRA CONTENT ONLINE