

Assessment of Potential Clinical Role for Exome Sequencing in Schizophrenia

Thivia Balakrishna¹ and David Curtis^{*,1,2}

¹UCL Genetics Institute, University College London, UK; ²Centre for Psychiatry, Barts and the London School of Medicine and Dentistry, UK

*To whom correspondence should be addressed; UCL Genetics Institute, University College London, Darwin Building, Gower Street, London WC1E 6BT, UK; tel: 00-44-7973-906-143, fax: 00-44-20-7679-3001, e-mail: d.curtis@ucl.ac.uk

Background: There is increasing evidence that certain genetic variants increase the risk of schizophrenia and other neurodevelopmental disorders. Exome sequencing has been shown to have a high diagnostic yield for developmental disability and testing for copy number variants has been advocated for schizophrenia. The diagnostic yield for exome sequencing in schizophrenia is unknown. **Method:** A sample of 591 exome-sequenced schizophrenia cases and their parents were screened for disruptive and damaging variants in autosomal genes listed in the Genomics England panels for intellectual disability and other neurological disorders. **Results:** Previously reported disruptive de novo variants were noted in *SETD1A*, *POGZ*, *SCN2A*, and *ZMYND11*. Although the loss of function of *ZMYND11* is a recognized cause of intellectual disability, it has not previously been noted as a risk factor for schizophrenia. A damaging de novo variant of uncertain significance was noted in *NRXN1*. A previously reported homozygous damaging variant in *BLM* is predicted to cause Bloom syndrome in 1 case and 1 case was homozygous for a damaging variant in *MCPHI*, a result of uncertain significance. There were more than 400 disruptive and damaging variants in the target genes in cases but similar numbers were seen among untransmitted parental alleles and none appeared to be clinically significant. **Conclusions:** The diagnostic yield from exome sequencing in schizophrenia is low. Disruptive and damaging variants seen in known neuropsychiatric genes should not be automatically assumed to have an etiological role if observed in a patient with schizophrenia.

Key words: variant/DNA/gene/trio

Introduction

It is now well established that deletions and duplications at specific chromosomal locations, termed copy number variants (CNVs), can have a substantial effect on the risk of developing schizophrenia and are found in 3%

of people with schizophrenia who have normal IQ and 7% of those with borderline IQ.^{1,2} Because of their clinical implications, it has been advocated that testing for these pathogenic CNVs should be a routine investigation for patients diagnosed with schizophrenia.³ Identifying a CNV with a substantial effect on risk could help by providing the patient with an explanation of why they have become unwell and might improve compliance with treatment, as well as having implications for relatives.⁴ As well as CNVs, polygenic risk factors and non-genetic risk factors, there is now good evidence that rare coding variants contribute to schizophrenia risk.⁵ There is strong statistical evidence that schizophrenia cases have an excess burden of rare variants predicted to damage the function of particular sets of genes, eg, those related to synaptic functioning.^{6,7} However, only a small number of individual genes are strongly implicated. Currently, we would argue that there is good evidence to support the claim that variants causing loss of function (LOF) of *SETD1A*, *RBM12*, and *NRXN1* can substantially increase the risk of schizophrenia.^{8–10} However, these are extremely rare and are collectively found in fewer than 1% of cases.

It has become apparent that the relationship between genotype and phenotype is often more complex than had originally been supposed and that variants which had been thought to have a simple Mendelian effect might manifest reduced penetrance.¹¹ In the context of psychiatric disorders, an important general theme which emerges from the findings to date is that a variant which confers increased risk of schizophrenia may also confer increased risk of intellectual disability (ID) or other neurodevelopmental disorder, suggesting a form of pleiotropy in that processes which disrupt neurodevelopment may lead to ID, schizophrenia, both, or neither. This applies to both CNVs and, eg, LOF variants of *SETD1A*, which are found both in schizophrenia cases and in subjects diagnosed with the developmental disorder.⁸

In general, we find that rare, damaging coding variants in schizophrenia cases affect genes implicated in other neurodevelopmental disorders.¹² Another example is a recently reported case with a de novo variant in *SCN2A* who initially presented with infantile-onset seizures, autistic features, and episodic ataxia but then developed psychotic symptoms as an adult.¹³

In contrast to schizophrenia, several hundred genes are implicated as risk factors for ID, in the sense that it is known that specific variants in these genes can produce a phenotype, which includes ID. Given the pleiotropy described earlier, if one detected in a patient with schizophrenia a variant known to cause ID then one might reasonably conclude that in this case, the variant was likely making an etiological contribution to the schizophrenia. Testing for variants in known ID genes, rather than only the handful of currently known schizophrenia genes, might produce a worthwhile diagnostic yield and this approach has been implemented in a pilot study incorporated in the 100 000 Genomes Project, a diagnostic service implemented by the UK National Health Service. Here, patients with schizophrenia accompanied by other features suggestive of a genetic etiology undergo whole-genome sequencing and are then checked for variants in genes known to be causative of ID and other neurological conditions. In children with severe developmental disorders, exome sequencing can produce an overall diagnostic yield of about 40%.¹⁴ We thought it would be of interest to gain a better understanding of what the likely yield might be for patients with schizophrenia by applying the proposed diagnostic testing protocol to a sample of research subjects for whom exome sequencing data were already available.

The Bulgarian sample consists mainly of a set of schizophrenia cases along with their parents who have undergone exome sequencing and who have been studied for the presence of de novo mutations and recessively acting variants.^{15,16} These previous reports were gene discovery projects aiming to build evidence to identify novel genes conferring susceptibility to schizophrenia, and indeed provided some of the evidence to implicate *SETD1A*. We set out to use this sample as if it were a clinical sample with the aim of identifying any variants, acting dominantly or recessively, which might in a clinical context be reported as pathogenic or likely pathogenic.

Methods and Materials

This sample comprised exome sequence data from 591 Bulgarian trios, consisting of probands with schizophrenia and their parents, 5 of whom were also affected.^{15,16} The authors of the original studies reported that all probands and parents provided written informed consent and that ethical committee approval was obtained from the hospitals from where the cases were recruited. The short read files were downloaded from dbGaP along

with family structure and phenotype information (https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000687.v1.p1). The short read files were converted to fastq files using the fastq-dump utility of the dbGaP SRA toolkit. Reads were then aligned to the hg19 human reference sequence (build GRCh37) using Novoalign V3.02.08 (NovoCraft Technologies), duplicate reads were marked using SAMBLASTER,¹⁷ and the BAM files were sorted using Novosort V1.03.09 (NovoCraft Technologies). Genotypes were called according to GATK best practices (Broad Institute). The HaplotypeCaller module of GATK V3.6 was used to produce gVCF files and these were then combined using the CombineGVCFs module. Initial calls were made using the GenotypeGVCFs module and then single nucleotide variants (SNPs) were filtered based on accuracy estimates produced by VariantRecalibrator and indels were filtered using the VariantFiltration module and the filtering expression “QD < 2.0 || FS > 50.0 || ReadPosRankSum < -20.0” to generate a PASS entry in the information field of the resultant VCF file. Variants were excluded if they did not have a PASS in the information field and individual genotype calls were excluded if they had a genotype quality score less than 20. Heterozygote calls were excluded if 1 allele accounted for less than 0.2 of total reads. Variants were also excluded if more than 10% of cases or of pseudocontrols failed this quality threshold or if the heterozygote count was smaller than both homozygote counts in both cohorts. Variants were annotated with using variant effect predictor (VEP), PolyPhen, and SIFT.^{18–20}

To identify sets of genes for which there was strong evidence of involvement in neurodevelopmental disorders, such that they might be used clinically, we downloaded panels generated for the 100 000 Genomes Project from the Genomics England Panel App (<https://panelapp.genomicsengland.co.uk/panels/>). We downloaded version 1.13 of these panels: Familial Focal Epilepsies, Familial Genetic Generalised Epilepsies, Genetic Epilepsy Syndromes, and ID. Attention was restricted to well-supported genes (color-coded Green), which were implicated in their own right rather than being listed only through lying within a pathogenic deletion or duplication. To handle transmissions from parents in a consistent manner, attention was restricted to autosomal genes. Genes reported to have a monoallelic effect, with the addition of *SETD1A*, were incorporated in a list of 328 putative dominant genes. A list of 942 putative recessive genes included the dominant genes along with the panel genes reported to have biallelic effects.

The GENEVARASSOC program (<https://github.com/davenomiddlenamecurtis/geneVarAssoc>) was used to export variants in each gene to SCOREASSOC.^{21–23} SCOREASSOC is able to use trio data to construct a sample of cases and pseudocontrols with genotypes consisting of the untransmitted alleles from the parents

of each affected proband. It can also detect de novo variants, which are present in offspring but not parents and can detect compound heterozygotes where a gene contains both paternally and maternally inherited variants. Stop gained, frameshift, splice site, and transcript ablation variants were classified as disruptive whereas nonsynonymous variants characterized by SIFT as deleterious or by PolyPhen as probably damaging were classified as damaging. Version hg19 of the reference human genome sequence and RefSeq genes were used to extract variants on a gene-wise basis. The gnomAD database (<http://gnomad.broadinstitute.org/>) was used to obtain minor allele frequencies (MAFs) for variants in populations with different ancestries.¹¹ For analyses of dominant effects, variants were excluded if they had $MAF > 0.005$ in any population or if more than 6 cases or pseudocontrols were heterozygous or homozygous for the minor allele. For the analyses of recessive effects, the same exclusions were applied, but using a threshold of $MAF > 0.05$. We reasoned that any variants with higher allele frequencies would have been discovered in previous genome-wide association studies if they had appreciable effects on risk.

Qualifying damaging or disruptive variants occurring in the dominant genes, or as homozygotes or compound heterozygotes in the recessive genes, were investigated further by consulting OMIM (<https://omim.org>), ClinVar,²⁴ GeneCards,²⁵ and PubMed (<https://www.ncbi.nlm.nih.gov/pubmed>) to determine whether there would be sufficient evidence to warrant reporting them in a clinical context. The interpretation of the clinical significance of variants remains a challenging task, which at times is inevitably subjective. When attempting this task, we endeavored to take account of the gene concerned, the related phenotypes, and previous reports of the occurrence of either the variant itself or else of variants with similar predicted effect on protein function.

This study differed from previous analyses of this dataset in that it sought to explore how an application of our current knowledge regarding genes involved in schizophrenia and ID might be used in a clinical context. In addition, although previous studies investigated de novo mutations and recessively acting variants, here we also considered single inherited variants, which might have a dominant effect. For variants of interest, we carried out more intensive consultation of the literature, as would be done in a clinical situation.

Results

The de novo variants have been reported on previously,²⁶ and in panel genes consisted of disruptive variants in *POGZ*, *SCN2A*, *SETD1A*, and *ZMYND11* and damaging variants in *ARID1B* and *NRXN1*. Each of these occurred in a single subject and all had been confirmed by Sanger sequencing in the original study, which had also

reported that both *SCN2A* and *POGZ* were established autism genes.²⁶ LOF variants in *SETD1A* have subsequently been recognized as a well-established risk factor for schizophrenia.^{8,27} LOF of *ZMYND11* has not previously been reported in schizophrenia cases but has been reported to be associated with ID and sometimes features of behavioral and/or mood disturbance, with 1 case being given diagnoses of “rapid cycling bipolar disorder, borderline personality disorder and pervasive developmental disorder, with psychosis and alcohol and drugs abuse”.²⁸ Taken together, it seems that these 4 disruptive variants would be classified as pathogenic or likely pathogenic. The damaging variant in *ARID1B* is 6:157495166 T>C, which is absent from gnomAD and which produces a methionine to threonine substitution at position 426 in the product of transcript ENST00000319584. It is predicted to be “deleterious_low_confidence” and “benign”. LOF mutations of *ARID1B* are a known cause Coffin-Siris syndrome, which comprises ID with other abnormalities but no clear psychiatric phenotype, and isolated nonsynonymous mutations have also been reported in 2 cases.²⁹ Thus, one would probably categorize this variant as being of unknown significance. The damaging variant in *NRXN1* is 2:50724807 T>C, which has MAF 0.000007 in gnomAD but is absent from ClinVar. It produces a glutamic acid to glycine substitution at position 848 in the product of transcript ENST00000401669. It lies in a laminin G domain and is predicted to be “deleterious” and “probably damaging”. Given that there is evidence that disruption of *NRXN1* increases risk of schizophrenia, one would probably categorize this variant as being of unknown significance but one might regard it as likely pathogenic.

In the analysis of dominant genes excluding de novo variants, disruptive variants were seen on 21 occasions in pseudocontrols and 12 in cases. Thus, of disruptive variants present in the parents of schizophrenia probands, there was no tendency for them to be preferentially transmitted to the proband. The disruptive variants seen in cases are listed in [table 1](#). Three cases have LOF mutations in *ABCC9*. One parent had an additional frameshift variant in *ABCC9*, which was not transmitted to a case. Missense mutations of *ABCC9* can cause Cantu syndrome, which can include ID as part of the phenotype but no clinical effects have been reported for LOF variants, except for a single case of dilated cardiomyopathy, hence we do not think the observed variants are likely to be clinically significant.³⁰ A single subject has a frameshift variant in *ASXL1*. De novo stop mutations in transcript NM_015338 of this gene cause Bohring-Opitz syndrome, a severe developmental and malformation disorder characterized by profound mental retardation among other features.³¹ The variant we observe, p.(Thr1217ThrX), is predicted to form a truncated product of this transcript and lies close to one of the previously reported pathogenic variants, p. (Ser1028X).

Table 1. Rare Disruptive Variants Noted in Schizophrenia Cases Along With Their VEP Annotation, Minor Allele Frequency (MAF) in Non-Finnish Europeans and ClinVar Entry

Gene	Variant	Effect	MAF	ClinVar
<i>ABCC9</i>	12:22028600-G>A	stop_gained	0.000004	Uncertain significance
<i>ABCC9</i>	12:22048250-CTG>C	splice_acceptor_variant	0.00003	Absent
<i>ABCC9</i>	12:22086827-A>AT	frameshift_variant	0.000004	Absent
<i>ASXL1</i>	20:31024165-C>CAA	frameshift_variant	0.000004	Absent
<i>HNRNPU</i>	1:245019204-GCCT>G	splice_donor_variant	0.0006	Benign/Likely_benign
<i>HNRNPU</i>	1:245019204-GCCT>G	splice_donor_variant	0.0006	Benign/Likely_benign
<i>KCNT1</i>	9:138671283-CA>C	frameshift_variant	0.000009	Absent
<i>NFI</i>	17:29645324-A>G	splice_donor_variant	0.00002	Absent
<i>PDE4D</i>	5:58270510-T>TCATCTATGACA	stop_gained	0.00001	Absent
<i>PTCHI</i>	9:98279020-C>CG	frameshift_variant	0.000009	Absent
<i>SAMD9</i>	7:92735007-G>GAACCTTT	stop_gained	0.000009	Absent
<i>SMARCA2</i>	9:2191408-T>TGTA	frameshift_variant	0.00002	Absent

However, the normal product is only 1541 amino acids in length and all the pathogenic known variants lead to a more drastic truncation so if this finding does not represent a genotyping error, it is possible that the small truncation it causes does not have severe consequences. In any event, its clinical significance seems uncertain. Two cases have 1:245019204-GCCT>G, which is a splice site variant for some transcripts of *HNRNPU* but is upstream or downstream of other transcripts. A variety of types of variant causing LOF of *HNRNPU* cause developmental delay and seizures³² but this particular variant is classified by ClinVar as “Benign/Likely_benign.” One case has a frameshift variant in *KCNT1*, which codes for a sodium-activated potassium channel. Although heterozygous gain-of-function mutations in this gene can cause infantile epilepsy,³³ there is no phenotype associated with LOF. One subject had a variant in *NFI*, which affects the splice site of some transcripts but is upstream or intronic for others. Although mutations of *NFI* can cause neurofibromatosis with learning disability,³⁴ this particular variant is absent from ClinVar and its significance is unclear. One subject has a stop gained and frameshift variant in *PDE4D*. Heterozygous missense variants in this gene can cause acrodysostosis and acroscaphodysplasia with ID but these seem to be through gain-of-function effects and no phenotype is associated with LOF.^{35,36} One subject has a frameshift variant in *PTCHI*. There is a report of a mother and son with holoprosencephaly like phenotype who had a duplication of this gene but there is no evidence for a psychiatric phenotype associated with LOF.³⁷ One subject has a stop gained and frameshift variant in *SAMD9*. Missense variants in this gene can cause MIRAGE system, a multisystem disorder that can have ID as part of the phenotype.³⁸ These effects seem to be due to gain of function, whereas LOF variants have been reported to be associated with adult myelodysplastic syndrome but no psychiatric phenotype.³⁹ One subject has a frameshift variant in *SMARCA2*. De novo missense (but not LOF) variants in this gene, which is involved in chromatin remodeling and neural development, cause

Nicolaides-Baraitser syndrome, which includes severe mental retardation.⁴⁰ There is a report of SNPs in *SMARCA2* being associated with schizophrenia and that psychotogenic drugs reduce its expression whereas anti-psychotic drugs increase it.⁴¹ Although the observation of this variant may be potentially of some scientific interest, there is not sufficient evidence to justify making any clinical interpretation of it. To summarize, more disruptive variants are seen in untransmitted parental alleles than occur in cases. Of the 12 which were seen in cases only 1, a truncating variant in *ASXL1*, might be viewed as possibly pathogenic and even for this variant, the case is very arguable. The others would all be viewed as benign or of uncertain significance.

With respect to damaging variants, these were seen on 429 occasions in pseudocontrols and 467 in cases, a nonsignificant difference. 594 were absent from ClinVar and the rest were judged to be benign, likely benign or of uncertain significance, split approximately equally between pseudocontrols and cases. None was reported by ClinVar to be likely to be pathogenic when occurring as a heterozygote.

In the recessive analyses, 1 case and 1 pseudocontrol had disruptive variants in both copies of a gene. The case was described in the original report and is homozygous for 15:91304245 C>T, which causes a stop mutation in the *BLM* gene. LOF variants in both copies of this gene cause Bloom syndrome and in addition to schizophrenia this subject was noted to have vitiligo and epilepsy. A pseudocontrol had 2 copies of 1:97915614 C>T, which produces a splice site variant in the *DYPD* gene. A subject homozygous for this variant would be expected to manifest dihydropyrimidine dehydrogenase deficiency but of course the pseudocontrol does not actually consist of a single individual but rather represents the genotype of the parental alleles not transmitted to the corresponding case.

In addition to these 2 subjects, a further 9 cases and 4 pseudocontrols had a damaging and/or disruptive variant in both copies of a gene. The variants that were seen

Table 2. Homozygote and Compound Heterozygote Genotypes of Damaging and/or Disruptive Variants Observed in Schizophrenia Cases

Gene	Haplotype	Variant	Transcript	Effect	SIFT	PolyPhen	ClinVar	MAF
<i>AGA</i>	Paternal	4:178360811-G>T	ENST00000264595	L105I	Tolerated (0.06)	probably_dam-aging(0.954)	Benign	0.012
<i>FH</i>	Maternal Paternal	(Homozygous) 1:241669398-T>C	ENST00000366560	Y270C	Deleterious (0.03)	benign(0.243)	Uncertain significance	0.0003
<i>LRP2</i>	Maternal Paternal	(Homozygous) 2:170042245-T>C	ENST00000263816	N3205D		probably_dam-aging(0.999)	Absent	0.00002
<i>MCPHI</i>	Maternal Paternal	(Homozygous) 8:6371240-C>G	ENST00000325203	L386F	Tolerated (0.15)	probably_dam-aging(0.972)	Absent	0.003
<i>MTRR</i>	Maternal Paternal	(Homozygous) 5:7885945-A>G	ENST00000264668	I372M	Deleterious (0.03)	benign(0.208)	Absent	0.0003
<i>PCCB</i>	Maternal Paternal	(Homozygous) 3:136002730-C>T	ENST00000469217	P219S	deleterious_low_confidence (0.01)	probably_dam-aging(0.998)	Uncertain significance	0.007
<i>RELN</i>	Maternal Paternal	(Homozygous) 7:103234828-G>C	ENST00000428762	I1217M	Deleterious (0)	possibly_dam-aging(0.643)	Conflicting interpretations of pathogenicity: Likely benign ³ or Uncertain significance ³	0.004
<i>SCN3A</i>	Maternal Paternal	7:103138569-G>A 2:165950891-C>T	ENST00000428762 ENST00000283254	T2933I R1510H	Deleterious (0.01) Deleterious (0.01)	possibly_dam-aging(0.493) probably_dam-aging(0.977)	Likely benign Absent	0.0002 0.000008
<i>SLX4</i>	Maternal Paternal Maternal	2:166003301-G>A 16:3639713-G>A (Homozygous)	ENST00000283254 ENST00000294008	S540F S1309F	Deleterious (0.04) Deleterious (0.01)	probably_dam-aging(0.986) probably_dam-aging(0.962)	Likely benign Uncertain significance	0.002 0.0004

For the compound heterozygotes, both the paternal and maternal allele are listed or else the genotype is noted to be homozygous. A transcript is given for which the stated SIFT and PolyPhen annotations were output. Also, shown are the ClinVar entry and minor allele frequency (MAF) in non-Finnish Europeans.

in cases are listed in table 2. In a clinical context, none could be interpreted as likely to be pathogenic. In addition, certain homozygous genotype calls may be inaccurate because it seems unlikely that one would observe a homozygote for a variant with allele frequency less than 0.001 in an outbred population. Again, in a clinical situation, any potentially pathogenic variants would need to be confirmed by Sanger sequencing. The observation of a subject homozygous for a probably damaging variant in *MCPHI*, a gene in which recessively acting variants can cause microcephaly, is possibly of some interest because of reports of variants in this gene being associated with schizophrenia and bipolar disorder.⁴²⁻⁴⁴ However, the effects of this particular variant are too uncertain for this finding to have any value in a clinical context.

Discussion

Exome sequencing of 591 schizophrenia cases and their parents and examining genes known to cause neuropsychiatric phenotypes identifies 4 cases with a disruptive de novo variant which would be categorized as likely pathogenic, in the genes *SETD1A*, *POGZ*, *SCN2A*, and *ZMYND11*. Two further cases have damaging de novo variants in *ARID1B* and *NRXN1* which would probably be classified as of unknown significance, although it could be argued that the *NRXN1* variant was likely pathogenic. An additional 12 subjects have inherited a disruptive variant in a known susceptibility gene but we argue that in a clinical context all of these would be classified as of unknown significance. Importantly, 21 of such disruptive variants were seen in parents but not transmitted to the schizophrenic proband. Likewise, there were 467 damaging variants that were inherited by cases but another 429 variants that were seen in parents but not transmitted. All of these appeared to be of benign or of unknown significance. With respect to recessively acting variants, we observed the previously reported homozygous variant disruptive of *BLM* but no other variants, which would be reported as likely pathogenic. The *BLM* finding is probably not of relevance to the diagnosis of schizophrenia, as psychosis is not a recognized phenotypic component of Bloom syndrome but should be regarded as an incidental finding with implications for management, eg, heightened surveillance for neoplasia.⁴⁵ The only other finding of note is a case homozygous for a damaging variant in *MCPHI* but in a clinical context, this would simply be reported as of unknown significance.

In summary, carrying out exome sequencing has yielded a clinically significant incidental finding in 1 case and there are an additional 4, or possibly 5, cases with a variant, which could be regarded as likely pathogenic. The use of trios rather than single cases means that it is possible to recognize that there is little or no tendency for disruptive and damaging variants in these genes to be seen at an increased frequency in

schizophrenia cases. They are not more common among cases than among the untransmitted parental alleles. We regard this as an important finding. One could easily envisage a clinical situation in which a patient with schizophrenia underwent sequencing and a disruptive or damaging variant was detected in a gene well established to cause a neuropsychiatric disorder. The sequencing might have been requested by a clinician or have been purchased from a commercial provider by the patient and their family. Our results caution that such a variant should not be regarded as likely pathogenic for schizophrenia unless there is very strong and specific evidence to support this interpretation. The vast majority of such variants should be regarded as simply being of unknown significance.

Our results also suggest that there may be considerable value in sequencing parents alongside cases, especially when there is no family history of schizophrenia. If a variant with ambiguous significance is inherited from a parent with a normal phenotype one may feel more confident that it is unlikely to be clinically relevant. Conversely, if a de novo variant is observed in a gene with a neuropsychiatric phenotype one may be more likely to interpret it as likely pathogenic.

Our study differs from an actual clinical scenario in a number of ways. We have not considered sex chromosome variants. We have not validated variants using Sanger sequencing, although this was previously done for the de novo mutations. We have used exome sequence data and it is possible that whole-genome sequencing would provide more complete coverage of all genes as well as providing information about non-coding variants, CNVs, and karyotype abnormalities.

We note a related study, which has taken a similar approach but using a different panel of genes, consisting of 37 genes known to cause inborn errors of metabolism in which psychosis could occur as the first presenting feature.⁴⁶ An exome sequenced sample of 2545 schizophrenia cases and 2545 controls was investigated. Overall, there was a 50% enrichment of rare disruptive variants in cases and although this finding is scientifically interesting no variants were identified for which one could make a clinical interpretation that they were likely pathogenic with respect to the schizophrenia phenotype.

The diagnostic yield for exome sequencing of known neuropsychiatric genes in this sample is about 1%, lower than the expected yield for testing for pathogenic CNVs of 3% or more and much lower than the yield of 40% obtained in severe developmental disorders. We can expect that the situation may change if and when more genes are identified as having an etiological role in schizophrenia and if better methods can be developed for predicting the effects of specific variants. The main conclusion of this investigation is a negative one. It is not the case that a disruptive or damaging variant in a gene known to have a neuropsychiatric phenotype should be

viewed as likely to be pathogenic when seen in a patient with schizophrenia and hence the diagnostic yield from exome sequencing is currently low.

Funding

Work at Cardiff was supported by Medical Research Council (MRC) Centre (G0800509) and Program Grants (G0801418), the European Community's Seventh Framework Programme (HEALTH-F2-2010241909 (Project EU-GEI)). Work at the Icahn School of Medicine at Mount Sinai was supported by the Friedman Brain Institute, the Institute for Genomics and Multiscale Biology and National Institutes of Health grants R01HG005827 (S.M.P.) and R01MH071681 (P.S.). Work at the Wellcome Trust Sanger Institute was supported by The Wellcome Trust (WT089062 and WT098051). The recruitment of the trios in Bulgaria was funded by the Janssen Research Foundation.

Acknowledgments

We are grateful to three anonymous reviewers for their thoughtful comments.

The Bulgarian Trio Sequencing study is an accumulation of exome sequencing performed and/or funded by the Broad Institute, Cardiff University, Icahn School of Medicine at Mount Sinai, and the Wellcome Trust Sanger Institute. Work at the Broad Institute was funded by Fidelity Foundations, the Sylvan Herman Foundation and philanthropic gifts from Kent and Liz Dauten, Ted and Vada Stanley, and an anonymous donor to the Stanley Center for Psychiatric Research. The authors declare they have no conflicts of interest.

References

1. Rees E, Walters JT, Georgieva L, et al. Analysis of copy number variations at 15 schizophrenia-associated loci. *Br J Psychiatry*. 2014;204(2):108–114.
2. Lowther C, Merico D, Costain G, 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.
3. Baker K, Costain G, Fung WL, Bassett AS. Chromosomal microarray analysis—a routine clinical genetic test for patients with schizophrenia. *Lancet Psychiatry*. 2014;1(5):329–331.
4. Curtis D, Adlington K, Bhui KS. Pursuing parity: genetic tests for psychiatric conditions in the UK National Health Service. *Br J Psychiatry*. 2019;214(5):248–250.
5. Coelewij L, Curtis D. Mini-review: update on the genetics of schizophrenia. *Ann Hum Genet*. 2018;82(5):239–243.
6. Genovese G, Fromer M, Stahl EA, et al. Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. *Nat Neurosci*. 2016;19(11):1433–1441.
7. Curtis D, Coelewij L, Liu SH, Humphrey J, Mott R. Weighted burden analysis of exome-sequenced case-control sample implicates synaptic genes in schizophrenia aetiology. *Behav Genet*. 2018;48(3):198–208.
8. Singh T, Kurki MI, Curtis D, et al.; Swedish Schizophrenia Study; INTERVAL Study; DDD Study; UK10 K Consortium. Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. *Nat Neurosci*. 2016;19(4):571–577.
9. Steinberg S, Gudmundsdottir S, Sveinbjornsson G, et al. Truncating mutations in RBM12 are associated with psychosis. *Nat Genet*. 2017;49(8):1251–1254.
10. Marshall CR, Howrigan DP, Merico D, et al.; Psychosis Endophenotypes International Consortium; CNV and Schizophrenia Working Groups of the Psychiatric Genomics Consortium. Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nat Genet*. 2017;49(1):27–35.
11. Lek M, Karczewski KJ, Minikel EV, et al.; Exome Aggregation Consortium. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536(7616):285–291.
12. Singh T, Walters JTR, Johnstone M, et al.; INTERVAL Study; UK10K Consortium. The contribution of rare variants to risk of schizophrenia in individuals with and without intellectual disability. *Nat Genet*. 2017;49(8):1167–1173.
13. Suddaby JS, Silver J, So J. Understanding the schizophrenia phenotype in the first patient with the full SCN2A phenotypic spectrum. *Psychiatr Genet*. 2019;29(3):91–94.
14. Wright CF, McRae JF, Clayton S, et al.; DDD Study. Making new genetic diagnoses with old data: iterative reanalysis and reporting from genome-wide data in 1,133 families with developmental disorders. *Genet Med*. 2018;20(10):1216–1223.
15. Fromer M, Pocklington AJ, Kavanagh DH, et al. De novo mutations in schizophrenia implicate synaptic networks. *Nature*. 2014;506(7487):179–184.
16. Rees E, Kirov G, Walters JT, et al.; Taiwanese Trios Exome Sequencing Consortium. Analysis of exome sequence in 604 trios for recessive genotypes in schizophrenia. *Transl Psychiatry*. 2015;5:e607.
17. Faust GG, Hall IM. SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics*. 2014;30(17):2503–2505.
18. McLaren W, Gil L, Hunt SE, et al. The ensembl variant effect predictor. *Genome Biol*. 2016;17(1):122.
19. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet*. 2013;Chapter 7:Unit7.20.
20. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*. 2009;4(7):1073–1081.
21. Curtis D. A rapid method for combined analysis of common and rare variants at the level of a region, gene, or pathway. *Adv Appl Bioinform Chem*. 2012;5:1–9.
22. Curtis D; UK10K Consortium. Practical experience of the application of a weighted burden test to whole exome sequence data for obesity and schizophrenia. *Ann Hum Genet*. 2016;80(1):38–49.
23. Curtis D. A weighted burden test using logistic regression for integrated analysis of sequence variants, copy number variants and polygenic risk score. *Eur J Hum Genet*. 2019;27(1):114–124.
24. Landrum MJ, Lee JM, Benson M, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res*. 2018;46(D1):D1062–D1067.

25. Stelzer G, Rosen N, Plaschkes I, et al. The genecards suite: from gene data mining to disease genome sequence analyses. *Curr Protoc Bioinformatics*. 2016;54:1.30.1–1.30.33.
26. Kirov G, Pocklington AJ, Holmans P, et al. De novo CNV analysis implicates specific abnormalities of postsynaptic signaling complexes in the pathogenesis of schizophrenia. *Mol Psychiatry*. 2012;17(2):142–153.
27. Takata A, Xu B, Ionita-Laza I, Roos JL, Gogos JA, Karayiorgou M. Loss-of-function variants in schizophrenia risk and SETD1A as a candidate susceptibility gene. *Neuron*. 2014;82(4):773–780.
28. 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.
29. Tsurusaki Y, Okamoto N, Ohashi H, et al. Mutations affecting components of the SWI/SNF complex cause Coffin-Siris syndrome. *Nat Genet*. 2012;44(4):376–378.
30. Bienengraeber M, Olson TM, Selivanov VA, et al. ABCC9 mutations identified in human dilated cardiomyopathy disrupt catalytic KATP channel gating. *Nat Genet*. 2004;36(4):382–387.
31. Hoischen A, van Bon BW, Rodríguez-Santiago B, et al. De novo nonsense mutations in ASXL1 cause Bohring-Opitz syndrome. *Nat Genet*. 2011;43(8):729–731.
32. Shimada S, Oguni H, Otani Y, et al. An episode of acute encephalopathy with biphasic seizures and late reduced diffusion followed by hemiplegia and intractable epilepsy observed in a patient with a novel frameshift mutation in HNRNPU. *Brain Dev*. 2018;40(9):813–818.
33. Ishii A, Shioda M, Okumura A, et al. A recurrent KCNT1 mutation in two sporadic cases with malignant migrating partial seizures in infancy. *Gene*. 2013;531(2):467–471.
34. Beaussart ML, Barbarot S, Mauger C, Roy A. Systematic review and meta-analysis of executive functions in preschool and school-age children with neurofibromatosis type 1. *J Int Neuropsychol Soc*. 2018;24(9):977–994.
35. Briet C, Pereda A, Le Stunff C, et al. Mutations causing acrodysostosis-2 facilitate activation of phosphodiesterase 4D3. *Hum Mol Genet*. 2017;26(20):3883–3894.
36. Michot C, Le Goff C, Blair E, et al. Expanding the phenotypic spectrum of variants in PDE4D/PRKAR1A: from acrodysostosis to acroscaphodysplasia. *Eur J Hum Genet*. 2018;26(11):1611–1622.
37. Derwińska K, Smyk M, Cooper ML, Bader P, Cheung SW, Stankiewicz P. PTCH1 duplication in a family with microcephaly and mild developmental delay. *Eur J Hum Genet*. 2009;17(2):267–271.
38. Narumi S, Amano N, Ishii T, et al. SAMD9 mutations cause a novel multisystem disorder, MIRAGE syndrome, and are associated with loss of chromosome 7. *Nat Genet*. 2016;48(7):792–797.
39. Nagata Y, Narumi S, Guan Y, et al. Germline loss-of-function SAMD9 and SAMD9L alterations in adult myelodysplastic syndromes. *Blood*. 2018;132(21):2309–2313.
40. Van Houdt JK, Nowakowska BA, Sousa SB, et al. Heterozygous missense mutations in SMARCA2 cause Nicolaides-Baraitser syndrome. *Nat Genet*. 2012;44(4):445–9, S1.
41. Koga M, Ishiguro H, Yazaki S, et al. Involvement of SMARCA2/BRM in the SWI/SNF chromatin-remodeling complex in schizophrenia. *Hum Mol Genet*. 2009;18(13):2483–2494.
42. Oldmeadow C, Mossman D, Evans TJ, et al. Combined analysis of exon splicing and genome wide polymorphism data predict schizophrenia risk loci. *J Psychiatr Res*. 2014;52:44–49.
43. Leonenko G, Richards AL, Walters JT, et al. Mutation intolerant genes and targets of FMRP are enriched for nonsynonymous alleles in schizophrenia. *Am J Med Genet B Neuropsychiatr Genet*. 2017;174(7):724–731.
44. Al Eissa M, Sharp S, O'Brien N, et al. Genetic association and functional characterization MCPH1 gene variation in bipolar disorder and schizophrenia. *Am J Med Genet Part B Neuropsychiatr Genet*. 2019;180(4):258–265.
45. Flanagan M, Cunniff CM. Bloom syndrome. In: Adam MP, Ardinger HH, Pagon RA, et al., eds. *GeneReviews*. University of Washington, Seattle, WA. 1993–2019.
46. Trakadis YJ, Fulginiti V, Walterfang M. Inborn errors of metabolism associated with psychosis: literature review and case-control study using exome data from 5090 adult individuals. *J Inherit Metab Dis*. 2018;41(4):613–621.