American Journal of Medical Genetics Part B: Neuropsychiatric Genetics

WILEY

RESEARCH ARTICLE OPEN ACCESS

Contribution of Rare and Potentially Functionally Relevant Sequence Variants in Schizophrenia Risk-Locus Xq28,distal

I. Claus[1] | S. Sivalingam[2,3,4,5] | A. C. Koller[1] | A. Weiß[1] | C. M. Mathey[1] | L. Sindermann[1] | D. Klein[1] | L. Henschel[1,6] | K. U. Ludwig[1] | P. Hoffmann[1,7,8] | A. Heimbach[9] | S. Heilmann-Heimbach[1] | H. Vedder[10] | J. Kammerer-Ciernioch[10] | T. Stürmer[11] | F. Streit[12] | A. Maaser-Hecker[1,13,14,15] | I. Nenadić[16,17,18] | B. T. Baune[19,20,21] | A. M. Hartmann[22] | B. Konte[22] | I. Giegling[22] | U. Heilbronner[23] | M. Wagner[24] | A. Philipsen[24] | B. Schmidt[25] | D. Rujescu[22] | A. Buness[2,3,4] | T. G. Schulze[23,26] | M. Rietschel[12] | A. J. Forstner[1,27] | M. M. Nöthen[1] | F. Degenhardt[1,28]

[1]Institute of Human Genetics, University of Bonn, School of Medicine and University Hospital Bonn, Bonn, Germany | [2]Institute for Medical Biometry, Informatics and Epidemiology, Medical Faculty, University of Bonn, Bonn, Germany | [3]Institute for Genomic Statistics and Bioinformatics, Medical Faculty, University of Bonn, Bonn, Germany | [4]Core Unit for Bioinformatics Data Analysis, Medical Faculty, University of Bonn, Bonn, Germany | [5]Institute of Human Genetics, Medical Faculty, University Hospital of Düsseldorf, Heinrich Heine University of Düsseldorf, Düsseldorf, Germany | [6]German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany | [7]Department of Genomics, Life & Brain Center, University of Bonn, Bonn, Germany | [8]Division of Medical Genetics, University Hospital and Department of Biomedicine, University of Basel, Basel, Switzerland | [9]NGS Core Facility, Medical Faculty, University of Bonn, Bonn, Germany | [10]Psychiatric Center Nordbaden, Wiesloch, Germany | [11]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, USA | [12]Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, Mannheim, Germany | [13]Genetics and Aging Research Unit, Department of Neurology, Mass General Institute for Neurodegenerative Disease, Massachusetts General Hospital, Charlestown, Massachusetts, USA | [14]Harvard Medical School, Boston, Massachusetts, USA | [15]McCance Center for Brain Health, Massachusetts General Hospital, Boston, Massachusetts, USA | [16]Department of Psychiatry und Psychotherapy, University of Marburg, Marburg, Germany | [17]Center for Mind, Brain and Behavior (CMBB), University of Marburg, Marburg, Germany | [18]Department of Psychiatry and Psychotherapy, Jena University Hospital, Jena, Germany | [19]Department of Psychiatry, University of Münster, Münster, Germany | [20]Department of Psychiatry, Melbourne Medical School, The University of Melbourne, Central Melbourne, Australia | [21]The Florey Institute of Neuroscience and Mental Health, The University of Melbourne, Central Melbourne, Australia | [22]Department of Psychiatry and Psychotherapy, Comprehensive Center for Clinical Neurosciences and Mental Health, Medical University of Vienna, Vienna, Austria | [23]Institute of Psychiatric Phenomics and Genomics (IPPG), LMU University Hospital, LMU Munich, Munich, Germany | [24]Department of Psychiatry and Psychotherapy, University of Bonn, Bonn, Germany | [25]Institute for Medical Informatics, Biometry and Epidemiology (IMIBE), University Hospital Essen, University Duisburg-Essen, Essen, Germany | [26]Department of Psychiatry and Behavioral Sciences, SUNY Upstate Medical University, Syracuse, New York, USA | [27]Institute of Neuroscience and Medicine (INM-1), Research Center Jülich, Jülich, Germany | [28]Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, University Hospital Essen, University of Duisburg-Essen, Essen, Germany

Correspondence: M. M. Nöthen (markus.noethen@uni-bonn.de)

Received: 17 April 2024 | **Revised:** 1 September 2024 | **Accepted:** 17 September 2024

Funding: This study was supported by a SciMed-Doctoral Scholarship of the BONFOR Research Funding Program of the Faculty of Medicine, University of Bonn. T.G.S. is further supported by the Dr. Lisa Oehler Foundation (Kassel, Germany), the Bundesministerium für Bildung und Forschung (BMBF, Federal Ministry of Education and Research; projects: IntegraMent (01ZX1614K), BipoLife (01EE1404H), e:Med Program (01ZX1614K)), and the European Union's Horizon 2020 Research and Innovation Programme. The Heidelberg Cohort Study of the Elderly (represented by T.S.) was supported with the following grants from the German Research Foundation: STU 235/10-2, HE 2443/8-1 (DNA-extraction from 4000 mouthwash samples of the Heidelberg Cohort Study on Psychology and Chronic Disease), and AM 37/19-1 (Heidelberg Cohort Study on Diagnosis and Prognosis of Chronic Diseases).

Keywords: burden analysis | copy number variation | rare variants | schizophrenia | X-chromosome

M. M. Nöthen and F. Degenhardt contributed equally to this work.

ABSTRACT

Duplications of the Xq28,distal locus have been described in male and female patients with schizophrenia (SCZ) or intellectual disability. The Xq28,distal locus spans eight protein-coding genes (*F8*, *CMC4*, *MTCP1*, *BRCC3*, *VBP1*, *FUNDC2*, *CLIC2*, and *RAB39B*) and is flanked by recurrent genomic breakpoints. Thus, the issue of which gene/s at this locus is/are relevant in terms of SCZ pathogenesis remains unclear. The aim of this study was to investigate the contribution of rare and potentially functionally relevant sequence variants within the Xq28,distal locus to SCZ risk using the single-molecule molecular inversion probes (smMIP) method. Targeted sequencing was performed in a cohort of 1935 patients with SCZ and 1905 controls of European ancestry. The consecutive statistical analysis addressed two main areas. On the level of the individual variants, allele counts in the patient and control cohort were systematically compared with a Fisher's exact test: (i) for the entire present study cohort; (ii) for patients and controls separated by sex; and (iii) in combination with data published by the Schizophrenia Exome Meta-Analysis (SCHEMA) consortium. On the gene-wise level, a burden analysis was performed using the X-chromosomal model of the Optimal Unified Sequence Kernel Association Test (SKAT-O), with adjustment for possible sex-specific effects. Targeted sequencing identified a total of 13 rare and potentially functional variants in four patients and 11 controls. However, neither at the level of individual rare and potentially functional variants nor at the level of the eight protein-coding genes at the Xq28,distal locus was a statistically significant enrichment in patients compared to controls observed. Although inconclusive, the present findings represent a step toward improved understanding of the contribution of X-chromosomal risk factors in neuropsychiatric disorder development, which is an underrepresented aspect of genetic studies in this field.

1 | Introduction

Schizophrenia (SCZ) is a severe neuropsychiatric disorder with a lifetime prevalence of around 1% and an estimated heritability of 60%–80% (Owen, Sawa, and Mortensen 2016). Clinically, the characteristic symptoms of SCZ include: (1) positive symptoms, such as hallucinations, delusions, and disorganized speech and behavior; (2) negative symptoms, such as affective flattening and social as well as emotional withdrawal; and (3) cognitive symptoms such as impaired memory or poor executive functioning (American Psychiatric Association 2000). SCZ typically manifests in young adults and is associated with a high personal and socioeconomic burden of disease (Solmi et al. 2023), as well as an increased suicide rate and a reduced life expectancy (Plana-Ripoll et al. 2019). Sex-specific differences are well-established and have been reported for various aspects of the disease, in particular age of onset, clinical course, and prognosis. Overall, male SCZ patients tend to have an earlier disease onset, a less favorable clinical course, and a generally poorer prognosis (Gaebel and Wölwer 2010). Despite potential clinical implications (González-Rodríguez et al. 2020; Seeman 2021), these sex-specific differences have scarcely been elaborated in recent clinical practice guidelines (American Psychiatric Association 2020; Deutsche Gesellschaft für Psychiatrie und Psychotherapie Psychosomatik und Nervenheilkunde 2019), partly because their biological basis still needs to be elucidated. One hypothesis is that genetic risk factors on the X-chromosome play a role in the complex interplay between environmental and genetic factors that leads to SCZ development (Bache and DeLisi 2018).

Research has established that SCZ is a multifactorial and highly polygenic disorder (Marshall et al. 2017; Singh et al. 2022; Trubetskoy et al. 2022). To date, thousands of risk variants have been identified in hundreds of different genes (Owen et al. 2023).

Mainly secondary to technical and methodological limitations, neither sex-dependent effects (Blokland et al. 2022) nor sex-specific genetic risk factors located on the X-chromosome have been the focus of previous large-scale genetic studies (Khramtsova, Davis, and Stranger 2018).

Improving bioinformatic analysis tools and genotyping/sequencing methods now allow the analysis of genetic risk factors located on the X-chromosome (Marshall et al. 2017; Singh et al. 2022; Trubetskoy et al. 2022).

In 2017, the first potential risk-associated copy number variant (CNV) on the X-chromosome was identified in 18/21094 patients with SCZ and 2/20227 controls (Marshall et al. 2017). The authors found duplications of the Xq28,distal locus span eight protein-coding genes and confer SCZ risk in both sexes. The overall reported odds ratio (OR) was 8.9, with a greater effect size being observed in males (OR females = 6.3; OR males = ∞) (Marshall et al. 2017). All Xq28,distal duplication carriers reported in this study share the same genomic breakpoints. Hence, establishing which gene/sets of genes at this locus confers SCZ risk is challenging. Following up on the findings of Marshall et al., additional genetic evidence from independent studies might pinpoint the disease-relevant gene/set of genes.

Previous studies have suggested that gene-dosage effects contribute to SCZ pathogenesis and have highlighted several examples in which both an increased and a decreased gene dosage confers risk for SCZ, such as the 1q21.1 or 7q11.21(*ZNF92*) regions (Marshall et al. 2017; Rees et al. 2016). For duplications, research has shown that the pathomechanism can be mediated by either a genuine overexpression of candidate genes or haploinsufficiency (e.g., due to the disruption of a dosage-sensitive gene) (Rice and McLysaght 2017).

Even more specifically, research has shown that both duplications and rare protein-truncating variants within the same gene contribute to SCZ risk (Marshall et al. 2017; Singh et al. 2022). For example, while an association with increased SCZ risk has been reported for rare duplications affecting the gene *RB1CC1* (Degenhardt et al. 2013; Marshall et al. 2017), one of the largest whole exome sequencing studies of SCZ to

date also identified *RB1CC1* as 1 of only 10 genes in which rare, protein-truncating variants achieved exome-wide significance (Singh et al. 2022).

To determine which gene/sets of genes within the Xq28 locus contribute/s to SCZ risk, the present study investigated rare and potentially functional sequence variants in a region previously implicated by a recurrent duplication, under the premise that both could ultimately result in a reduction in gene expression. All eight protein-coding genes within the Xq28,distal locus were sequenced in a SCZ case–control cohort (1935 patients with SCZ and 1905 controls) using the single-molecule molecular inversion probes (smMIP) method (Hiatt et al. 2013; O'Roak et al. 2012). We hypothesized that phenotype-relevant genes will be implicated in SCZ development by an enrichment of rare and potentially functionally relevant variants in patients compared to controls.

2 | Materials and Methods

2.1 | Sample Description

The present study was performed in accordance with the principles of the Declaration of Helsinki. The study was approved by the respective institutional ethics committees, and all participants provided written informed consent prior to inclusion.

Patients with SCZ were recruited at multiple clinical centers across Germany. These comprised the Departments of Psychiatry at the Universities of Munich (Budde et al. 2019), Münster, Jena, and at the Central Institute of Mental Health in Mannheim and its collaborating psychiatric hospitals. Diagnoses were assigned following a structured clinical interview (First et al. 1996; First 1997), as conducted by a trained rater (Budde et al. 2019; Giegling et al. 2020), in accordance with the criteria of the Diagnostic and Statistical Manual of Mental Disorders (DSM)-IV (American Psychiatric Association 2000). Participants with an intellectual disability were systematically excluded from the present patient cohorts.

Controls were recruited at different sites across Germany. The majority were drawn from the Heidelberg Cohort Study of the Elderly (Amelang, Hasselbach, and Stürmer 2004; Stürmer, Hasselbach, and Amelang 2006) and the Heinz Nixdorf Recall study (Schmermund et al. 2006). Participants of these population-based studies did not undergo systematic screening for concomitant psychiatric disorders or premorbid cognitive functioning. In total, the present study included 1935 patients with SCZ (41% females and 59% males) and 1905 controls (44% females and 56% males).

2.2 | Selection of Candidate Genes

Targeted sequencing was performed for the eight protein-coding genes *F8*, *CMC4*, *MTCP1*, *BRCC3*, *VBP1*, *FUNDC2*, *CLIC2*, and *RAB39B* that are spanned by the duplications in Xq28,distal (Marshall et al. 2017) (chrX:153,800,000-154,225,000, NCBI36/hg18; chrX:154,146,806-154,571,806, GRCh37/hg19; https://genome.ucsc.edu/cgi-bin/hgLiftOver).

2.3 | Design of the smMIPs

For the eight genes, a total of 111 smMIPs were designed using the open-source software MIPgen (Boyle et al. 2014). These covered all 61 coding exons (±5 base pairs (bp)), and the capture size ranged from 210 to230 bp. All smMIPs were generated using the Genome Reference Consortium Human Build 37 (GRCh37/hg19) (Church et al. 2011), and correct alignment to the genomic target regions was visually verified using the UCSC Genome Browser (Kent et al. 2002). The smMIP oligonucleotides were obtained from Integrated DNA Technologies (IDT, Leuven, Belgium). For each smMIP, coverage was evaluated in a balancing step, which was conducted on a set of six samples using an Illumina MiSeq v2 nano kit (Illumina, San Diego, CA, USA). If required, the concentration of the corresponding smMIP was adjusted as a function of the mean coverage within this test run. The individual primer sequences of the smMIPs used in the present study are listed in Table S1.

2.4 | Multiplex Targeted Sequencing of the Xq28,distal Locus

DNA was extracted from whole venous blood. All subsequent genetic analyses were performed at the University of Bonn, Germany. Library preparation was performed on the basis of recommendations made in a previous publication (Eijkelenboom et al. 2016) and as described elsewhere (Mathey et al. 2022; Thieme et al. 2021). Sequencing experiments were performed on an Illumina HiSeq 2500 (Illumina, San Diego, CA, USA), in accordance with the manufacturer's instructions.

2.5 | Data Processing

Sequencing data were processed using an in-house pipeline (Mathey et al. 2022; Thieme et al. 2021), as based on the workflow developed by Hiatt and O'Roak (Hiatt et al. 2013; O'Roak et al. 2012), and was performed in accordance with the best practice guidelines of the Genome Analysis Tool Kit (GATK) (Van der Auwera et al. 2013). Briefly, collapsed reads were aligned to the reference genome using Burrows-Wheeler Aligner (BWA) (Li and Durbin 2009).

Variants were called using the GATK UnifiedGenotyper (Van der Auwera et al. 2013), and were annotated using Annovar (Wang, Li, and Hakonarson 2010) and the Combined Annotation Dependent Depletion (CADD) score v1.6 (Rentzsch et al. 2021).

2.6 | Quality Control

Quality control was performed in two steps. First, preexisting genotyping data of the study participants generated in the context of previous studies (Amelang, Hasselbach, and Stürmer 2004; Budde et al. 2019; Schmermund et al. 2006; Trubetskoy et al. 2022) were used to investigate population structure, sex mismatches, and relatedness using PLINK (Purcell et al. 2007) and KING (Manichaikul et al. 2010). To assess the population structure of the present case–control study, a multidimensional scaling analysis (MDS)

was performed to detect outliers of non-European ancestry. Samples were excluded if they deviated by more than three standard deviations (±3SD) from the mean distance. The present samples were projected against the Phase 3 data of the 1000 Genomes project (Auton et al. 2015). This generated a homogenous cluster, which demonstrated substantial overlap between the respective European samples.

Further samples were removed if a mismatch between reported and genotyped sex, or a kinship coefficient > 0.088, was found.

Second, the smMIP sequencing data were used to evaluate the overall mean coverage of each sample.

To be included in the downstream analyses, samples were required to have a coverage of ≥10X for at least 90% of the targeted bases, as analyzed using Picard (Broad Institute 2019) (see Figure S1).

In addition, the heterozygous/homozygous and transversion/transition ratios of the samples were calculated. All samples that deviated by more than ±3SD from the mean of the targeted sequencing data, as obtained by BCFTools (Li 2011), were excluded. Next, the GATK hard filter criteria were applied (Van der Auwera et al. 2013).

2.7 | Variant Filtering

To identify rare and potentially functional variants, the analyses focused on non-synonymous (missense, stopgain, stoploss, startloss, and splicing) single-nucleotide variants and small insertions and deletions (InDels, including frameshift and non-frameshift) with a minor allele frequency (MAF) ≤1% in the present study cohort and according to the frequency data for the non-psych, non-Finnish Europeans provided by the Exome Aggregation Consortium (ExAC) (Lek et al. 2016), as based on previous publications (Rees et al. 2019; Wang et al. 2020).

To be eligible for inclusion in the downstream burden analysis, all remaining variants were required to fulfill all three of the following criteria:

i. located within an exon (±5 bp) of a candidate gene;
ii. MAF ≤0.1% in the present study cohort and according to the frequency for the non-psych, non-Finnish Europeans provided by the ExAC data (Lek et al. 2016); and
iii. CADD score ≥20 (Rentzsch et al. 2021). In the case of indels, CADD scores are not provided. Therefore, the respective variants were included in the downstream analysis under the assumption that they are functionally—and in consequence, biologically—relevant.

While this variant filtering strategy was selected to minimize the number of false-positive variants and applied a commonly used variant prediction tool, it may still be biased toward deleterious loss-of-function variants, and its use might lead to an underrepresentation of, for example, more common gain-of-function variants.

2.8 | Verification of Identified Variants

To exclude potential false-positive findings, all variants subjected to the consecutive burden analysis were visually inspected using the Integrative Genomics Viewer Version 2.16.1 (Robinson et al. 2011) and were subjected to Sanger sequencing in addition to rigorous quality control.

Sanger sequencing was performed on an ABI Sequencer 3500 XL platform (Applied Biosystems, Waltham, MA, USA), and the generated sequences were analyzed using the JSI Sequence Pilot software Version 5.3.4 (JSI Medical Systems, Ettenheim, Germany). Further information on the primer sequences is available upon request.

2.9 | Analysis on the Single Variant Level

The single variant level analysis involved two steps. First, we filtered for rare variants with an MAF threshold of ≤1% and applied a two-sided Fisher's exact test to compare the absolute counts of heterozygous and hemizygous calls within both the entire patient and control group.

In a second step, we focused on rarer (MAF ≤0.1%) and potentially functional variants, that were subsequently included in the burden analysis. Here, a similar analysis was performed (i) for patients and controls separated by sex and (ii) combining the allele counts for the specific variants identified in the present study and data published by the SCHEMA Consortium (Singh et al. 2022).

Since the summary statistics of the SCHEMA browser (https://schema.broadinstitute.org/) contained no information on sex, the absolute allele counts of both data sets were totaled and then compared between patients and controls using the two-sided Fisher's exact test. The sex ratio of the case group (37.1% females and 62.9% males; personal communication Tarjinder Singh) was taken as a basis for the overall SCHEMA sample (24,248 patients with SCZ and 97,322 controls) in order to calculate the total number of alleles for the X-chromosome.

2.10 | Gene-Wise Burden Analysis

To focus on and test for an increased burden of rare and functionally relevant variants in the present case–control cohort, a gene-wise burden analysis was performed using the SKAT-O test (Lee, Wu, and Lin 2012; Lee et al. 2012), with sex as a covariate.

Within the SKAT-O framework, a specific statistical model for the X-chromosome that considers X-inactivation in female individuals was applied (Ma et al. 2015). This model also emphasizes the underlying biology and statistical power in light of different female-to-male-ratios. To correct for multiple testing, adjusted p-values for the SKAT-O test were calculated using the Benjamini–Hochberg method (Benjamini and Hochberg 1995). For each of the candidate genes, $p < 0.05$ was considered nominally significant.

All statistical analyses were performed in R v3.6.3 (R Core Team 2022).

3 | Results

3.1 | Data Processing, Quality Control, and Variant Filtering

After quality control, the initial dataset comprised 71 different exonic variants in 3383 samples (1757 patients with SCZ and 1626 controls).

In total, 69 of these variants were rare according to the initial MAF threshold of ≤1% and 63 variants were even rarer according to the MAF threshold (≤0.1%) used in the consecutive burden analysis. Of these, 13 different variants in the genes *F8*, *BRCC3*, *VBP1*, *FUNDC2*, and *RAB39B*, as identified in 18 individuals, were classified as potentially functional according to the CADD score threshold (≥20).

3.2 | Verification of Identified Variants

For four of five genes carrying a potentially rare and functionally relevant variant, all observations were confirmed via Sanger sequencing. In the gene *BRCC3*, one specific variant was called in five different individuals. Upon visual inspection, in three of these five individuals, this variant appeared to be a technical artifact. However, all five samples were included in the Sanger sequencing step. As expected from the visual inspection, the variant was only verified in two individuals, which suggests that the others were indeed likely to have represented technical artifacts. In summary, after quality control, filtering, and Sanger validation, a total of 13 different variants in five genes were found in four patients with SCZ and 11 controls. Table 1 lists all validated variants that were subjected to the consecutive burden analysis; together with more detailed information such as allele counts, as obtained from the SCHEMA browser.

3.3 | Analysis on the Single Variant Level

None of the 69 variants tested in the single-variant analysis with an MAF threshold of ≤1% were (nominally) associated with SCZ risk across the entire study cohort, as determined by Fisher's exact test. Individual results for each of the investigated variants are displayed in Table S2.

The majority of the rare (MAF ≤0.1%) and potentially functional variants (11/13) subjected to the burden analysis were detected as singletons, thus confirming their individual rarity. No significant enrichment was found in either the sex-specific analysis, or in the combined analysis that systematically added the allele counts provided by the SCHEMA Consortium to the data of the present case–control cohort. On the single-variant level, the lowest *p*-value for these variants was 0.209, as found in the combined analysis of the variants *F8*: c.599A>G;p.(Glu200Gly) and *FUNDC2*:c.85C>T;p.(Arg29Cys).

3.4 | Gene-Wise Burden Analysis

The SKAT-O test using the X-chromosomal model and sex as a covariate revealed no significant gene-wise association (all $p > 0.05$). However, the lowest *p*-value was observed for *BRCC3* ($p = 0.103$, *p*-adjusted = 0.515), which encodes BRCA1/BRCA2-Containing Complex Subunit 3. As the SKAT-O test considers several directions of a causal effect, this might indicate a potential protective effect. In the genes *CLIC2*, *MTCP1*, and *CMC4*, no rare and potentially functional variant could be identified with the present strict quality control and filter criteria. The respective genes could therefore not be included in the statistical analysis.

4 | Discussion

The present study involved a comprehensive analysis of the contribution of rare and potentially functionally relevant sequence variants within the first X-chromosomal region that was previously implicated as a genetic risk factor in the most recent large-scale CNV analysis of SCZ (Marshall et al. 2017). Although performed in a SCZ case–control cohort with a substantial number of participants and consideration of possible sex-specific effects, none of the individual rare and potentially functional variants identified or any of the eight protein-coding genes of the Xq28,distal locus showed statistically significant enrichment in patients compared to controls.

These results and information from current datasets, such as the latest meta-analysis of exome sequencing data in SCZ published by the SCHEMA Consortium (Singh et al. 2022), suggest that there is no strong evidence for the hypothesis that rare and potentially functional sequence variants in the Xq28,distal locus contribute to the risk for SCZ development.

However, clinical evidence for the potential involvement of this region in neuropsychiatric disease is available. Equivalent ~0.5 Mb duplications of the Xq28,distal locus (genomic positions according to GRCh37/hg19 coordinates: chrX:154,146,806-154,571,806 (Marshall et al. 2017) and chrX:154,100,000 to 154,600,000 (El-Hattab et al. 2011) respectively) have been associated with a rare neurodevelopmental syndrome comprising a variable degree of X-linked intellectual disability, dysmorphic facial features, and a neurobehavioral phenotype (Ballout et al. 2020; Ballout and El-Hattab 2021; El-Hattab et al. 2011, 2015; Vanmarsenille et al. 2014).

Among the 35 clinically characterized duplication carriers investigated in the context of intellectual disability to date, 16 displayed neuropsychiatric symptoms, ranging from autism spectrum features to the development of psychosis (Ballout et al. 2020; El-Hattab et al. 2011, 2015). In general, these tended to manifest more severely in hemizygous males compared to heterozygous females.

In contrast, to date, the reciprocal deletion has only been described in female carriers, and thus several authors have suggested a recessive embryonic lethality in males (El-Hattab et al. 2015; Marshall et al. 2017).

In the literature, both a further distal overlapping duplication encompassing *RAB39B* and *CLIC2*, as well as pathogenic missense variants within these two genes, have been associated with X-linked intellectual disability (Andersen et al. 2014; Giannandrea et al. 2010; Takano et al. 2012). This neuropsychiatric

TABLE 1 | Rare and potentially functionally relevant variants included in the burden analysis of the Xq28,distal locus.

HGVS	Internal AC (patients/controls)	Sex (female/male)	AC SCHEMA (patients/controls)	CADD v1.6	Internal MAF	MAF ExAC	VEP impact
F8* (*NM_000132*): *7 variants							
c.599A > G; p.(Glu200Gly)	1/0	m	—	25.0	0.00030	—	Moderate
c.524C > T; p.(Thr175Ile)	0/1	f	1/3	24.6	0.00015	0.000070	Moderate
c.5593G > A; p.(Asp1865Asn)	0/1	f	—	26.4	0.00015	—	Moderate
c.2083G > A; p.(Glu695Lys)	0/1	m	—	26.2	0.00030	—	Moderate
c.1724A > G; p.(Lys575Arg)	0/1	f	—	25.3	0.00015	—	Moderate
c.461A > C; p.(Thr154Lys)	0/1	m	—	24.2	0.00030	—	Moderate
c.90A > C; p.(Glu30Asp)	0/1	f	—	23.4	0.00015	—	Moderate
FUNDC2* (*NM_023934*): *3 variants							
c.85C > T; p.(Arg29Cys)	1/0	m	—	22.9	0.00030	0.000083	Moderate
c.211G > A; p.(Gly71Arg)	0/1	m	1/0	23.9	0.00030	0.000033	Moderate
c.487_489del; p.(Glu164del)	0/1	f	—	—	0.00015	—	Moderate
BRCC3* (*NM_001018055*): *1 variant							
c.27_35del; p.(Gln10_Val12del)	0/2	f/m	—	—	0.00089	0.000700	Moderate
VBP1* (*NM_001303543*): *1 variant							
c.577T > A; p.(Ser193Thr)	1/1	m/m	1/4	20.2	0.00059	—	Moderate
RAB39B* (*NM_171998*): *1 variant							
c.627_629del; p.(Arg210del)	1/0	m	0/5	—	0.00030	—	Moderate

Note: This table lists all variants ($n = 13$) that were subjected to the burden analysis of the Xq28,distal locus after strict quality control and filtering measures. AC SCHEMA, absolute allele counts retrieved from the SCHEMA web browser (24,248 patients with schizophrenia and 97,322 controls) (Singh et al. 2022); CADD v 1.6, phred-scaled Combined Annotation Dependent Depletion (CADD) score v1.6 (Rentzsch et al. 2021); f, female; HGVS; variant description according to the nomenclature of the Human Genome Variation Society (Den Dunnen et al. 2016); internal AC, absolute allele counts in the present case–control cohort; internal MAF, minor allele frequency in the present combined case–control cohort (1757 patients with schizophrenia and 1626 controls); m, male; MAF ExAC, frequencies for the non-psychiatric, non-Finnish Europeans provided by the Exome Aggregation Consortium (ExAC) (Lek et al. 2016); VEP Impact, impact predicted by the Ensembl Variant Effect Predictor (McLaren et al. 2016).

phenotype is similar to that of clinically characterized carriers of a duplication at Xq28,distal (Ballout et al. 2020; Ballout and El-Hattab 2021; El-Hattab et al. 2011, 2015).

The fact that the present study identified only a small number of variants has several potential explanations. On the one hand, the low variant detection rate may have been attributable to several, primarily technical, limitations, and/or to direct consequences of the study design.

First, the filter criteria were strict. Although the CADD score is a commonly applied tool for assessing the pathogenicity of single-nucleotide variants, it has limited relevance for our specific research question, which focuses on a recurrent duplication. This is because a gain-of-function mechanism might be more consistent with the concept of a duplicated gene. With the aim of differentiating between pathogenic and benign genetic variation, CADD prioritizes variants subjected to selective pressure. In principle, CADD is non-directive and could be applied to both loss-of function and gain-of-function variants.

Nevertheless, by incorporating information from scores geared toward negative effects on protein function, such as SIFT and PolyPhen (Flanagan, Patch, and Ellard 2010), currently available metapredictors, such as CADD, are more accurate predictors of pathogenicity in an underlying loss-of-function mechanism (Sevim Bayrak et al. 2021). This also applies to other commonly used in silico prediction tools such as REVEL (Hopkins et al. 2023), while the situation for novel machine-learning-based scores, such as AlphaMissense, still awaits evaluation (Cheng et al. 2023). The latter represents a promising future approach in terms of improving understanding of different molecular disease mechanisms.

However, the fact that three observations could not be confirmed by visual inspection and Sanger sequencing provides evidence that some genomic regions remain technically challenging. This highlights the need for the replication of results, even from large cohorts, despite advances in sequencing technologies and bioinformatic methods.

Second, the sample size was limited, as shown by the efforts of the SCHEMA Consortium, which required tens of thousands of study participants to achieve exome-wide significance for disease-associated genes (Singh et al. 2022). Ultimately, functionally relevant variants within the Xq28,distal locus that contribute to SCZ risk may be ultra-rare and consequently only identifiable with certainty in even larger cohorts.

Third, the analyses were restricted to individuals of European ancestry in order to ensure a study cohort that was as genetically homogenous as possible. To assess the transferability of our findings, future studies should investigate the contribution of rare sequence variants at the Xq28,distal region to SCZ in samples with diverse ancestries. Nevertheless, this remains a key challenge in genetic studies (Derks, Thorp, and Gerring 2022). International efforts are essential to identify population-specific risk factors and address global transferability appropriately, thus ensuring equitable applicability of findings from psychiatric genetic research across diverse populations (Martin et al. 2019; Peterson et al. 2019).

Fourth, disease-relevant sequence variants could also be located within the non-coding regions of the Xq28,distal locus, which were not investigated in the present study. Whole genome sequencing in ethnically diverse cohorts would be desirable and could expand knowledge of the contribution of rare genetic variation to SCZ development across the global population.

Fifth, several factors might have contributed to a selection bias in the study design. Most prominently, participants with an intellectual disability were systematically excluded from the present patient cohorts, despite the fact that a broad phenotypical overlap between SCZ and neurodevelopmental disorders is well-established and previous phenotypic analyses of CNV carriers demonstrated a strong enrichment for premorbid cognitive abnormalities in this particular patient subgroup (Foley et al. 2020). This might have led to an underrepresentation of variants with clear evidence for pathogenicity in our data. Moreover, the present controls were not systematically screened for psychiatric disorders or premorbid cognitive functioning. Nonetheless, given that the prevalence of SCZ is around 1%, the degree of power loss secondary to the use of unscreened controls can be assumed to be low (Moskvina et al. 2005). Additionally, the key variables adjusted for in the patient and control cohort in terms of comparability were ancestry and biological sex, and did not include further demographic information such as mean age.

Next, the statistical power may have been further diminished by the inclusion of a slightly lower proportion of females with two X-chromosomes in both the patient and the control group. This was demonstrated by the developers of the SKAT-O test (Ma et al. 2015), and represents a challenge even for large-scale systematic genetic analysis in psychiatric disorders (Pirastu et al. 2021).

On the other hand, the small number of variants detected may also imply that rare and potentially functional sequence variants at Xq28,distal do indeed play a limited role in SCZ risk.

This might point toward a more general role of the Xq28,distal locus in neuronal development via alternative pathomechanisms that were not covered by the present study design.

For example, a previous study of *RAB39B* demonstrated a specific increase in gene expression in the blood lymphocytes of individual carriers of a duplication at Xq28,distal, as well as in corresponding cultures of genetically modified primary mouse hippocampal neurons (Vanmarsenille et al. 2014).

Besides a genuine overexpression of any of the implicated genes themselves, another plausible mechanism could be the disruption of topologically associated domains (TADs) (Spielmann, Lupiáñez, and Mundlos 2018), which might lead to the dysregulation of several genes—even outside the Xq28,distal locus—via the formation of new TAD boundaries, and thus to altered gene enhancer interactions. This has been exemplified previously by X-chromosomal duplications in the context of 46,XY gonadal dysgenesis (Meinel et al. 2022).

In conclusion, the results of the present study provide no evidence for an effect of rare and potentially functionally relevant sequence variants in the Xq28,distal locus on the risk of SCZ.

However, given the restricted statistical power and the technical and methodological limitations of the study, the possibility that rare sequence variants in any of the implicated genes contribute to the development of SCZ cannot be excluded.

Given its known associations with neuropsychiatric phenotypes, this region may still yield important insights. Thus, further investigation of the Xq28,distal region, and systematic investigations of X-chromosomal risk factors in general, are still warranted in genetic SCZ research.

A deeper understanding of sex-dependent and sex-specific risk factors might shed further light onto the biological mechanisms that underlie phenotypic differences between female and male patients with SCZ and pave the way for the implementation of more profound recommendations for both sexes into future evidence-based and sex-specific clinical practice guidelines.

Author Contributions

I.C. and F.D. designed the study. I.C., S.S., A.C.K., A.W., C.M.M., D.K., L.H., K.U.L., P.H., A.H., S.H.-H., H.V., J.K.-C., T.S., F.S., A.M.-H., I.N., B.T.B., A.M.H., B.K., I.G., U.H., M.W., A.P., B.S., D.R., A.B., T.G.S., M.R., A.J.F., M.M.N., and F.D. participated in data acquisition, preparation, and quality control. I.C., S.S., A.B., and F.D. performed the statistical analyses. I.C., S.S., A.B., F.D., A.J.F., and M.M.N. analyzed and interpreted the data. I.C. wrote the first draft of the manuscript; L.S., C.M.M., A.M.-H., U.H., F.D., A.J.F., and M.M.N. revised it. All authors contributed to and have approved the final manuscript.

Acknowledgments

The authors thank all patients and controls for their participation and all clinical collaborators for their cooperation with recruitment. We thank the Next Generation Sequencing Core Facility and the Core Unit for Bioinformatics Data Analysis of the University of Bonn for their contribution to the generation and statistical analysis of the sequence data. Open Access funding enabled and organized by Projekt DEAL.

Conflicts of Interest

DR served as a consultant for Janssen; received honoraria from Boehringer-Ingelheim, Gerot Lannacher, Janssen, and Pharmagenetix; received travel support from Angelini, Janssen, and Schwabe; and served on the advisory boards of AC Immune, Boehringer-Ingelheim, Roche, and Rovi. The remaining authors have no conflicts of interest to declare.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Amelang, M., P. Hasselbach, and T. Stürmer. 2004. "Personality, Cardiovascular Disease, and Cancer." *Zeitschrift Für Gesundheitspsychologie* 12, no. 3: 102–115. https://doi.org/10.1026/0943-8149.12.3.102.

American Psychiatric Association. 2000. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5-TR)*. Washington, DC: American Psychiatric Association.

American Psychiatric Association. 2020. "The American Psychiatric Association Practice Guideline for the Treatment of Patients With Schizophrenia." In *Practice Guidelines*, 3rd ed. Washington, DC: American Psychiatric Association Publishing. https://doi.org/10.1176/appi.books.9780890424841.

Andersen, E. F., E. E. Baldwin, S. Ellingwood, R. Smith, and A. N. Lamb. 2014. "Xq28 Duplication Overlapping the int22h-1/int22h-2 Region and Including RAB39B and CLIC2 in a Family With Intellectual and Developmental Disability." *American Journal of Medical Genetics, Part A* 164, no. 7: 1795–1801. https://doi.org/10.1002/ajmg.a.36524.

Auton, A., G. R. Abecasis, D. M. Altshuler, et al. 2015. "A Global Reference for Human Genetic Variation." *Nature* 526, no. 7571: 68–74. https://doi.org/10.1038/nature15393.

Bache, W. K., and L. E. DeLisi. 2018. "The Sex Chromosome Hypothesis of Schizophrenia: Alive, Dead, or Forgotten? A Commentary and Review." *Complex Psychiatry* 4, no. 2: 83–89. https://doi.org/10.1159/000491489.

Ballout, R. A., C. Dickerson, M. J. Wick, et al. 2020. "Int22h1/Int22h2-Mediated Xq28 Duplication Syndrome: de Novo Duplications, Prenatal Diagnoses, and Additional Phenotypic Features." *Human Mutation* 41, no. 7: 1238–1249. https://doi.org/10.1002/humu.24009.

Ballout, R. A., and A. W. El-Hattab. 2021. "The int22h1/int22h2-Mediated xq28 Duplication Syndrome: An Intersection Between Neurodevelopment, Immunology, and Cancer." *Genes* 12, no. 6: 860. https://doi.org/10.3390/genes12060860.

Benjamini, Y., and Y. Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society: Series B: Methodological* 57, no. 1: 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Blokland, G. A. M., J. Grove, C.-Y. Chen, et al. 2022. "Sex-Dependent Shared and Nonshared Genetic Architecture Across Mood and Psychotic Disorders." *Biological Psychiatry* 91, no. 1: 102–117. https://doi.org/10.1016/j.biopsych.2021.02.972.

Boyle, E. A., B. J. O'Roak, B. K. Martin, A. Kumar, and J. Shendure. 2014. "MIPgen: Optimized Modeling and Design of Molecular Inversion Probes for Targeted Resequencing." *Bioinformatics* 30, no. 18: 2670–2672. https://doi.org/10.1093/bioinformatics/btu353.

Broad Institute. 2019. "Picard Toolkit." https://broadinstitute.github.io/picard/.

Budde, M., H. Anderson-Schmidt, K. Gade, et al. 2019. "A Longitudinal Approach to Biological Psychiatric Research: The PsyCourse Study." *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics* 180, no. 2: 89–102. https://doi.org/10.1002/ajmg.b.32639.

Cheng, J., G. Novati, J. Pan, et al. 2023. "Accurate Proteome-Wide Missense Variant Effect Prediction With AlphaMissense." *Science* 381, no. 6664: eadg7492. https://doi.org/10.1126/science.adg7492.

Church, D. M., V. A. Schneider, T. Graves, et al. 2011. "Modernizing Reference Genome Assemblies." *PLoS Biology* 9, no. 7: 1–5. https://doi.org/10.1371/journal.pbio.1001091.

Degenhardt, F., L. Priebe, S. Meier, et al. 2013. "Duplications in RB1CC1 Are Associated With Schizophrenia; Identification in Large European Sample Sets." *Translational Psychiatry* 3, no. 11: e326. https://doi.org/10.1038/tp.2013.101.

Den Dunnen, J. T., R. Dalgleish, D. R. Maglott, et al. 2016. "HGVS Recommendations for the Description of Sequence Variants: 2016 Update." *Human Mutation* 37, no. 6: 564–569. https://doi.org/10.1002/humu.22981.

Derks, E. M., J. G. Thorp, and Z. F. Gerring. 2022. "Ten Challenges for Clinical Translation in Psychiatric Genetics." *Nature Genetics* 54, no. 10: 1457–1465. https://doi.org/10.1038/s41588-022-01174-0.

Deutsche Gesellschaft für Psychiatrie und Psychotherapie Psychosomatik und Nervenheilkunde. 2019. "*S3-Leitlinie Schizophrenie. Langfassung 2019, Version 1.0, zuletzt geändert am 15.* März 2019" DGPPN e.V. (Hrsg.) für die Leitliniengruppe. https://register.awmf.org/de/leitlinien/detail/038-009.

Eijkelenboom, A., E. J. Kamping, A. W. Kastner-van Raaij, et al. 2016. "Reliable Next-Generation Sequencing of Formalin-Fixed, Paraffin-Embedded Tissue Using Single Molecule Tags." *Journal of Molecular Diagnostics* 18, no. 6: 851–863. https://doi.org/10.1016/j.jmoldx.2016.06.010.

El-Hattab, A. W., P. Fang, W. Jin, et al. 2011. "Int22h-1/int22h-2-Mediated Xq28 Rearrangements: Intellectual Disability Associated With Duplications and In Utero Male Lethality With Deletions." *Journal of Medical Genetics* 48, no. 12: 840–850. https://doi.org/10.1136/jmedgenet-2011-100125.

El-Hattab, A. W., C. P. Schaaf, P. Fang, et al. 2015. "Clinical Characterization of int22h1/int22h2-Mediated Xq28 Duplication/Deletion: New Cases and Literature Review." *BMC Medical Genetics* 16, no. 1: 12. https://doi.org/10.1186/s12881-015-0157-2.

First, M. B. 1997. *User's Guide for the Structured Clinical Interview for DSM-IV Axis II Personality Disorders: SCID-II*. Washington DC: American Psychiatric Press.

First, M. B., R. L. Spitzer, M. Gibbon, and J. B. Williams. 1996. "Structured Clinical Interview for DSM-IV Axis I Disorders, Clinician Version (SCID-CV)." In *Encyclopedia of Behavioral Medicine*. Washington DC: American Psychiatric Press.

Flanagan, S. E., A. M. Patch, and S. Ellard. 2010. "Using SIFT and PolyPhen to Predict Loss-Of-Function and Gain-Of-Function Mutations." *Genetic Testing and Molecular Biomarkers* 14, no. 4: 533–537. https://doi.org/10.1089/gtmb.2010.0036.

Foley, C., E. A. Heron, D. Harold, et al. 2020. "Identifying Schizophrenia Patients Who Carry Pathogenic Genetic Copy Number Variants Using Standard Clinical Assessment: Retrospective Cohort Study." *British Journal of Psychiatry* 216, no. 5: 275–279. https://doi.org/10.1192/bjp.2019.262.

Gaebel, W., and W. Wölwer. 2010. *Gesundheitsberichterstattung des Bundes - Themenheft 50: "Schizophrenie"*. Berlin, Germany: Robert Koch Institut. https://doi.org/10.25646/3154.

Giannandrea, M., V. Bianchi, M. L. Mignogna, et al. 2010. "Mutations in the Small GTPase Gene RAB39B Are Responsible for X-Linked Mental Retardation Associated With Autism, Epilepsy, and Macrocephaly." *American Journal of Human Genetics* 86, no. 2: 185–195. https://doi.org/10.1016/j.ajhg.2010.01.011.

Giegling, I., A. M. Hartmann, J. Genius, et al. 2020. "Polymorphisms in CRYBB2 Encoding βB2-Crystallin Are Associated With Antisaccade Performance and Memory Function." *Translational Psychiatry* 10, no. 10: 113. https://doi.org/10.1038/s41398-020-0791-0.

González-Rodríguez, A., A. Guàrdia, A. Á. Pedrero, et al. 2020. "Women With Schizophrenia Over the Life Span: Health Promotion, Treatment and Outcomes." *International Journal of Environmental Research and Public Health* 17, no. 15: 1–13. https://doi.org/10.3390/ijerph17155594.

Hiatt, J. B., C. C. Pritchard, S. J. Salipante, B. J. O'Roak, and J. Shendure. 2013. "Single Molecule Molecular Inversion Probes for Targeted, High Accuracy Detection of Low Frequency Variation." *Genome Research* 23, no. 5: 843–854. https://doi.org/10.1101/gr.147686.112.

Hopkins, J. J., M. N. Wakeling, M. B. Johnson, S. E. Flanagan, and T. W. Laver. 2023. "REVEL Is Better at Predicting Pathogenicity of Loss-Of-Function Than Gain-Of-Function Variants." *Human Mutation* 2023: 1–6. https://doi.org/10.1155/2023/8857940.

Kent, W. J., C. W. Sugnet, T. S. Furey, et al. 2002. "The Human Genome Browser at UCSC." *Genome Research* 12, no. 6: 996–1006. https://doi.org/10.1101/gr.229102.

Khramtsova, E. A., L. K. Davis, and B. E. Stranger. 2018. "The Role of Sex in the Genomics of Human Complex Traits." *Nature Reviews Genetics* 20: 173–190. https://doi.org/10.1038/s41576-018-0083-1.

Lee, S., M. J. Emond, M. J. Bamshad, et al. 2012. "Optimal Unified Approach for Rare-Variant Association Testing With Application to Small-Sample Case-Control Whole-Exome Sequencing Studies." *American Journal of Human Genetics* 91, no. 2: 224–237. https://doi.org/10.1016/j.ajhg.2012.06.007.

Lee, S., M. C. Wu, and X. Lin. 2012. "Optimal Tests for Rare Variant Effects in Sequencing Association Studies." *Biostatistics* 13, no. 4: 762–775. https://doi.org/10.1093/biostatistics/kxs014.

Lek, M., K. J. Karczewski, E. V. Minikel, et al. 2016. "Analysis of Protein-Coding Genetic Variation in 60,706 Humans." *Nature* 536, no. 7616: 285–291. https://doi.org/10.1038/nature19057.

Li, H. 2011. "A Statistical Framework for SNP Calling, Mutation Discovery, Association Mapping and Population Genetical Parameter Estimation From Sequencing Data." *Bioinformatics* 27, no. 21: 2987–2993. https://doi.org/10.1093/bioinformatics/btr509.

Li, H., and R. Durbin. 2009. "Fast and Accurate Short Read Alignment With Burrows-Wheeler Transform." *Bioinformatics* 25, no. 14: 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

Ma, C., M. Boehnke, S. Lee, and the GoT2D investigators. 2015. "Evaluating the Calibration and Power of Three Gene-Based Association Tests of Rare Variants for the X Chromosome." *Genetic Epidemiology* 39, no. 7: 499–508. https://doi.org/10.1002/gepi.21935.Evaluating.

Manichaikul, A., J. C. Mychaleckyj, S. S. Rich, K. Daly, M. Sale, and W. M. Chen. 2010. "Robust Relationship Inference in Genome-Wide Association Studies." *Bioinformatics* 26, no. 22: 2867–2873. https://doi.org/10.1093/bioinformatics/btq559.

Marshall, C. R., D. P. Howrigan, D. Merico, et al. 2017. "Contribution of Copy Number Variants to Schizophrenia From a Genome-Wide Study of 41,321 Subjects." *Nature Genetics* 49, no. 1: 27–35. https://doi.org/10.1038/ng.3725.

Martin, A. R., M. Kanai, Y. Kamatani, Y. Okada, B. M. Neale, and M. J. Daly. 2019. "Clinical Use of Current Polygenic Risk Scores May Exacerbate Health Disparities." *Nature Genetics* 51, no. 4: 584–591. https://doi.org/10.1038/s41588-019-0379-x.

Mathey, C. M., C. Maj, A. B. Scheer, J. Fazaal, B. Wedi, and D. Wieczorek. 2022. "Molecular Genetic Screening in Patients With ACE Inhibitor / Angiotensin Receptor Blocker-Induced Angioedema to Explore the Role of Hereditary Angioedema Genes." *Frontiers in Genetics* 13. https://doi.org/10.3389/fgene.2022.914376.

McLaren, W., L. Gil, S. E. Hunt, et al. 2016. "The Ensembl Variant Effect Predictor." *Genome Biology* 17, no. 1. https://doi.org/10.1186/s13059-016-0974-4.

Meinel, J. A., V. Yumiceba, A. Künstner, et al. 2022. "Disruption of the Topologically Associated Domain at Xp21.2 Is Related to 46,XY Gonadal Dysgenesis." *Journal of Medical Genetics* 60, no. 5: 469–476. https://doi.org/10.1136/jmg-2022-108635.

Moskvina, V., P. Holmans, K. M. Schmidt, and N. Craddock. 2005. "Design of Case-Controls Studies With Unscreened Controls." *Annals of Human Genetics* 69, no. 5: 566–576. https://doi.org/10.1111/j.1529-8817.2005.00175.x.

O'Roak, B. J., L. Vives, W. Fu, et al. 2012. "Multiplex Targeted Sequencing Identifies Recurrently Mutated Genes in Autism Spectrum Disorders." *Science* 338, no. 6114: 1619–1622.

Owen, M. J., S. E. Legge, E. Rees, J. T. R. Walters, and M. C. O'Donovan. 2023. "Genomic Findings in Schizophrenia and Their Implications." *Molecular Psychiatry* 28: 3638–3647. https://doi.org/10.1038/s41380-023-02293-8.

Owen, M. J., A. Sawa, and P. B. Mortensen. 2016. "Schizophrenia." *Lancet* 388, no. 10039: 86–97. https://doi.org/10.1016/S0140-6736(15)01121-6.

Peterson, R. E., K. Kuchenbaecker, R. K. Walters, et al. 2019. "Genome-Wide Association Studies in Ancestrally Diverse Populations: Opportunities, Methods, Pitfalls, and Recommendations." *Cell* 179, no. 3: 589–603. https://doi.org/10.1016/j.cell.2019.08.051.

Pirastu, N., M. Cordioli, P. Nandakumar, et al. 2021. "Genetic Analyses Identify Widespread Sex-Differential Participation Bias." *Nature Genetics* 53, no. 5: 663–671. https://doi.org/10.1038/s41588-021-00846-7.

Plana-Ripoll, O., C. B. Pedersen, E. Agerbo, et al. 2019. "A Comprehensive Analysis of Mortality-Related Health Metrics Associated With Mental Disorders: A Nationwide, Register-Based Cohort Study." *Lancet* 394, no. 10211: 1827–1835. https://doi.org/10.1016/S0140-6736(19)32316-5.

Purcell, S., B. Neale, K. Todd-Brown, et al. 2007. "PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses." *American Journal of Human Genetics* 81, no. 3: 559–575. https://doi.org/10.1086/519795.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.r-project.org/.

Rees, E., N. Carrera, J. Morgan, et al. 2019. "Targeted Sequencing of 10,198 Samples Confirms Abnormalities in Neuronal Activity and Implicates Voltage-Gated Sodium Channels in Schizophrenia Pathogenesis." *Biological Psychiatry* 85, no. 7: 554–562. https://doi.org/10.1016/j.biopsych.2018.08.022.

Rees, E., K. Kendall, A. F. Pardiñas, et al. 2016. "Analysis of Intellectual Disability Copy Number Variants for Association With Schizophrenia Supplemental Content at Jamapsychiatry.Com." *JAMA Psychiatry* 73, no. 9: 963–969. https://doi.org/10.1001/jamapsychiatry.2016.1831.

Rentzsch, P., M. Schubach, J. Shendure, and M. Kircher. 2021. "CADD-Splice—Improving Genome-Wide Variant Effect Prediction Using Deep Learning-Derived Splice Scores." *Genome Medicine* 13, no. 1: 31. https://doi.org/10.1186/s13073-021-00835-9.

Rice, A. M., and A. McLysaght. 2017. "Dosage-Sensitive Genes in Evolution and Disease." *BMC Biology* 15, no. 1: 1–10. https://doi.org/10.1186/s12915-017-0418-y.

Robinson, J. T., H. Thorvaldsdóttir, W. Winckler, et al. 2011. "Integrative Genomics Viewer." *Nature Biotechnology* 29, no. 1: 24–26. https://doi.org/10.1038/nbt.1754.

Schmermund, A., S. Möhlenkamp, S. Berenbein, et al. 2006. "Population-Based Assessment of Subclinical Coronary Atherosclerosis Using Electron-Beam Computed Tomography." *Atherosclerosis* 185, no. 1: 177–182. https://doi.org/10.1016/j.atherosclerosis.2005.06.003.

Seeman, M. V. 2021. "Sex Differences in Schizophrenia Relevant to Clinical Care." *Expert Review of Neurotherapeutics* 21, no. 4: 443–453. https://doi.org/10.1080/14737175.2021.1898947.

Sevim Bayrak, C., D. Stein, A. Jain, et al. 2021. "Identification of Discriminative Gene-Level and Protein-Level Features Associated With Pathogenic Gain-of-Function and Loss-Of-Function Variants." *American Journal of Human Genetics* 108, no. 12: 2301–2318. https://doi.org/10.1016/j.ajhg.2021.10.007.

Singh, T., T. Poterba, D. Curtis, et al. 2022. "Rare Coding Variants in Ten Genes Confer Substantial Risk for Schizophrenia." *Nature* 604, no. 7906: 509–516. https://doi.org/10.1038/s41586-022-04556-w.

Solmi, M., G. Seitidis, D. Mavridis, et al. 2023. "Incidence, Prevalence, and Global Burden of Schizophrenia - Data, With Critical Appraisal, From the Global Burden of Disease (GBD) 2019." *Molecular Psychiatry* 28: 5319–5327. https://doi.org/10.1038/s41380-023-02138-4.

Spielmann, M., D. G. Lupiáñez, and S. Mundlos. 2018. "Structural Variation in the 3D Genome." *Nature Reviews Genetics* 19, no. 7: 453–467. https://doi.org/10.1038/s41576-018-0007-0.

Stürmer, T., P. Hasselbach, and M. Amelang. 2006. "Personality, Lifestyle, and Risk of Cardiovascular Disease and Cancer: Follow-Up of Population Based Cohort." *British Medical Journal* 332, no. 7554: 1359–1362. https://doi.org/10.1136/bmj.38833.479560.80.

Takano, K., D. Liu, P. Tarpey, et al. 2012. "An x-Linked Channelopathy With Cardiomegaly due to a CLIC2 Mutation Enhancing Ryanodine Receptor Channel Activity." *Human Molecular Genetics* 21, no. 20: 4497–4507. https://doi.org/10.1093/hmg/dds292.

Thieme, F., L. Henschel, N. L. Hammond, et al. 2021. "Extending the Allelic Spectrum at Noncoding Risk Loci of Orofacial Clefting." *Human Mutation* 42, no. 8: 1066–1078. https://doi.org/10.1002/humu.24219.

Trubetskoy, V., A. F. Pardiñas, T. Qi, et al. 2022. "Mapping Genomic Loci Implicates Genes and Synaptic Biology in Schizophrenia." *Nature* 604, no. 7906: 502–508. https://doi.org/10.1038/s41586-022-04434-5.

Van der Auwera, G. A., M. O. Carneiro, C. Hartl, et al. 2013. "From fastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline." *Current Protocols in Bioinformatics* 43: 11.10.1-11.10.33. https://doi.org/10.1002/0471250953.bi1110s43.

Vanmarsenille, L., M. Giannandrea, N. Fieremans, et al. 2014. "Increased Dosage of RAB39B Affects Neuronal Development and Could Explain the Cognitive Impairment in Male Patients With Distal Xq28 Copy Number Gains." *Human Mutation* 35, no. 3: 377–383. https://doi.org/10.1002/humu.22497.

Wang, K., M. Li, and H. Hakonarson. 2010. "ANNOVAR: Functional Annotation of Genetic Variants From High-Throughput Sequencing Data." *Nucleic Acids Research* 38, no. 16: 1–7. https://doi.org/10.1093/nar/gkq603.

Wang, T., K. Hoekzema, D. Vecchio, et al. 2020. "Large-Scale Targeted Sequencing Identifies Risk Genes for Neurodevelopmental Disorders." *Nature Communications* 11, no. 1: 1–13. https://doi.org/10.1038/s41467-020-18723-y.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.