Biological Psychiatry

Archival Report

A Regional Burden of Sequence-Level Variation in the 22q11.2 Region Influences Schizophrenia Risk and Educational Attainment

Elemi J. Breetvelt, Karel C. Smit, Jessica van Setten, Daniele Merico, Xiao Wang, Ilonca Vaartjes, Anne S. Bassett, Marco P.M. Boks, Peter Szatmari, Stephen W. Scherer, René S. Kahn, and Jacob A.S. Vorstman

ABSTRACT

BACKGROUND: Genomic loci where recurrent pathogenic copy number variants are associated with psychiatric phenotypes in the population may also be sensitive to the collective impact of multiple functional low-frequency single nucleotide variants (SNVs).

METHODS: We examined the cumulative impact of low-frequency, functional SNVs within the 22q11.2 region on schizophrenia risk in a discovery cohort and an independent replication cohort (N = 1933 and N = 11,128, respectively), as well as the impact on educational attainment (EA) in a third, independent, general population cohort (N = 2081). In the discovery and EA cohorts, SNVs were identified using genotyping arrays; in the replication cohort, whole-exome sequencing was available. For verification, we compared the regional SNV count for schizophrenia cases in the discovery cohort with a normative count distribution derived from a large population dataset (N = 26,500) using bootstrap procedures.

RESULTS: In both schizophrenia cohorts, an increased regional SNV burden (≥4 low-frequency SNVs) in the 22q11.2 region was associated with schizophrenia (discovery cohort: odds ratio = 7.48, p = .039; replication cohort: odds ratio = 1.92, p = .004). In the EA cohort, an increased regional SNV burden at 22q11.2 was associated with decreased EA (odds ratio = 4.65, p = .049). Comparing the SNV count for schizophrenia cases with a normative distribution confirmed the unique nature of the distribution for schizophrenia cases (p = .002).

CONCLUSIONS: In the general population, an increased burden of low-frequency, functional SNVs in the 22q11.2 region is associated with schizophrenia risk and a decrease in EA. These findings suggest that in addition to structural variation, a cumulative regional burden of low-frequency, functional SNVs in the 22q11.2 region can also have a relevant phenotypic impact.

https://doi.org/10.1016/j.biopsych.2021.11.019

The genetic architecture of schizophrenia is highly complex. Genetic variation contributing to its estimated heritability (64%–81%) (1–3) encompasses a spectrum of extremely rare to common variants with variable effect sizes (4,5). The cumulative effect of common and rare single nucleotide variants (SNVs) only explains a small fraction of the estimated heritability (6), suggesting a role for other types of genetic variations. Structural variation, including copy number variants (CNVs), represents one specific type of variation associated with the disease risk (7,8). A subset of recurrent CNVs shows a strong association with neurodevelopmental phenotypes, including schizophrenia (7,9–11), intellectual disability (ID), and autism spectrum disorder (12–14).

The observation that some CNVs are pathogenic while many others can occur with no apparent phenotypic consequences indicates that genomic regions differ with regard to the likelihood of phenotypic impact in the case of dosage change. At the same time, the neurodevelopmental phenotypes, e.g., schizophrenia or ID, associated with these large-effect structural variants are thought to be the result of the involvement of several genes affected by the CNV. Hence, a single deleterious single nucleotide variant (SNV) affecting only 1 gene at that locus typically does not have the same phenotypic consequences. Nevertheless, such genomic regions may be more liable than other regions to display phenotypic effects when affected by multiple SNVs (Figure 1). The question addressed by this study is whether a burden of SNVs in regions where pathogenic CNVs recurrently occur can, cumulatively, explain some of the genetic risk for neurodevelopmental outcomes.

We will examine this question with regard to the 22q11.2 region. CNVs recurrently occur in this region in the population, with the largest phenotypic impact exerted by the deletion associated with 22q11.2 deletion syndrome (22q11DS). With a prevalence estimated in the range of 1 in 3000 to 6000 live births (15), phenotypic manifestations of

SEE COMMENTARY ON PAGE 692

ISSN: 0006-3223

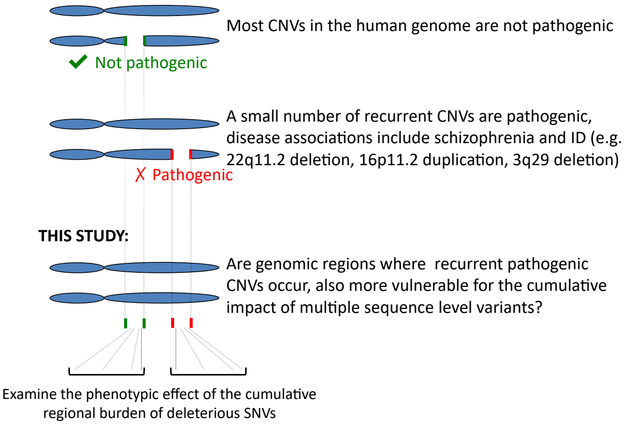
Most CNVs in the human genome are not pathogenic
✓ Not pathogenic
A small number of recurrent CNVs are pathogenic, disease associations include schizophrenia and ID (e.g. 22q11.2 deletion, 16p11.2 duplication, 3q29 deletion)
✗ Pathogenic
THIS STUDY:
Are genomic regions where recurrent pathogenic CNVs occur, also more vulnerable for the cumulative impact of multiple sequence level variants?
Examine the phenotypic effect of the cumulative regional burden of deleterious SNVs

Figure 1. Schematic overview of the research question. CNV, copy number variant; ID, intellectual disability; SNV, single nucleotide variant.

22q11DS are highly variable and can affect multiple organ systems (16). Among others, the syndrome is strongly and consistently associated with an increased risk for both schizophrenia and other neurodevelopmental disorders, including ID (10,17).

Findings thus far suggest that a sequence-level disruption of a single gene within the 22q11.2 region is, in and of itself, not a strong risk factor for schizophrenia or ID. Accordingly, a recent study showed that the risk for schizophrenia in individuals with 22q11DS was not modified by SNVs in the remaining allele of the 22q11.2 region (18). Furthermore, genome-wide association study findings in the general population have not revealed any signal exerted by common risk variants in the 22q11.2 region (3). Taken together, these observations suggest that if deleterious SNVs in the 22q11.2 region contribute to the risk of schizophrenia or ID, then this would unlikely be the result of a single deleterious SNV. Therefore, we hypothesize that such effects may arise when several SNVs co-occur in this region, thereby mimicking the impact by a CNV in the same region. Hence, the cumulative burden of several SNVs in the 22q11.2 region may be of relevance.

Schizophrenia and ID, common neurodevelopmental phenotypes associated with the 22q11.2 deletion, do not occur entirely independently. Both low childhood cognitive level (19–22) and a decline thereof (22,23) are associated with the risk of developing schizophrenia in individuals with 22q11DS, echoing similar observations in the general population (24–26). In addition, studies are starting to reveal substantial overlap between genetic variants associated with both phenotypes in the general population (27–29) and in 22q11DS (30). In the light of this growing body of evidence indicating a strong connection between academic/cognitive performance and schizophrenia (31), we examined both phenotypes in relation to the hypothesized phenotypic impact of a cumulative burden of putatively deleterious SNVs in the 22q11.2 region.

Figure 2 provides a schematic overview of this study. First, we examined the impact of the burden of SNVs in the 22q11.2 region on schizophrenia status in a schizophrenia case-control sample of 1933 individuals (discovery cohort). Second, we sought to replicate our findings in an independent schizophrenia case-control cohort (N = 11,128) while also extending the analysis from exome chip–based data to exome sequencing data (replication cohort). Third, in an independent population cohort (N = 2081), we examined whether the cumulative effect of multiple SNVs in the 22q11.2 region was also associated with decreased levels of educational attainment (EA) (validation cohort). Fourth, we conducted several post hoc analyses to further examine our observations.

METHODS AND MATERIALS

Study Populations

Discovery Cohort. The discovery cohort (N = 1933) consisted of 1002 patients with schizophrenia (cases) and 931 control subjects from the Netherlands (32). This cohort is relatively homogeneous because all participants were of Dutch

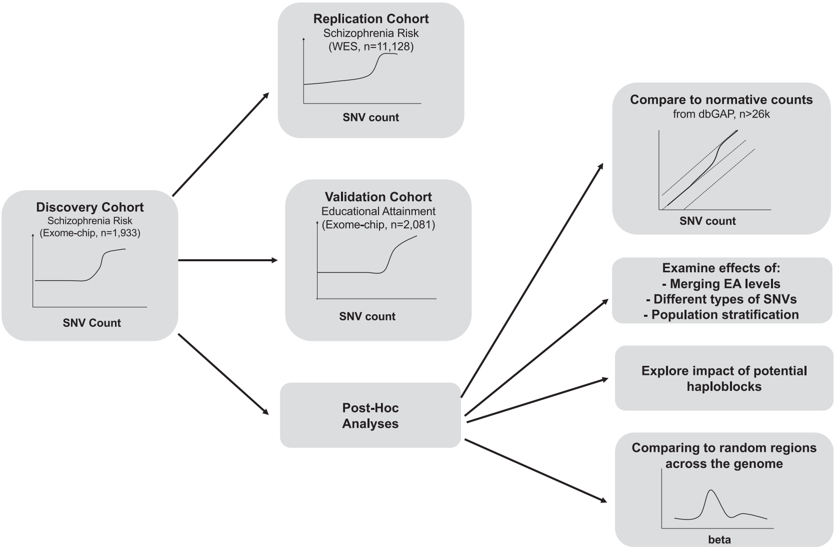

Figure 2. An overview of the study. The study consists of three independent cohorts: a discovery cohort, a replication cohort, and a validation cohort. We examined the impact of an incremental burden of low-frequency, functional SNVs in the 22q11.2 region on schizophrenia risk in the discovery and replication cohorts and on the level of EA in the validation sample. In the discovery and validation cohorts, data from the HumanExome BeadChip were used, and in the replication cohort, data from a WES study were used. Furthermore, we performed several post hoc analyses. dbGaP, Database of Genotypes and Phenotypes; EA, educational attainment; SNV; single nucleotide variant; WES, whole-exome sequencing.

descent (defined as having at least 3 grandparents of Dutch ancestry). Cases were recruited from psychiatric institutions throughout the Netherlands, and diagnosis was based on a DSM-IV diagnosis of schizophrenia (33). Control subjects were volunteers who had no psychiatric history.

Replication Cohort. In the replication cohort (*N* = 11,128), we examined the association with schizophrenia in a second, independent cohort, details of which have been described previously (8,34). This cohort consisted of 4909 individuals with schizophrenia and 6219 control subjects from Sweden. Psychiatric cases with a diagnosis of schizophrenia were ascertained from the Swedish National Hospital Discharge Register as described in previous studies (5,35); this captures all inpatient hospitalizations. Control subjects were randomly selected from population registers.

Validation Cohort. For our study on the effect on EA, we used genotype and phenotype data available from 2081 participants of the Utrecht Health Project, an ongoing, population-based, longitudinal cohort study conducted in a newly developed residential area in the Netherlands (36). Data collected include demographics (sex, ethnicity, and age) and EA categorized into low (primary school), middle (high school), or high (college or university) EA. The number of participants in the low-education group was relatively modest (19%); we therefore merged low and middle EA into one group (lower EA) in the initial analysis.

Definition of 22q11.2 Region

The 22q11.2 region is a structurally complex area of the genome that is characterized by multiple low-copy repeats (37–39). The two largest blocks, LCR22A and LCR22D, flank the most common ~3-Mb deletion in this region. The A-D deletion, a result of nonallelic homologous recombination between LCR22A and LCR22D, is present in approximately 85% of patients with 22q11DS (38,40). The ~1.5-Mb proximal deletion, flanked by LCR22A and LCR22B, is referred to as the minimal critical region because the approximately 5% to 10% of patients with 22q11DS with this smaller sized A-B deletion display the same (variable) phenotype, including cognitive and neurodevelopmental features, in particular schizophrenia. Consequently, we selected the minimal critical region, flanked by low-copy repeats A and B, for our analysis (10); this region contains approximately 30 protein-coding genes (16), (chromosome 22: 19.02 and chromosome 22: 20.26 Mb, respectively, GRCh37/hg19), henceforth referred to as the LCR22qA-B region.

Genotyping

DNA from participants in the discovery and validation cohorts were genotyped using the Illumina Infinium HumanExome BeadChip (version 1.1). This chip provides focused coverage of potentially functional protein-altering exonic variants, including missense, stop-gain, and splice site alteration SNVs with low minor allele frequencies ($<$0.05%). Genotypes were called with Illumina GenomeStudio software, and no-calls were called using zCall. Strict sample and genotype quality control had been ensured for these cohorts as part of previous studies with these data. In short, markers with low call rates ($\leq$95%) and subjects with high rates of missing genotypes ($>$5%) were excluded from our study (41–43).

Genotyping of the replication cohort was based on whole-exome sequencing (WES) data, as described previously (34). We applied several filters to ensure variant quality (see section 3 in Supplement 1).

For all three cohorts, we excluded all noncoding SNVs and SNVs with minor allele frequency $>$5%. In the discovery and validation cohorts, we queried 250 successive possible variants in the LCR22qA-B region, using the filtered list of SNVs detected by Illumina Infinium HumanExome BeadChip (section 1 in Supplement 2). Given that genotypes in the replication cohort were derived from exome sequencing data, the same region revealed a higher number of SNVs (773); of these, 426 (55%) were singletons, i.e., they were found only in 1 individual. The quality control steps of genotyping are similar to those used in genome-wide association studies and sequence studies, bar the standard correction for linkage disequilibrium, given that inherent to our study, we focused on the potential cumulative impact of several or multiple SNVs in the same genomic region. However, to address the potential impact of linkage disequilibrium, we examined the occurrence of haploblocks and compared our findings against a normative distribution of the SNV burden generated in a large general population sample; see Post Hoc Analyses below. Data were analyzed using R3.3.1, PLINK (version 1.90) (44), and SPSS (version 22.0).

Analytic Methods

In the discovery cohort, we determined the number of subjects per incrementally increasing count of SNVs (0, 1, 2, 3, and so on). For each SNV count, we calculated the ratio of individuals with schizophrenia to control subjects. Subsequently, we log-transformed the proportion of cases per SNV count, while grouping SNV counts above 5 into one bin (0, 1, 2, 3, 4, 5, or more). Based on this analysis, we established a threshold effect at a SNV burden of $\geq$4 SNVs at which the difference between cases and control subjects increases steeply. We used the two-sided Fisher's exact test to test for significance.

In the replication cohort, we applied the same counting procedures as in the discovery sample. Subsequently, we used binary logistic regression with schizophrenia status as the outcome and an above-threshold SNV count as the predictor. The sample size of our replication cohort also allowed us to investigate the regional burden of SNVs without dichotomization around a threshold value (section 4 in Supplement 1). To this end, we repeated the analysis with the normalized SNV count as the predictor. In both analyses, we used sex and the first three components from the population stratification multidimensional scaling analysis (PLINK version 1.90b6), restricted to only the European samples, as covariates.

Validation and Extending the Observations. To further validate our findings, we extended our analysis, with EA as the phenotype of interest. We reiterated the procedure followed in the discovery cohort and compared the numbers of subjects with lower EA versus high EA. We used the same cutoff for the SNV burden (≥4 SNVs) as in the discovery cohort. We also used this cutoff as predictor in a binary logistic regression model, with lower EA as outcome, and sex and age as covariates.

Post Hoc Analyses

We conducted the following post hoc analyses to examine the possible methodological or data-related factors that could influence our observations.

Comparing the Observed to a Normative Count Distribution. First, we examined the count distribution (i.e., number of subjects with 0, 1, 2, 3, and so on, SNVs) in each cohort to confirm their approximation of a Poisson distribution. We then used Database of Genotypes and Phenotypes (dbGaP) data to create an empirical normative distribution and compared this with the number of schizophrenia cases per SNV count in our discovery cohort. We used a dataset from dbGaP with more than 26,000 individuals from non-Hispanic European descent with exome data available. Importantly, these data were obtained with the same ExomeChip as the one used in the discovery cohort, allowing us to compare count distributions. We used a bootstrap procedure, taking 10,000 random samples from the dbGaP data, each time drawing the same number of subjects as the schizophrenia cases in the discovery cohort (*N* = 1002). Then, we determined the counts for each iteration and created empirical cutoffs of the mean and standard deviations for the expected number of subjects for each SNV count (0, 1, 2, 3, and so on).

Potential Impact of EA Levels, SNV Types, Population Stratification, and Genome-wide SNV Rate. To rule out any confounding due to our merging of educational levels, we repeated our analysis using the original three levels of EA as the outcome, ≥4 SNV burden as the predictor, and age and sex as covariates in an ordered logistic regression model.

The replication sample was sufficiently large to also reiterate the analysis for missense variants and for loss-of-function variants separately.

We considered the potential impact of population stratification in the three samples. In our discovery cohort, no significant influence of population stratification or cryptic relatedness had been shown by previous studies. In the replication cohort, we corrected for population stratification by using the first three components from the population stratification multidimensional scaling analysis. In the validation cohort, we examined possible confounding effects due to ethnicity by reanalyzing the data with the reported European origin as a covariate. In addition, to examine the possible effects of population stratification, we performed a principal component analysis and an analysis to detect relatedness in the validation cohort (section 8 in Supplement 1). We also examined whether the genome-wide SNV rate could account for the observed effect of the LCR22qA-B regional burden; no associations were observed between genome-wide and an increased regional burden in LCR22qA-B; in addition, no associations were observed between genome-wide SNV rates and phenotypic outcomes (section 9 in Supplement 1).

Explore the Potential Impact of Small Haploblocks. We explored whether collapsing potential haploblocks in the region altered the association of the regional SNV burden in the LCR22A-B region with schizophrenia and EA.

Explore Impact of Increased Regional Burden of SNVs Across the Genome. We examined the extent to which the observed phenotypic impact of an increased regional SNV burden is specific to the 22q11.2 LCR22A-B region. We reiterated the exact same procedure to calculate the phenotypic effects of the above-threshold regional SNV burden for a set of randomly generated regions across the genome, each containing a similar number of consecutive SNV loci as observed in the 22q11.2 region. This permutation allowed us to place the impact of the SNV burden in 22q11.2 in the context of the distribution of SNV burden effects in similarly defined regions across the genome (section 5 in Supplement 1). In addition, we explored the impact of the regional SNV burden in other genomic regions where recurrent CNVs are associated with schizophrenia (section 7 in Supplement 1).

RESULTS

Primary Analysis

Figure 3A, B shows the relationship between schizophrenia risk and the number of functional SNVs in the LCR22qA-B region. In our discovery cohort, a regional burden of four SNVs in the LCR22qA-B region marked a threshold beyond which a steep incline was observed for the proportion of schizophrenia cases, from 81.2% (4 SNVs) to 100% (≥5 SNVs). The incline in our independent replication cohort was more incremental but nevertheless indicative of a similar threshold. Based on the distribution of counts in the WES data, we identified a cutoff of 5 SNVs or more in our replication sample. Applying these cutoffs to the discovery and replication cohorts yielded comparable proportions of subjects with a similarly defined excess regional SNV burden (0.7% and 0.5%, respectively). The odds ratio (OR) of an above-threshold regional SNV burden in 22q11.2 on schizophrenia risk was 7.48 in the discovery cohort (95% CI = 1.00–331.98, *p* = .039) and 1.92 in the replication cohort (95% CI = 1.21–3.12, *p* = .004). The large sample size of the replication cohort also allowed us to examine the effect of a regional SNV burden in 22q11.2 without applying a threshold; beta was 0.052 (95% CI = 0.013–0.090, *p* = .008).

Validation and Extension to EA

The validation sample was derived from the general population (*N* = 2081), was 45% male, had a mean age of 39.3 (SD = 13.2) years, and was of 86% Dutch ethnicity. Regarding EA levels, 40.5% had attended college or university, and 59.5% had lower EA. Again, a regional burden of 4 SNVs in the LCR22qA-B region marked a threshold beyond which a steep incline was observed for the proportion of participants with a lower EA, from 59.5% lower EA in those with 0, 1, or 3 SNVs in 22q11.2 to 83.4% lower EA in those with 4 SNVs in this region.

Figure 4 shows the relationship between EA and SNV counts in the LCR22qA-B region. The effect size (OR) was 4.65 (95% CI = 1.00–21.56, *p* = .049).

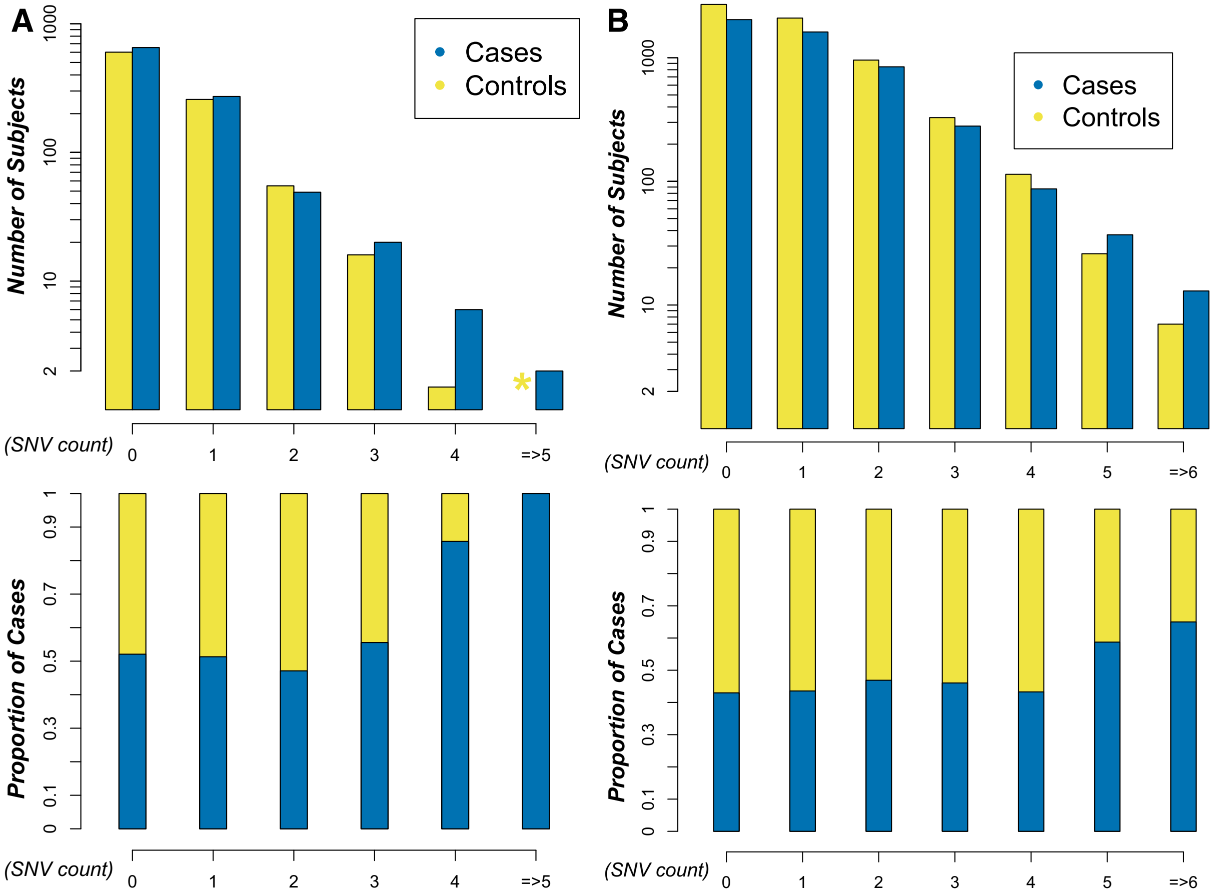

Figure 3. The effect of a SNV burden at the 22q11.2 A-B locus on schizophrenia risk. **(A)** Results for the discovery cohort, a schizophrenia case-control cohort using the HumanExome BeadChip. The log-transformed number of subjects for each SNV count is given for cases and control subjects. Furthermore, the proportion of cases is given for each SNV count. The plot shows that until a SNV count of 3, cases and control subjects do not differentiate from each other. The number of control subjects with 4 SNVs decreases more rapidly compared with cases, and there are no control subjects with 5 or more SNVs. This is reflected in a steep incline in the proportion of cases at the threshold of 4 or more SNVs (from ~50% to ~86%). Note that the data do not indicate an incremental dosage effect. **(B)** Results for the replication cohort, a schizophrenia case-control cohort using whole-exome sequencing data. The log-transformed number of subjects for each SNV count is given for cases and control subjects. Furthermore, the proportion of cases is given for each SNV count. The plot shows that until a SNV count of 4, cases and control subjects do not differentiate from each other. At 5 SNVs, the number of control subjects is for the first time smaller than that of cases, and the number of control subjects with 5 or more SNVs decreases more rapidly than the number of cases. This is reflected in an incline in the proportion of cases at the threshold of 5 or more SNVs (from ~43% to ~60%). In this case, the data indicate an incremental dosage effect but with an additional increase after the threshold. SNV, single nucleotide variant.

Post Hoc Analyses

Compare to Normative Count Distribution. For all three cohorts, the counts followed a distribution approximating a Poisson distribution with a lambda of 0.46 for the discovery cohort, 0.93 for the replication cohort, and 0.46 for the validation cohort. Figure 5 shows the results from the comparison between counts for schizophrenia cases in the discovery cohort to empirical count distribution based on dbGaP data. These results confirmed the earlier findings in that with ≤3 SNVs, the number of patients with schizophrenia follow the distribution as observed in the dbGaP data. This is consistent with our observation that the proportion of patients with schizophrenia is equal to control subjects in these lower SNV counts. However, with ≥4 SNVs, a difference becomes apparent, consistent with the start of the deviation from the normal distribution derived from dbGaP (Figure 5). From the 10,000 count distributions from the bootstrap, 17 had similar counts as the observed counts in schizophrenia cases, i.e., an initial trajectory around the mean (±10 percentile points) for the lower counts and followed by a deviation exceeding 1 SD (p = .002). Finally, we repeated the comparison for lower EA, showing a similar pattern (p = .018).

Potential Impact of EA Levels, SNV Types, and Population Stratification. The results from the ordered logistic regression confirmed the observed association between a regional burden of ≥4 SNVs at 22q11.2 (OR = 3.55, 95% CI = 1.34–9.74, p = .012) (section 2 in Supplement 1).

When restricting our analysis in the validation cohort to only missense variants, we observed a similar regional burden SNV

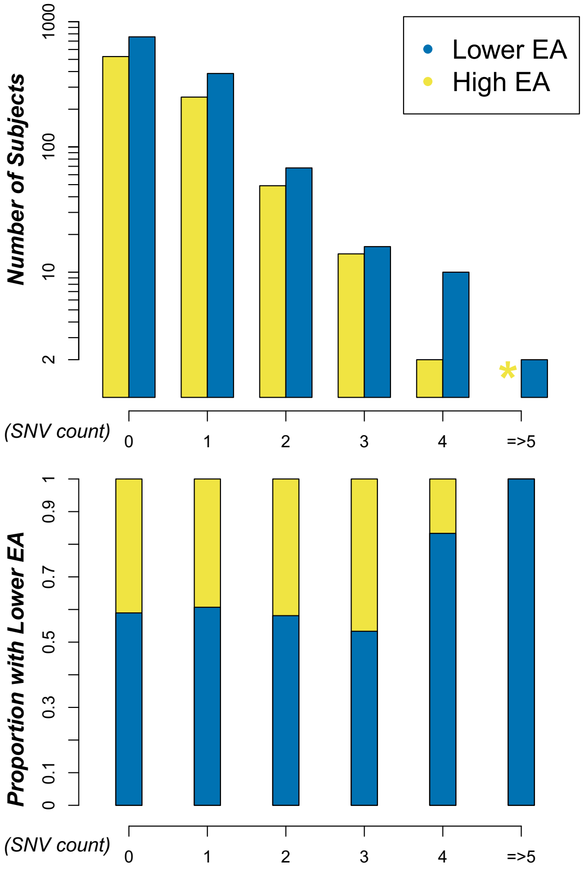

Figure 4. The effect of a SNV burden at the 22q11.2 A-B locus on the level of EA. The results are given for the validation cohort, a population-based cohort using the Human Exome BeadChip. The log-transformed number of subjects for each SNV count is given for subjects with a lower EA level compared with subjects with a high EA level. Furthermore, the proportion of subjects with a lower EA level is given for each SNV count. The plot shows that until a SNV count of 3, the two groups do not differentiate from each other. The number of subjects with a high EA level with 4 SNVs decreases more rapidly than for cases, and there are no subjects with a high EA level with 5 or more SNVs. This is reflected in a steep incline in the proportion of cases at the threshold of 4 or more SNVs (from ~60% to ~83.4%). Note that the data do not indicate an incremental dosage effect. EA, educational attainment; SNV, single nucleotide variant.

effect in the LCR22qA-B region on schizophrenia risk (OR = 1.94, *p* = .003). Limiting the analysis to only loss-of-function variants yielded 387 subjects with one loss-of-function variant and only 2 subjects with two variants, precluding a regional burden analysis. The cases were not overrepresented in 387 subjects with one loss-of-function variant. The only subject with two loss-of-function variants in the LCR22qA-B region was a patient with schizophrenia.

The results of the principal component analysis confirmed the self-reported ethnicity status of the participants in the validation cohort and showed no indication for population stratification (section 7 in Supplement 1). There was also no indication of cryptic relatedness within the validation cohort.

Finally, we confirmed that our findings regarding the LCR22qA-B regional burden were not driven by increased genome-wide rates of SNVs (section 9 in Supplement 1).

Explore the Impact of Potential Small Haploblocks. Our exploratory analysis of the potential haploblocks contributing to the regional SNV burden in the LCR22qA-B region indicated that 7 of the 9 subjects in the discovery cohort and 10 of the 14 subjects in the validation cohort with the above-threshold regional SNV burden shared a haploblock consisting of 3 co-occurring SNVs, all located within the *CLTCL1* gene. Note that the presence of this haploblock of 3 SNVs in *CLTCL1* alone, i.e., without additional SNVs in the LCR22qA-B region, was not associated with schizophrenia (*p* = .66) or with lower EA (*p* = 1.00) (section 1 in Supplement 1). In the replication cohort, we observed the same *CLTCL1* haploblock but also other correlated SNVs (combination of SNVs that co-occur in multiple subjects). The *CLTCL1* haploblock was by itself again not associated with increased risk for schizophrenia (*p* = 1.00). Correcting for the presence of the observed haploblocks did not alter the results for the replication sample, indicating that the *CLTCL1* haploblock alone had no clinical impact; only in combination with additional SNVs did the association with the phenotype emerge.

Explore Impact of Increased Regional Burden of SNVs Across the Genome. In the distributions of effect sizes obtained in the replication and validation samples, the

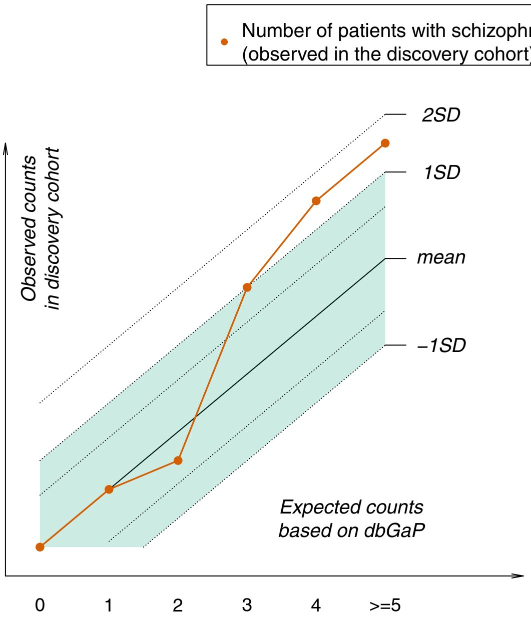

Figure 5. Comparison between SNV counts for schizophrenia cases in the discovery cohort and empirical count distribution. The results are given for the comparison between SNV counts for schizophrenia cases to a normative count distribution derived from control subjects from the Database of Genotypes and Phenotypes. We performed a bootstrap procedure taking 10,000 random samples from the Database of Genotypes and Phenotypes data with the same sample size as the schizophrenia cases in cohort 1 (*N* = 1002). We compared the number of schizophrenia cases for each SNV count to this normative count distribution. We see that for SNV counts 0, 1, 2, and 3, the schizophrenia cases do not deviate from the mean of the normative distribution but that for the SNV counts of 4 and 5 or more, the numbers of schizophrenia cases is above 1 SD. Finding this specific pattern based on changes is smaller than 0.17%. SNV, single nucleotide variant.

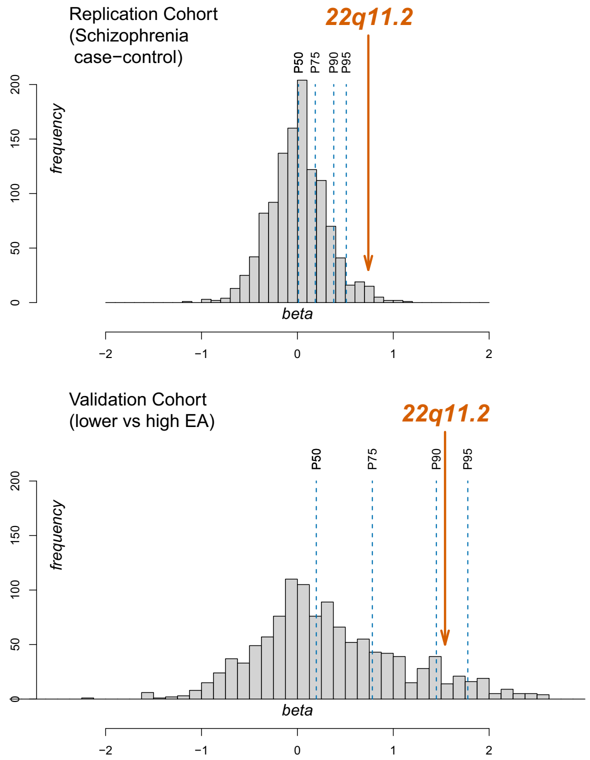

Figure 6. Comparing results for the 22q11.2 region to random regions across the genome. Distribution of effect sizes of the phenotypic impact of the above-threshold, regional, single nucleotide variant burden in permuted genomic regions comparable to the 22q11.2 LCR22qA-B region. The orange arrow indicates the beta exerted by the above-threshold, regional, single nucleotide variant burden in the 22q11.2 region. Vertical dotted lines indicate distribution percentiles (p50, p75, p90, p95). EA, educational attainment.

effect size of the phenotypic effect of the regional SNV burden in the 22q11.2 region is above the 98th and 92.5th percentiles in the replication and validation cohorts, respectively (Figure 6). The a priori probability of these findings in two independent cohorts is .006 (section 5 in Supplement 1). Our explorative analyses into the overlap between genomic regions across the genome and other genomic regions where recurrent CNVs are associated with schizophrenia shows that five of these eight regions (including LCR22qA-B) fell in the tail of the probability distribution (>85th percentile) for both replication and validation cohorts (section 5 in Supplement 1).

DISCUSSION

We investigated the phenotypic impact of the cumulative burden of SNVs in regions with known recurrent pathogenic CNVs in three independent cohorts focusing on the 22q11.2 region, with schizophrenia and EA as the primary phenotypes of interest.

We showed that an elevated burden of low-frequency SNVs in the 22q11.2 region is associated with an increased risk of schizophrenia and a lower level of EA in the general population. These findings suggest that LCR22qA-B, a genomic region in which CNVs have a documented, strong phenotypic impact on schizophrenia risk and intelligence, is also vulnerable to the effect of a cumulative burden of low-frequency SNVs.

Our observation that an effect can be observed in the tail of the distribution of SNV counts (in just under 1%), suggests a threshold pattern, which is somewhat unexpected because studies examining the cumulative phenotypic impact of common variants generally indicate an additive-effect model (3,45–47). Possibly, the effect observed in this study is essentially different from genome-wide polygenic effects such as those implemented in polygenic risk scores. Instead, multiple hits in a vulnerable genomic region may be more similar to the phenotypic impact of a structural genetic variant in such regions. We suggest that this might also explain why classical genetic association studies have yet to identify causative SNVs in the 22q11.2 region, which seems at odds with the fact that this same region, when affected by a deletion, constitutes the single largest genetic risk factor of schizophrenia.

The convergence of phenotypic impact of SNV burden in the 22q11.2 region on both schizophrenia and EA is consistent with the hypothesized association between schizophrenia and cognitive ability (31). Recent evidence for shared, common genetic variants between schizophrenia and EA (27–29,48) is consistent with these results. For example, our results suggest a role of *CLTCL1* in both schizophrenia risk and EA level. The potential relevance of *CLTCL1* for brain development is suggested by the observation of a rare homozygous missense *CLTCL1* mutation in a patient with severe ID (49) and a compound heterozygous mutation in a patient with infantile spasm (50). *CLTCL1* is also one of the 709 genes associated with cognitive functions in a genome-wide association study in more than 300,000 individuals (27). Consistent with a burden effect, we found that carriers of a *CLTCL1* haploblock were at increased risk of lower EA or schizophrenia only if the same haploblock concurred with at least 1 additional SNV in the LCR22qA-B region.

The limitations of our study include the relatively small sample size of the discovery cohort, resulting in large CIs. In addition, the predetermined, and somewhat restricted, number of SNVs on the exome chip may be considered as a limitation to the generalizability of our finding in the discovery cohort. However, we have addressed both limitations by seeking replication of our discovery findings in a large replication cohort with WES data. In this sample, we observed essentially the same effect, albeit with a higher SNV count threshold, which is not unexpected given the several-fold higher number of SNV calls in the WES data compared with the exome chip data. We surmise that extending the observed impact of a regional SNV burden in the 22q11.2 region across two phenotypes further strengthens our findings.

Winners' curse probably plays a role in the observed, smaller effect size in the replication cohort; another possible explanation is that the control group in the latter included subjects with other psychiatric diagnoses as well as individuals with lower EA. Finally, similar to most large-scale genetic studies, our study population consisted mainly of participants of European descent. It is unclear whether the relationship between increased regional burden and risk of schizophrenia differs across populations, but it needs to be explored in future studies (51).

Results from our permutation analysis indicate that the phenotypic impact of a regional SNV burden in the 22q11.2 region is higher than in most other regions across the genome with a similar number of consecutive SNVs. While the focus of this study is on the 22q11.2 region, the existence of other genomic regions with similar susceptibility for the cumulative effect of multiple regional SNVs is highly probable. Our future studies will expand on the principle presented here, among others, by examining the phenotypic impact of a regional SNV burden in other genomic regions recurrently affected by pathogenic CNVs. Indeed, the 22q11.2 region is only one of multiple hot spot regions in the human genome where segmental duplications increase the probability of recurrent pathogenic CNVs (52). As a result, CNVs can recur at these loci in the population, despite their negative impact on neurocognitive function and reproductive fitness in affected individuals. Many genes in these regions with human-specific segmental duplications are associated with brain development, function, and disease (53,54).

The regional burden approach could potentially contribute to explaining a part of the missing heritability, possibly with observed effect sizes between the strong phenotypic impacts for some of the rare variants (6) and the small effect sizes observed in common variants.

In summary, we report that an increased regional burden of protein-altering SNVs in the LCR22qA-B region confers a risk of schizophrenia as well as a negative impact on EA in the general population. Our results suggest that it may be worthwhile to study the regional burden of low-frequency SNVs in other genomic regions, in particular those known to be vulnerable to recurrent pathogenic CNVs with neurodevelopmental impact.

ACKNOWLEDGMENTS AND DISCLOSURES

This work was supported by the Mr. Rajeeb Mukherjee B.Sc. (Hons.) (London School of Economics) Fellowship in Schizophrenia (to EJB), an academic scholarship award from the Department of Psychiatry at the University of Toronto (to EJB), the O'Brien Scholarship (to EJB), and the National Institutes of Mental Health (Grant No. 5U01MH101722-02 [to JASV]).

DM is a full-time employee of Deep Genomics Inc. and is entitled to a stock option. JASV serves as a consultant for NoBias Therapeutics Inc. SWS is on the Scientific Advisory Committees of Deep Genomics and Population Bio and is a highly cited Academic Consultant for the King Abdul Aziz University. All other authors report no biomedical financial interests or potential conflicts of interests.

ARTICLE INFORMATION

From the Department of Psychiatry (EJB, PS, JASV), The Hospital for Sick Children; Department of Psychiatry (EJB, ASB, PS, JASV), University of Toronto; Center for Applied Genomics, Genetics and Genome Biology (DM, XW, SWS), The Hospital for Sick Children; Deep Genomics Inc. (DM); Dalglish Family 22q Clinic for Adults with 22q11.2 Deletion Syndrome (ASB), Toronto General Hospital, University Health Network; Clinical Genetics Research Program (ASB), Centre for Addiction and Mental Health; Division of Clinical and Metabolic Genetics (ASB), The Hospital for Sick Children; Medical Genetics and Genomics Residency Training Program (ASB), University of Toronto; Toronto General Research Institute (ASB); Campbell Family Mental Health Research Institute (ASB); McLaughlin Centre and Department of Molecular Genetics (SWS), University of Toronto; the Program in Genetics and Genome Biology (JASV), Research Institute, The Hospital for Sick Children, Toronto, Ontario, Canada; Department of Psychiatry (KCS, MPMB, RSK, JASV), University Medical Center Utrecht Brain Center; Department of Cardiology (JvS), and Department of Medical Oncology (KCS), University Medical Center Utrecht, University of Utrecht; Department of Epidemiology (KCS, IV), Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, University of Utrecht, Utrecht, The Netherlands; and Department of Psychiatry (RSK), Icahn School of Medicine at Mount Sinai, NewYork, New York.

Address correspondence to Elemi J. Breetvelt, M.D., Ph.D., at Elemi.Breetvelt@sickkids.ca.

Received Apr 17, 2021; revised Oct 25, 2021; accepted Nov 12, 2021.

Supplementary material cited in this article is available online at https://doi.org/10.1016/j.biopsych.2021.11.019.

REFERENCES

1. Sullivan PF, Kendler KS, Neale MC (2003): Schizophrenia as a complex trait: Evidence from a meta-analysis of twin studies. Arch Gen Psychiatry 60:1187–1192.
2. Cross-Disorder Group of the Psychiatric Genomics Consortium, Lee SH, Ripke S, Neale BM, Faraone SV, Purcell SM, *et al.* (2013): Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat Genet 45:984–994.
3. Schizophrenia Working Group of the Psychiatric Genomics (2014): Biological insights from 108 schizophrenia-associated genetic loci. Nature 511:421–427.
4. Sullivan PF, Daly MJ, O'Donovan M (2012): Genetic architectures of psychiatric disorders: The emerging picture and its implications. Nat Rev Genet 13:537–551.
5. Purcell SM, Moran JL, Fromer M, Ruderfer D, Solovieff N, Roussos P, *et al.* (2014): A polygenic burden of rare disruptive mutations in schizophrenia. Nature 506:185–190.
6. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, *et al.* (2009): Finding the missing heritability of complex diseases. Nature 461:747–753.
7. Stefansson H, Rujescu D, Cichon S, Pietiläinen OPH, Ingason A, Steinberg S, *et al.* (2008): Large recurrent microdeletions associated with schizophrenia. Nature 455:232–236.
8. Buizer-Voskamp JE, Muntjewerff JW, Genetic Risk Outcome in Psychosis (GROUP) Consortium Members, Strengman E, Sabatti C, Stefansson H, *et al.* (2011): Genome-wide analysis shows increased frequency of copy number variation deletions in Dutch schizophrenia patients. Biol Psychiatry 70:655–662.
9. International Schizophrenia Consortium (2008): Rare chromosomal deletions and duplications increase risk of schizophrenia. Nature 455:237–241.
10. Rees E, Walters JTR, Georgieva L, Isles AR, Chambert KD, Richards AL, *et al.* (2014): Analysis of copy number variations at 15 schizophrenia-associated loci. Br J Psychiatry 204:108–114.
11. Marshall CR, Howrigan DP, Merico D, Thiruvahindrapuram B, Wu W, Greer DS, *et al.* (2017): Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects [published correction appears in Nat Genet 2017; 49:651] [published correction appears in Nat Genet 2017; 49:1558]. Nat Genet 49:27–35.
12. Vorstman JAS, Ophoff RA (2013): Genetic causes of developmental disorders. Curr Opin Neurol 26:128–136.
13. Malhotra D, Sebat J (2012): CNVs: Harbingers of a rare variant revolution in psychiatric genetics. Cell 148:1223–1241.
14. Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J, *et al.* (2008): Structural variation of chromosomes in autism spectrum disorder. Am J Hum Genet 82:477–488.
15. Oskarsdóttir S, Vujic M, Fasth A (2004): Incidence and prevalence of the 22q11 deletion syndrome: A population-based study in Western Sweden. Arch Dis Child 89:148–151.
16. McDonald-McGinn DM, Sullivan KE, Marino B, Philip N, Swillen A, Vorstman JAS, *et al.* (2015): 22q11.2 deletion syndrome. Nat Rev Dis Primers 1:15071.
17. Levinson DF, Duan J, Oh S, Wang K, Sanders AR, Shi J, *et al.* (2011): Copy number variants in schizophrenia: Confirmation of five previous findings and new evidence for 3q29 microdeletions and VIPR2 duplications. Am J Psychiatry 168:302–316.

18. Cleynen I, Engchuan W, Hestand MS, Heung T, Holleman AM, Johnston HR, *et al.* (2021): Genetic contributors to risk of schizophrenia in the presence of a 22q11.2 deletion. Mol Psychiatry 26:4496–4510.
19. Gothelf D, Schneider M, Green T, Debbané M, Frisch A, Glaser B, *et al.* (2013): Risk factors and the evolution of psychosis in 22q11.2 deletion syndrome: A longitudinal 2-site study. J Am Acad Child Adolesc Psychiatry 52:1192–1203.e3.
20. Hooper SR, Curtiss K, Schoch K, Keshavan MS, Allen A, Shashi V (2013): A longitudinal examination of the psychoeducational, neurocognitive, and psychiatric functioning in children with 22q11.2 deletion syndrome. Res Dev Disabil 34:1758–1769.
21. Kates WR, Russo N, Wood WM, Antshel KM, Faraone SV, Fremont WP (2015): Neurocognitive and familial moderators of psychiatric risk in velocardiofacial (22q11.2 deletion) syndrome: A longitudinal study. Psychol Med 45:1629–1639.
22. Vorstman JAS, Breetvelt EJ, Duijff SN, Eliez S, Schneider M, Jalbrzikowski M, *et al.* (2015): Cognitive decline preceding the onset of psychosis in patients with 22q11.2 deletion syndrome. JAMA Psychiatry 72:377–385.
23. Gothelf D, Eliez S, Thompson T, Hinard C, Penniman L, Feinstein C, *et al.* (2005): COMT genotype predicts longitudinal cognitive decline and psychosis in 22q11.2 deletion syndrome. Nat Neurosci 8:1500–1502.
24. Keefe RSE, Perkins DO, Gu H, Zipursky RB, Christensen BK, Lieberman JA (2006): A longitudinal study of neurocognitive function in individuals at-risk for psychosis. Schizophr Res 88:26–35.
25. Meier MH, Caspi A, Reichenberg A, Keefe RSE, Fisher HL, Harrington H, *et al.* (2014): Neuropsychological decline in schizophrenia from the premorbid to the postonset period: Evidence from a population-representative longitudinal study. Am J Psychiatry 171:91–101.
26. Woodberry KA, Giuliano AJ, Seidman LJ (2008): Premorbid IQ in schizophrenia: A meta-analytic review. Am J Psychiatry 165:579–587.
27. Davies G, Lam M, Harris SE, Trampush JW, Luciano M, Hill WD, *et al.* (2018): Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function [published correction appears in Nat Commun 2019; 10:2068. Nat Commun 9:2098.
28. McIntosh AM, Gow A, Luciano M, Davies G, Liewald DC, Harris SE, *et al.* (2013): Polygenic risk for schizophrenia is associated with cognitive change between childhood and old age. Biol Psychiatry 73:938–943.
29. Smeland OB, Frei O, Kauppi K, Hill WD, Li W, Wang Y, *et al.* (2017): Identification of genetic loci jointly influencing schizophrenia risk and the cognitive traits of verbal-numerical reasoning, reaction time, and general cognitive function. JAMA Psychiatry 74:1065–1075.
30. Davies RW, Fiksinski AM, Breetvelt EJ, Williams NM, Hooper SR, Monfeuga T, *et al.* (2020): Using common genetic variation to examine phenotypic expression and risk prediction in 22q11.2 deletion syndrome. Nat Med 26:1912–1918.
31. Kahn RS, Keefe RSE (2013): Schizophrenia is a cognitive illness: Time for a change in focus. JAMA Psychiatry 70:1107–1112.
32. Genome of the Netherlands Consortium (2014): Whole-genome sequence variation, population structure and demographic history of the Dutch population. Nat Genet 46:818–825.
33. American Psychiatric Association (1994). In: Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) Washington, DC: American Psychiatric Publishing.
34. Genovese G, Fromer M, Stahl EA, Ruderfer DM, Chambert K, Landén M, *et al.* (2016): Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. Nat Neurosci 19:1433–1441.
35. Ripke S, O'Dushlaine C, Chambert K, Moran JL, Kähler AK, Akterin S, *et al.* (2013): Genome-wide association analysis identifies 13 new risk loci for schizophrenia. Nat Genet 45:1150–1159.
36. Grobbee DE, Hoes AW, Verheij TJM, Schrijvers AJP, van Ameijden EJC, Numans ME (2005): The Utrecht Health Project: Optimization of routine healthcare data for research. Eur J Epidemiol 20:285–287.
37. Edelmann L, Pandita RK, Spiteri E, Funke B, Goldberg R, Palanisamy N, *et al.* (1999): A common molecular basis for rearrangement disorders on chromosome 22q11. Hum Mol Genet 8:1157–1167.
38. Shaikh TH, Kurahashi H, Saitta SC, O'Hare AM, Hu P, Roe BA, *et al.* (2000): Chromosome 22-specific low copy repeats and the 22q11.2 deletion syndrome: Genomic organization and deletion endpoint analysis. Hum Mol Genet 9:489–501.
39. Saitta SC, Harris SE, Gaeth AP, Driscoll DA, McDonald-McGinn DM, Maisenbacher MK, *et al.* (2004): Aberrant interchromosomal exchanges are the predominant cause of the 22q11.2 deletion. Hum Mol Genet 13:417–428.
40. Edelmann L, Pandita RK, Morrow BE (1999): Low-copy repeats mediate the common 3-Mb deletion in patients with velo-cardio-facial syndrome. Am J Hum Genet 64:1076–1086.
41. Loohuis LMO, Vorstman JAS, Ori AP, Staats KA, Wang T, Richards AL, *et al.* (2015): Genome-wide burden of deleterious coding variants increased in schizophrenia. Nat Commun 6:7501.
42. de Kovel CG, Mulder F, van Setten J, van 't Slot R, Al-Rubaish A, Alshehri AM, *et al.* (2016): Exome-wide association analysis of coronary artery disease in the Kingdom of Saudi Arabia population. PLoS One 11:e0146502.
43. Moons T, De Hert M, Gellens E, Gielen L, Sweers K, Jacqmaert S, *et al.* (2016): Genetic evaluation of schizophrenia using the Illumina HumanExome chip. PLoS One 11:e0150464.
44. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, *et al.* (2007): PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559–575.
45. Huguet G, Schramm C, Douard E, Tamer P, Main A, Monin P, *et al.* (2021): Genome-wide analysis of gene dosage in 24,092 individuals estimates that 10,000 genes modulate cognitive ability. Mol Psychiatry 26:2663–2676.
46. Brainstorm Consortium, Anttila V, Bulik-Sullivan B, Finucane HK, Walters RK, Bras J, *et al.* (2018): Analysis of shared heritability in common disorders of the brain. Science 360:eaap8757.
47. Psychiatric GWAS Consortium Steering Committee (2009): A framework for interpreting genome-wide association studies of psychiatric disorders. Mol Psychiatry 14:10–17.
48. Le Hellard S, Wang Y, Witoelar A, Zuber V, Bettella F, Hugdahl K, *et al.* (2017): Identification of gene loci that overlap between schizophrenia and educational attainment. Schizophr Bull 43:654–664.
49. Nahorski MS, Al-Gazali L, Hertecant J, Owen DJ, Borner GHH, Chen YC, *et al.* (2015): A novel disorder reveals clathrin heavy chain-22 is essential for human pain and touch development. Brain 138:2147–2160.
50. Wang Q, Liu Z, Lin Z, Zhang R, Lu Y, Su W, *et al.* (2019): De novo germline mutations in *SEMA5A* associated with infantile spasms. Front Genet 10:605.
51. Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, *et al.* (2017): Human demographic history impacts genetic risk prediction across diverse populations [published correction appears in Am J Hum Genet 2020; 107:788–789]. Am J Hum Genet 100:635–649.
52. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, *et al.* (2004): Detection of large-scale variation in the human genome. Nat Genet 36:949–951.
53. Sassa T (2013): The role of human-specific gene duplications during brain development and evolution. J Neurogenet 27:86–96.
54. Levchenko A, Kanapin A, Samsonova A, Gainetdinov RR (2018): Human accelerated regions and other human-specific sequence variations in the context of evolution and their relevance for brain development. Genome Biol Evol 10:166–188.