Penetrance for copy number variants associated with schizophrenia

Evangelos Vassos, David A Collier, Simon Holden, Christine Patch, Dan Rujescu, David St Clair, Cathryn M Lewis

Supplementary Methods

Since in many samples there were zero observations of CNVs in control cohorts, classical statistical methods using the binomial distribution or direct analysis of contingency tables cannot be applied and odds ratios are not defined. We therefore took a Bayesian approach, to derive a posterior distribution of likely values for the frequency of CNVs in case and control cohorts using the observed CNV frequencies from published data. We then simulated from this distribution to obtain empirical estimates of penetrance and their credible intervals.

We assumed a prior distribution for CNV frequency (p) for both cases and controls of p ~Beta(α, β), which is defined for $0<p<1$, as required, with $\alpha, \beta > 0$. Then, with n CNVs observed in N cases, and m CNVs observed in M controls (for $n, m \geq 0$), the posterior distributions for CNV frequency in cases and controls are Beta($\alpha+n$, $\beta+N-n$) and Beta($\alpha+m$, $\beta+M-m$), respectively. We sampled pairs of CNV frequencies from these distributions, and calculate the penetrance, i.e. the probability of developing schizophrenia (disease, D) for individuals carrying the CNV (genotype, G), from

$$P(D \mid G) = \frac{P(G \mid D)P(D)}{P(G \mid D)P(D) + P(G \mid \overline{D})P(\overline{D})}$$, where $\overline{D}$ represents controls, who do not have schizophrenia, and P(D) the lifetime morbid risk for schizophrenia.

An uninformative prior of a uniform distribution (Beta(1,1)) was used, which assumes that the frequency of CNV could equally take any value between 0 and 1. For the estimation of penetrance of schizophrenia we adopted the conservative median global value for lifetime morbid risk ($P(D)$ above) of 0.72% from a recent comprehensive meta-analysis. Simulation

was performed using the R statistical package (http://cran.r-project.org/), with 2.5%, 50% and 97.5% quantiles extracted to obtain the median penetrance, and its approximate 95% credible intervals.

At first we employed the above formula replacing $P(G|D)$ with n/N (CNV frequency in cases) and $P(G|\overline{D})$ with m/M (CNV frequency in controls). This approach assumes that the controls of the original studies did not include cases with schizophrenia. However, since not all the controls were screened for schizophrenia and some controls may develop the disease after their recruitment, we examined the possibility that the control groups may include "hidden" cases with schizophrenia. Let *L* be the proportion of controls screened absent for schizophrenia, and assume, conservatively, that the remaining proportion of controls *(1-L)* carry the general population risk for schizophrenia. Thus, the total number of cases was estimated to be *N+M*(1-L)*P(D)*, the number of screened controls *M- M*(1-L)*P(D*), the expected number of CNVs in cases *n+M*(1-L)*P(D)*n/N* and the number of CNVs in controls *m- M*(1-L)*P(D)*n/N*. Using these formulae we estimated penetrance for various values of *L* including the extreme values of *L=1* (no schizophrenia cases hidden in controls, as assumed previously) and *L=0* (general population controls, unscreened for schizophrenia). The results are presented in supplement table A.

Sensitivity analysis was carried out, using more informative priors of Beta(2,5) and Beta(0.5,2), which assume CNVs have lower frequencies in the population (mean values of 0.28 and 0.2, respectively). These provided very similar results, with little change in the penetrance estimates, and with little decrease in the width of confidence intervals. A bootstrapping simulation study was also performed for studies with sufficient numbers of CNVs observed in both cases and controls. We simulated representative cohorts from the observed CNV carriers/non-carriers, with replacement, and calculated penetrance. For those studies for which this procedure was possible, the penetrance estimates and confidence intervals were very similar to those obtained using the Bayesian method.

Supplementary Table

Table S1: Frequency and penetrance of CNVs for schizophrenia.

CNV	Reference	observed CNVs (frequency) Cases	observed CNVs (frequency) Controls	Penetrance (95% CI) *L*=1	Penetrance (95% CI) *L*=0.5	Penetrance (95% CI) *L*=0
1q21.1 del	Stefansson	11/4718 (0.00233)	8/41199 (0.00019)	0.078 (0.035-0.174)	0.081 (0.036-0.181)	0.084 (0.037-0.188)
1q21.1 del	ISC	10/3391 (0.00295)	1/3181 (0.00031)	0.041 (0.011-0.241)	0.042 (0.011-0.250)	0.043 (0.011-0.257)
1q21.1 del	Kirov*	17/7918 (0.00215)	11/46502 (0.00024)	0.061 (0.030-0.122)	0.062 (0.031-0.125)	0.064 (0.032-0.129)
2p16.3 del (NRXN1)	Rujescu	12/2977 (0.00403)	49/33746 (0.00145)	0.020 (0.011-0.036)	0.021 (0.011-0.036)	0.021 (0.011-0.036)
2p16.3 del (NRXN1)	Vrijenhoek	4/752 (0.00532)	1/706 (0.00142)	0.019 (0.004-0.130)	0.018 (0.004-0.131)	0.019 (0.004-0.134)
15q11.2 del	Stefansson	26/4718 (0.00551)	79/41199 (0.00192)	0.021 (0.013-0.031)	0.021 (0.013-0.031)	0.021 (0.013-0.031)
15q11.2 del	Kirov*	49/7918 (0.00619)	103/46497 (0.00222)	0.020 (0.014-0.028)	0.020 (0.014-0.028)	0.020 (0.014-0.028)
15q13.3 del	Stefansson	7/4213 (0.00166)	8/39800 (0.00020)	0.056 (0.022-0.136)	0.057 (0.022-0.139)	0.059 (0.023-0.142)
15q13.3 del	ISC	9/3391 (0.00265)	0/3181 -	0.087 (0.015-0.730)	0.086 (0.015-0.728)	0.086 (0.015-0.726)
15q13.3 del	Kirov*	15/7413 (0.00202)	8/45103 (0.00018)	0.074 (0.034-0.160)	0.076 (0.035-0.166)	0.078 (0.036-0.173)
16p13.1 dup	Ingason	13/4345 (0.00299)	32/35079 (0.00091)	0.024 (0.012-0.043)	0.024 (0.013-0.043)	0.024 (0.013-0.043)
16p13.1 dup	Kirov	3/471 (0.00637)	6/2792 (0.00215)	0.023 (0.006-0.074)	0.023 (0.006-0.074)	0.023 (0.006-0.074)
16p11.2 dup	McCarthy	26/8590 (0.00303)	8/28406 (0.00028)	0.069 (0.035-0.144)	0.071 (0.035-0.149)	0.073 (0.036-0.156)
17p12 del	Kirov*	8/5089 (0.00157)	6/38884 (0.00015)	0.067 (0.026-0.170)	0.069 (0.026-0.175)	0.071 (0.027-0.181)
22q11 del (VCFS)	Stefansson	8/3838 (0.00208)	0/39299 -	0.481 (0.128-0.963)	0.479 (0.128-0.963)	0.478 (0.127-0.963)
22q11 del (VCFS)	ISC	13/3391 (0.00383)	0/3181 -	0.118 (0.022-0.790)	0.118 (0.022-0.788)	0.117 (0.022-0.787)
22q11 del (VCFS)	Kirov*	18/7038 (0.00256)	0/44602 -	0.553 (0.177-0.972)	0.552 (0.176-0.972)	0.550 (0.176-0.971)

*Kirov's study reports combined observations from Stefansson, ISC and new data.

CI: Credible intervals, *L*: Proportion of controls that have been screened for schizophrenia.

[*L*=1 corresponds to results presented in Table 1 in main paper.]