Penetrance for copy number variants associated with schizophrenia

Evangelos Vassos¹,²,*, David A. Collier¹, Simon Holden³, Christine Patch³,⁴, Dan Rujescu⁵, David St Clair⁶ and Cathryn M. Lewis¹,⁷

¹Institute of Psychiatry, MRC SGDP Centre and ²Institute of Psychiatry, Division of Psychological Medicine and Psychiatry, King’s College London, SE5 8AF, UK, ³Clinical Genetics and ⁴NIHR, Biomedical Research Centre, Guy’s and St Thomas’ NHS Foundation Trust and King’s College London, UK, ⁵Department of Psychiatry, Ludwig-Maximilians-University, Nußbaumstrasse 7, 80336 Munich, Germany, ⁶Department of Mental Health, University of Aberdeen, Royal Cornhill Hospital, Aberdeen AB25 2ZD, UK and ⁷Department of Medical and Molecular Genetics, King’s College London, SE1 9RT, UK

Received March 22, 2010; Revised June 10, 2010; Accepted June 21, 2010

The discovery of ‘high-risk’ de novo copy number variants (CNVs) associated with neuropsychiatric disorders such as schizophrenia offers the opportunity to translate these findings into useful tools for clinical geneticists. However, this will require estimation of penetrance for these variants, which has not yet been properly considered. To facilitate this process, we estimated the penetrance of CNVs associated with schizophrenia, at 15q13.3, 1q21.1, 15q11.2, 17p12, 2p16.3, 16p13.1 and 16p11.2 with a novel Bayesian method applied to pooled data from published case–control studies. For these CNVs, penetrance for schizophrenia was between 2 and 7.4%, which contrasts with the much higher penetrance for schizophrenia of the 22q11.2 deletions found in velo-cardio-facial syndrome. The highest penetrance was for 15q13.3 deletion (6–9% in individual studies) and the lowest was for 15q11.2 (2%). CNVs confer much higher risk for schizophrenia than common variants, but their penetrance is substantially lower than Mendelian disorders or other syndromic conditions. Since these CNVs predispose to multiple disorders, including epilepsy, autism and intellectual impairment, penetrance estimates will also need to take into account diagnostic specificity, and their overall penetrance for any neuropsychiatric disorder is likely to be much higher. Thus, although CNVs are still far from being clinically useful or relevant to genetic counselling for specific disorders, their detection may hold an important clinical value in predicting negative developmental outcomes.

INTRODUCTION

Many medical disciplines have been able to benefit from recent technological developments in genetics, and are developing pathways for personalized patient care. Recent progress in the genetics and genomics of psychiatric disorders such as autism and schizophrenia has been impressive, and it is timely to assess the utility of these advances in clinically relevant applications such as diagnostics and genetic counselling. In particular, there has been a flurry of excitement at the emergence of ‘high-risk’ de novo copy number variants (CNVs), a type of genomic variation in which segments of DNA of more than 1000 bp are duplicated or deleted, as genomic risk factor for common complex brain disorders, including schizophrenia, autism and mental retardation (1–4). One CNV, a deletion of ~3 Mb, on chromosome 22, which causes the 22q11 deletion syndrome (velo-cardio-facial syndrome or VCFS), is well established as a risk factor for schizophrenia and other neuropsychiatric phenotypes (5). More recent findings demonstrate that VCFS is not unique and provide strong support for a model of schizophrenia that includes an excess of rare, de novo CNVs across the genome as a whole (6,7), as well as associations at specific loci, including 1q21.1, 15q11.2, 15q13.3, 17p12, 16p11.2, 16p13.1 and the neurexin1 gene (2,3,8–13). These CNVs are not disorder specific, as they can give rise to a range of phenotypes, from language disorder

*To whom correspondence should be addressed at: Institute of Psychiatry, PO Box P063, De Crespigny Park, London SE5 8AF, UK. Tel: +44 2078480592; Fax: +44 2078480287; Email: evangelos.vassos@iop.kcl.ac.uk

© The Author 2010. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org
to epilepsy (4,14). The pleiotropic effects of common CNVs challenge psychiatric diagnoses and current classification systems (15).

These pathogenic CNVs tend to be rare in the population: deletions at 1q21.1, 15q13.3 and 17p12 occur in ~1 in 500 patients with schizophrenia compared with fewer than 1 in 5000 controls without neuropsychiatric illness, whereas the commoner 2p16.3 and 15q11.2 deletions occur in ~1 in 200 patients and 1 in 500 controls. Duplications at 16p13.1 and 16p11.2 occur in ~1 in 300 cases compared with 1 in 1000 and 1 in 3500 controls, respectively. The odds ratio for risk is high; CNVs at 1q21.1, 15q13.3, 17p12 and 16p11.2 confer an approximate 10-fold increased risk of schizophrenia. This is in contrast to the expected pattern of genetic susceptibility for common, complex disorders such as schizophrenia, involving many common, small effect variants: the ‘common disease—common variant’ model. Over 1000 genetic association studies have been published in schizophrenia based on this model, with largely inconsistent results, but a recent meta-analysis has identified 16 genes with nominally significant effects and an average pooled odds ratio of 1.23 (16). There are also examples of susceptibility loci emerging from genome-wide association, in the major histocompatibility complex region, and at the ZNF804a, neuregranin and TCF4 genes (17–20), all conferring between 1.1- and 1.24-fold increased risk, which is typical for low-risk variants of common complex diseases (21). Estimates from one of these studies (19) suggests that common, low-risk variants may account collectively for at least one-third of the genetic risk for schizophrenia, but high-risk CNVs and other rare variants are also likely to play a significant role in the susceptibility.

Genomic rearrangements have long been associated with rare genetic syndromes that include psychiatric manifestations, most notably VCFS for which the penetrance for schizophrenia is estimated to be 25% (22). For autistic spectrum disorders, penetrance is >40% in 15q duplication, Angelman syndrome, and 90% in Potocki–Lupski syndrome (23). However, unlike clinical geneticists dealing with penetrant disorders caused by high-risk variants, complex disease geneticists do not usually consider the issue of penetrance, as for common, low-risk alleles this will be vanishingly small. The larger odds ratios of de novo CNVs (3,24,25) have potential implications in the conceptualization of novel disease models, diagnosis and genetic counselling, because they are expected to be more predictive of disease status than common variants. Even though causal relationship to psychiatric disorders has not been proved so far, CNVs can be of particular interest to clinical geneticists as the use of genome-wide arrays CGH tests, which are able to routinely detect them, is now commonplace in clinical genetic settings (26). However, the clinical utility of the recent findings will require proper estimation of penetrance, especially for use in genetic counselling and risk prediction.

RESULTS

Data from six studies were used in the estimation of penetrance for CNVs associated with schizophrenia (2,3,9–12). Our method shows that the penetrance is moderate, at between 2 and 7.4% for most CNVs and most data sets, with 95% credible intervals from the larger data sets, indicating that penetrance is unlikely to be >20% (Table 1 and Supplementary Material, Table S1). These penetrance estimates are much lower than for syndromic conditions with psychiatric manifestations: for VCFS, we estimate penetrance to be 55%, based on data from pooled studies, although credible intervals are broad, as no CNVs are observed in control cohorts. The distributions for penetrance for other CNVs are tight, indicating that data provide good information for an accurate estimation of penetrance.

Since the size and the number of disrupted genes differed between the seven CNVs studied, we examined potential association of these factors with our penetrance estimates. We did not observe significant association of penetrance with the CNV size ($P = 0.61$) nor with the number of genes affected ($P = 0.30$).

Studies of these CNVs in other neuropsychiatric disorders (1,4,14,27–31) used diversely ascertained clinical samples and complex co-morbid phenotypes, such as autism plus dysmorphic features or intellectual impairment plus epilepsy. Thus, estimates of prevalence for the case ascertainment criteria are unavailable, and we were unable to estimate CNV penetrance for these disorders (Table 2). CNVs associated with both schizophrenia and other neuropsychiatric disorders have similar frequency in cases across disorders (between 0.2 and 0.6%), with the exception of the 15q13.3 deletion, which was more common (1%) in cases with idiopathic generalized epilepsy (28). Since these CNVs are not disorder specific, future estimates of risk will need to take into account not only the penetrance for each individual disorder, but also the overall risk of developing any neuropsychiatric phenotype.

DISCUSSION

On the basis of our penetrance estimations, we can conceptualize CNVs as genetic variants that bridge the gap between highly penetrant mutations in Mendelian, single-gene diseases and the common low-risk genetic variants typically associated with complex genetic disorders. The CNVs we describe are neither necessary nor sufficient for the development of a neuropsychiatric disorder, but they substantially increase the risk of disease. The variable expressivity and incomplete penetrance suggest that their impact is modified by other genetic loci or environmental factors (32). The level of penetrance of the CNVs described in this study is not sufficient for them to be considered as useful clinical tools in genetic counselling and testing, as well as in diagnosis at present. Before this can be implemented, more detailed information on the factors that modulate penetrance and diagnostic specificity are required. If these parameters can be established, then specific benefits to patients would include (i) improved diagnostic validity, (ii) enhanced prognostic accuracy, (iii) personalized care planning and (iv) individualized risk assessment and genetic screening/counselling.

Modulation of the penetrance of CNV deletions could involve a number of factors, such as the exposure of recessive alleles on the intact chromosome because of the
resulting hemizygosity, the effects of genetic background, environmental or epigenetic factors. Penetration of a specific CNV may depend on the functionality of the genes within the locus, with loci containing dose-sensitive critical genes being less tolerant to any disruption and having greater penetrance (33). For example, in HFE-related haemochromatosis, mediation of the penetrance appears to result from concomitant mutations in other genes that influence iron metabolism (34). Likewise, in a mouse model of 22q11.2 deletion syndrome, genetic background also appears to have a substantial effect on the penetrance of various malformations seen in the syndrome; mutations in Crkl, which lies within 22q11.2, and in Fgf8 and chordin, which lie elsewhere, both modulate the developmental phenotype by enhancing the effects of a Tbx1 null mutation on craniofacial abnormalities (35–37). Thus, for CNVs predisposing to neuropsychiatric disorders, it should be possible to identify modifier genes through the careful analysis of related biological pathways and the use of mouse models. The clinical utility of these CNVs will be enhanced by further understanding of their interactions with common genetic variants, and the roles of environmental and epigenetic risk factors.

A limitation of our study is that the accuracy of the penetrance estimation depends on the quality of the original studies. Although the associations with schizophrenia are almost certainly robust, given the rarity of CNVs their frequency estimates still lack precision, despite the large sample sizes. With the current detection methods it is difficult to define accurately the CNV breakpoints. The studies reviewed identified CNVs with dosage detection methods using the intensities from single nucleotide polymorphism probes on Illumina Bead-Arrays or Affymetrix Gene-Chips and provide adequate quality control evaluation. Since these detection methods have low resolution and potential issues with measurement precision, the CNVs identified may be an underestimate of the total number of CNVs. However, if there is no systematic bias in the detection of CNVs (i.e. use of same detection method in both cases and controls), a proportional inflation of CNVs would not substantially change frequency estimates still lack precision, despite the large sample sizes. With the current detection methods it is difficult to define accurately the CNV breakpoints. The studies reviewed identified CNVs with dosage detection methods using the intensities from single nucleotide polymorphism probes on Illumina Bead-Arrays or Affymetrix Gene-Chips and provide adequate quality control evaluation. Since these detection methods have low resolution and potential issues with measurement precision, the CNVs identified may be an underestimate of the total number of CNVs. However, if there is no systematic bias in the detection of CNVs (i.e. use of same detection method in both cases and controls), a proportional inflation of CNVs would not substantially change the penetrance estimations.

Another limitation is that the penetrance estimation depends on the characteristics of cases and controls used in the original studies. One consideration is whether cases are representative of the clinical population as a whole. Cases in research studies are usually screened with structured interviews and probably
have ‘cleaner’ phenotypes than patients in clinical practice. If
the cases included in the studies represent the more severe end
of the spectrum, one may hypothetize that penetrance is over-
estimated. On the other hand, given the pleiotropic effects of
CNVs, exclusion of cases with congenital malformations,
learning disability or neurological symptoms would have led
to underestimation of penetrance. In all studies, cases fulfilled
DSM-IV or ICD10 criteria for schizophrenia and explicit
description of phenotypic characteristics of cases with
CNVs, when available, were comparable with the general
schizophrenic patient population. As regard to controls, in
some samples, they were screened to exclude schizophrenia,
while in others, controls were drawn from the population
without psychiatric screening and possibly include some
schizophrenia cases. We repeated the analyses assuming popu-
lation sampling for controls, and obtained very similar pene-
trance estimates (data presented in Supplementary Material,
Table S1).

Patients with schizophrenia and their family members may
want to avail themselves of genetic counselling and require
more information about the risk of developing schizophrenia
(38). The CNVs identified to date account for only a small pro-
portion of genetic susceptibility to schizophrenia, but many
more are likely to be discovered over the next few years
(25). Ideally, to overcome the limitations of case ascertain-
ment and to be able to explore the whole spectrum of pheno-
types associated with CNVs, future studies should investigate
large birth cohorts followed up prospectively beyond the risk
period for the development of associated phenotypes. Since
the predictive value of genetic variants depends on their pene-
trance, our study shows that the recently discovered CNVs are
still far from being clinically useful or relevant to genetic
counselling for specific disorders. CNVs are neither necessary
nor sufficient for specific diseases and current evidence does
not support the referral of patients with schizophrenia for
genetic screening. However, once better characterized and
more precisely associated with diseases, their detection may
hold important clinical value in predicting negative develop-
mental outcomes. They may be useful in contributing to the
formulation of empiric risk models and perhaps, more impor-
tantly, have a role in the understanding of genetic and biologi-
cal processes underpinning the development of schizophrenia
and other neuropsychiatric disorders.

MATERIALS AND METHODS

Data sources

We identified all published case–control studies that investi-
gate the association of CNVs with schizophrenia and other
neuropsychiatric disorders. For each study, we extracted the
inclusion criteria for cases and controls, the number of
observed CNVs and the total number of participants in each
cohort. Criteria for study selection for our analyses were the
association of CNVs at specific loci (as opposed to measure-
ment of the collective burden of rare CNVs) and the presence
of more than one observation of each CNV in cases. For
studies with overlapping samples, we analysed data from the
larger study and for studies that included meta-analyses we
used the pooled numbers. The sizes and location of CNVs
were variable and the end points differed between studies.
For our analysis, we took the straightforward approach of
counting all CNVs in each locus as equivalent events, irre-
respect of their size and exact position.

Statistical analysis

Using the observed CNV frequencies from published data, we
employed a Bayesian approach to derive posterior distribu-
tions of likely values for the frequency of CNVs in case
and control cohorts. We then sampled pairs of CNV frequen-
cies from these distributions, and calculated the penetrance,
i.e. the probability of developing schizophrenia (disease, \(D\))
for individuals carrying the CNV (genotype, \(G\)), from

\[
P(D|G) = \frac{P(G|D)P(D)}{P(G|D)P(D) + P(G|\overline{D})P(\overline{D})},
\]

where \(\overline{D}\) represents controls, who do not have schizophrenia,
and \(P(D)\) the lifetime morbid risk for schizophrenia. 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 comprehensive meta-analysis
(39). 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
≈95% credible intervals. Complete details of the statistical
methods are given as Supplementary Material.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG online.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by EU grant LSHM-CT-2006-
037761 (Project SGENE), Guy’s and St Thomas’ Charity
grant, and NARSAD Young Investigator Award (E.V.). The
author’s work was independent from the funders.

REFERENCES

   Association between microdeletion and microduplication at 16p11.2 and
2. Stefansson, H., Rujescu, D., Cichon, S., Pietilainen, O.P., Ingason, A.,
   Steinberg, S., Fossdal, R., Sigurdsson, E., Sigmundsson, T.,
   Buizer-Voskamp, J.E. et al. (2008) Large recurrent microdeletions
3. ISC (2008) Rare chromosomal deletions and duplications increase risk
   rearrangements of chromosome 1q21.1 and variable pediatric phenotypes.
   Psychiatry, 56, 940–945.
17. Stefansson, H., Ophoff, R.A., Steinberg, S., Andreassen, O.A., Cichon, S.,
    et al. (2008) Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. Science, 320, 539–543.
    comparative genome hybridization suggests a role for NRXN1 and
13. McCarthy, S.E., Makarov, V., Kirov, G., Addington, A.M., McClellan, J.,
    et al. (2009) Support for the involvement of large copy number variants in the
10. Rujescu, D., Ingason, A., Cichon, S., Pietilainen, O.P., Barnes, M.R.,
    Touloupidou, T., Picchioni, M., Vassos, E., Ettinger, U., Bramon, E. et al. (2009)
    Disruption of the neurexin 1 gene is associated with schizophrenia. Hum. Mol. Genet., 18, 988–996.
9. Kirov, G., Grozeva, D., Norton, N., Ivanov, D., Mantrpragada, K.K.,
    Holmans, P., Craddock, N., Owen, M.J. et al. and International
    Schizophrenia Consortium, the Wellcome Trust Case Control Consortium
    (2009) Support for the involvement of large copy number variants in the
8. Vrijenhoek, T., Buizer-Voskamp, J.E., van der Stelt, I., Strengman, E.,
    Sabatti, C., Geurts van Kessel, A., Brunner, H.G., Ophoff, R.A. and
13. McCarthy, S.E., Makarov, V., Kirov, G., Addington, A.M., McClellan, J.,
    et al. (2009) Support for the involvement of large copy number variants in the
10. Rujescu, D., Ingason, A., Cichon, S., Pietilainen, O.P., Barnes, M.R.,
    Touloupidou, T., Picchioni, M., Vassos, E., Ettinger, U., Bramon, E. et al. (2009)
    Disruption of the neurexin 1 gene is associated with schizophrenia. Hum. Mol. Genet., 18, 988–996.
9. Kirov, G., Grozeva, D., Norton, N., Ivanov, D., Mantrpragada, K.K.,
    Holmans, P., Craddock, N., Owen, M.J. et al. and International
    Schizophrenia Consortium, the Wellcome Trust Case Control Consortium
    (2009) Support for the involvement of large copy number variants in the
8. Vrijenhoek, T., Buizer-Voskamp, J.E., van der Stelt, I., Strengman, E.,
    Sabatti, C., Geurts van Kessel, A., Brunner, H.G., Ophoff, R.A. and
13. McCarthy, S.E., Makarov, V., Kirov, G., Addington, A.M., McClellan, J.,
    et al. (2009) Support for the involvement of large copy number variants in the
10. Rujescu, D., Ingason, A., Cichon, S., Pietilainen, O.P., Barnes, M.R.,
    Touloupidou, T., Picchioni, M., Vassos, E., Ettinger, U., Bramon, E. et al. (2009)
    Disruption of the neurexin 1 gene is associated with schizophrenia. Hum. Mol. Genet., 18, 988–996.
9. Kirov, G., Grozeva, D., Norton, N., Ivanov, D., Mantrpragada, K.K.,
    Holmans, P., Craddock, N., Owen, M.J. et al. and International
    Schizophrenia Consortium, the Wellcome Trust Case Control Consortium
    (2009) Support for the involvement of large copy number variants in the
8. Vrijenhoek, T., Buizer-Voskamp, J.E., van der Stelt, I., Strengman, E.,
    Sabatti, C., Geurts van Kessel, A., Brunner, H.G., Ophoff, R.A. and
13. McCarthy, S.E., Makarov, V., Kirov, G., Addington, A.M., McClellan, J.,
    et al. (2009) Support for the involvement of large copy number variants in the
10. Rujescu, D., Ingason, A., Cichon, S., Pietilainen, O.P., Barnes, M.R.,
    Touloupidou, T., Picchioni, M., Vassos, E., Ettinger, U., Bramon, E. et al. (2009)
    Disruption of the neurexin 1 gene is associated with schizophrenia. Hum. Mol. Genet., 18, 988–996.
9. Kirov, G., Grozeva, D., Norton, N., Ivanov, D., Mantrpragada, K.K.,
    Holmans, P., Craddock, N., Owen, M.J. et al. and International
    Schizophrenia Consortium, the Wellcome Trust Case Control Consortium
    (2009) Support for the involvement of large copy number variants in the
8. Vrijenhoek, T., Buizer-Voskamp, J.E., van der Stelt, I., Strengman, E.,
    Sabatti, C., Geurts van Kessel, A., Brunner, H.G., Ophoff, R.A. and
13. McCarthy, S.E., Makarov, V., Kirov, G., Addington, A.M., McClellan, J.,
    et al. (2009) Support for the involvement of large copy number variants in the
10. Rujescu, D., Ingason, A., Cichon, S., Pietilainen, O.P., Barnes, M.R.,
    Touloupidou, T., Picchioni, M., Vassos, E., Ettinger, U., Bramon, E. et al. (2009)
    Disruption of the neurexin 1 gene is associated with schizophrenia. Hum. Mol. Genet., 18, 988–996.
9. Kirov, G., Grozeva, D., Norton, N., Ivanov, D., Mantrpragada, K.K.,
    Holmans, P., Craddock, N., Owen, M.J. et al. and International
    Schizophrenia Consortium, the Wellcome Trust Case Control Consortium
    (2009) Support for the involvement of large copy number variants in the
8. Vrijenhoek, T., Buizer-Voskamp, J.E., van der Stelt, I., Strengman, E.,
    Sabatti, C., Geurts van Kessel, A., Brunner, H.G., Ophoff, R.A. and