A new wave in the genetics of psychiatric disorders: the copy number variant tsunami

Ridha Joober, MD, PhD; Patricia Boksa, PhD

Joober, Boksa — Department of Psychiatry, McGill University and the Douglas Mental Health University Institute, Montréal, Que.

Because genes are major determinants of psychiatric disorders, our community is hoping that, as in many other branches of medicine, molecular genetics will lead to important findings that will revolutionize our understanding of these disorders and ultimately improve prevention, diagnosis and treatments. From the early days of simple genetic association studies with phenotypic or molecular markers (blindness or blood groups) to the most recent and sophisticated genome-wide association studies, the amount of genetic psychiatric data generated has been increasing drastically. A quick PubMed search indicates that roughly 1 of every 6 papers published on schizophrenia, bipolar disorder or autism refers to genetics. Two decades ago, this ratio was 1 of 8, and 3 decades ago it was 1 of 10.

Over the years, the trend in genetic studies has been to use much larger sample sizes (several thousand patients and controls are being assembled for genome-wide association studies) and to genotype a very large number of markers (5000–10 000). This is because the genetic effects that we are trying to pin down are presumably conferred by common and numerous genetic variants (polygenes), each increasing the risk by a very modest factor (relative risk of about 1.1–1.2). This common disease/common variant scenario is where most of the efforts have been directed in recent years. Remarkably, this worked very well for many somatic complex diseases, such as type II diabetes and inflammatory bowel disease, where many loci have now been confirmed as risk factors with a very comfortable confidence. Whether or not this will work for psychiatric disorders remains to be tested. The success will depend on the magnitude of many other parameters that are not well understood: environmental effects, diagnostic uncertainties, pathogenic heterogeneity and bias. Our ability to detect common genetic variants with modest effects will be highly dependent on the magnitude of these factors.

In contrast to common genetics variants, there are rare and highly penetrant mutations. These are mutations that are sufficient to cause the disease regardless of whether or not other factors are present. It is these Mendelian mutations that allowed us to infer with extraordinary clarity that genetic alterations can cause particular diseases (e.g., cystic fibrosis, Huntington disease). In the last few years, the field of psychiatric genetics turned away from this paradigm, mainly because linkage results in psychiatric disorders were difficult to replicate. However, recent new developments in the field are re-ving the notion of genes as potent determinants of psychiatric disorders, largely independent of other factors, and we are now witnessing an interesting debate between the common disease/common variant and the Mendelian proponents.

It has been known for a long time that chromosomal aneuploidies (chromosome number difference compared with the normal 46 XX or 46 XY), segmental aneusomies (chromosomal deletions and duplications) or rearrangements (balanced or unbalanced translocations, inversions, etc.) are associated with various developmental disorders. The recurrence of the same chromosomal abnormality in association with specific morphologic and/or behavioural features allowed the description of thousands of developmental syndromes. Some of these abnormalities have been associated with psychiatric disorders. The most well-known of these chromosomal abnormalities in psychiatry is a deletion in the long arm of chromosome 22 (22q11–12), because it has often been associated with psychotic, attention and mood symptoms. Reports of chromosomal anomalies associated with pervasive developmental disorders, particularly syndromic cases, are also very common, especially on chromosome 15. In most cases, these abnormalities are de novo mutations, meaning mutations that happened in the germ cells of either of the parents. The recurrence of these events in specific locations of the genome are often due to genomic architectural features rendering a segment of the genome unstable, and the resulting disorders are referred to as genomic disorders.

Recently, technological advances, mainly in the form of genomic hybridization arrays, allowed cytogeneticists to go beyond the microscopic level of resolution and to detect chromosomal rearrangements ever smaller than before. In
fact, one of the most profound insights we have glimpsed from the human genome project is how our genome, although much less rich in genes (25 000) than initially expected, is extremely complex with regard to its structure and its variations. A recent finding by Redon and colleagues published in 2006 in the journal Nature identified 1447 copy number variable (CNV) regions encompassing 12% of the human genome. A CNV is operationally defined as a segment of DNA, of 1000 base-pairs or more, with variable copy number compared with the reference genome. Redon and colleagues found that CNV regions contained hundreds of genes, disease loci and functional elements. In fact, it has been shown that these variations are 2 to 3 times more important in scope than the single nucleotide polymorphisms (SNP) that are used in genome-wide association studies. The high mutant rate of CNVs (1.7 × 10^-8 – 1.2 × 10^-8) compared with single nucleotide mutations (2 × 10^-9) makes CNVs, particularly recurrent ones, potential important contributors to diseases with relatively high frequencies, as is the case for most psychiatric disorders. Thus it is not surprising that many authors have focused on the topic of CNVs and psychiatric disorders in recent years.

In a landmark paper, Sebat and colleagues compared the frequency of de novo CNVs in children with autism (n = 118) and normally developing children (n = 196). They found de novo CNVs in 10% of children with autism compared with 1% of normally developing children. These mutations were mostly submicroscopic, affected various genomic regions, and some of them affected one single gene. Interestingly, each particular mutation was observed only once, thus potentially explaining singular cases. Some of the genes located in these CNVs (e.g., SHANK3) have, in fact, been identified as potential candidates for autism because they were found in previous studies to harbour de novo and inherited mutations in children with autism. In an independent study, the Autism Genome Project Consortium has also reported CNV data from the largest ever conducted linkage study of autism. In this study, 254 CNVs were detected in 196 patients with autism (1.29 CNV per patient), and 370 CNVs were detected in 292 unaffected individuals (1.25 CNV per person). Among the CNVs identified in patients with autism, 10 (about 5%) were de novo CNVs. Here again, it was reported that some genes (e.g., neurexin 1) located in some of the discovered CNVs interact with genes previously implicated in autism. The authors of this study focused their attention mainly on the CNVs discovered in autistic patients to highlight their potential pathogenicity. However, they also reported that some of the CNVs identified in patients with autism have also been identified in their nonaffected siblings, which weakens the involvement of CNVs in autism. No information was reported on gene content of CNVs that were present exclusively in nonaffected individuals, precluding a proper evaluation of the specificity and sensitivity of these events in autism.

Finally, a third independent study identified a recurrent microdeletion and its reciprocal microduplication in 4 different samples of children with autism spectrum disorder at a frequency of about 1%. The deletion, located on 16p11.2, was about 600 Kb in size and was observed in only 2 of 18 834 control participants from Iceland who were not screened for mental disorders, indicating that the deletion is 100 times more frequent in developmentally impaired children than controls. It is of note that this deletion was observed in other psychiatric disorders including bipolar disorder (1/420), attention-deficit hyperactivity disorder (1/203), schizophrenia (1/648),
The copy number variant

dyslexia (1/748) and anxiety, panic, depression or addiction (1/3000). Also of note, none of the CNVs reported in the studies by Sebat and colleagues14 and by Szatmari and colleagues28 was found to recur in this sample.

Four major studies of CNV and schizophrenia have been reported in the last few months, the first in the journal Science, the second in Nature Genetics and the third and fourth in Nature.13–22 Walsh and colleagues20 reported that rare (i.e., not described in the literature or in the specialized databases as of November 2007) chromosomal structural variants may be highly penetrant events each explaining a few, if not singular, cases of schizophrenia. They reported a significant 3-fold increase (22/150 v. 13/268) in the rate of these CNVs in patients compared with controls. Further, they found that the rate of these CNVs is significantly higher in chromosomal transmitted from parents to their children affected with childhood onset schizophrenia compared with the rate of these CNVs in the nontransmitted chromosomes. In this study, the authors identified only 2 de novo CNVs (2.4%) in the childhood onset schizophrenia sample. In addition, at least 56% of parents in this sample had CNVs, and it was not reported whether the rate of schizophrenia was proportionally high in these parents, as might be expected under the high penetrance hypothesis. In a more recent paper, Xu and colleagues22 reported a significant (p < 0.001) 8-fold increase in de novo CNVs in a sample of Afrikaner patients with sporadic schizophrenia (15/152) compared with a sample of racially matched controls (2/159). No such increase was identified in familial cases. Contrary to the study by Walsh and colleagues,20 the rate of inherited CNVs was not different in patients versus controls (46/152 v. 32/159). Two more recent studies appeared back to back in the journal Nature. The first, by Stone and colleagues21 indicates that the total burden of CNVs that are rare (less than 1% in frequency and more than 100 Kb in length) and disturb genes are slightly (1.15-fold) but significantly increased in patients with schizophrenia compared with controls. In addition to confirming the higher frequency of the 22q11 deletion in patients and controls, 2 deletions were found to recur more frequently in patients compared with controls: 1 on 15q13.3 (p = 0.046) and 1 on 1q21.1 (p = 0.046). In the second study, Stefansson and colleagues22 focussed on recurrent de novo CNV mutations, because reduced fertility in major psychiatric disorders would exert a negative pressure on causative mutations. The authors hypothesized that such mutations would be highly relevant for schizophrenia (and possibly other psychiatric disorders) since the persistence of schizophrenia at relatively high levels in the population in spite of the reduced fertility may be explained, at least in part, by the recurrence of these mutations. In a large population control sample, they identified 66 de novo CNVs with very low frequency compared with their theoretical expected frequency. These CNVs were thus considered to be subjected to negative selection and were compared in a sample of 1433 patients with schizophrenia and 33 250 controls (assembled from various other genetic studies). Three chromosomal locations were found to harbor CNVs more frequently in patients with schizophrenia and related psychoses compared with controls: 1q21.1, 15q11.2 and 15q13.3, thus confirming the results of the study by Stone and colleagues.21 However, when a strict diagnosis of schizophrenia was applied, only the CNV located on 1q21.1 (0.23%) remained significant compared with controls (0.02%, p < 0.001). This study also confirmed a high rate of the 22q11 deletion in patients with schizophrenia.

More recently, in a large cohort of pediatric patients with developmental problems (n = 5218), it was reported that chromosomal rearrangement in 1q21.1 was associated with a large array of phenotypic expression ranging from mild to moderate mental retardation, autism spectrum manifestations, cardiac abnormalities, cataract and other abnormal developmental features.23 Similarly, chromosomal rearrangements in the 15q13.2q13.3 region have been recently associated with autism spectrum disorders, cognitive impairments, learning disabilities, anxiety and mood disorders in a sample of 1445 children with various developmental delays and 1441 children with autism spectrum disorders.24 In both studies, the clinical characteristics of carriers of these mutations were reported, and none had psychotic features. Thus these recent studies question the specificity of the association between the 1q21.1 and 15q13.3 rearrangements and schizophrenia.

In addition to these major studies published in high-impact journals (Nature and Science), other investigations of CNVs in schizophrenia and autism have been reported with various designs and various results.2–5

This new wave of genetic studies posits that major psychiatric disorders could be caused by highly penetrant and very rare (possibly singular) genetic events (inherited or not) in a substantial proportion of cases. It is hoped that with increasing resolution of the technology, more and more cases will be attributable to such molecular events.

Both genome-wide association studies and CNV studies query only a small fraction of the human genome, and most of the genome sequence is not investigated. Ideally, one would aspire to sequence the entire human genome in patients and controls to gain the most insight on the role of genes in mental illnesses. This approach of deep sequencing is probably not far, because 2 customized human genomes (the genome of the double helix discoverer, James Watson,25 and that of the genomic leader, Greg Venter26) have recently been completed. Such an approach applied to a large number of patients and controls would certainly provide a tremendous advantage to both genome-wide association studies that may detect the common variants underlying psychiatric disorders and CNV studies that may lead to the identification of a number of genomic disorders clustering within each major psychiatric syndrome. Although still not yet feasible, the proof of concept of this approach is being established. Indeed, a number of deep sequencing studies in selected genes have identified mutations that cosegregate notably with autism. The most prominent candidates are SHANK3 and NLG3–4, which were corroborated to a certain extent in the CNV studies. Other research groups have selected thousands of genes (about 1500) that are important for synaptic development, maintenance and function for deep sequencing in hundreds of patients with schizophrenia, autism or mental retardation.27 It is expected that this approach will lead to the
identification of very rare, highly penetrant mutations that will explain a few cases of these diseases. If any of these mutations turn out to be recurrent in even a few patients excluding controls, this will again define a genetic sub-syndrome within a given psychiatric disorder.

It took the field of psychiatric genetics 3 decades and tens of thousands of publications to come to the conclusion that the complexity underlying the genetics of these disorders cannot be solved by 1 team or even several collaborating teams. Under the common disease/common variant hypothesis, it is possible that very large sample sizes (possibly in the hundreds of thousands) and commensurate analytical power may be needed to convincingly incriminate a polymorphism in increasing susceptibility to psychiatric disorders. This begs for international collaborative studies in which possibly all the samples collected need to be assembled in an effort to overcome the “dustiness” of genetic effects.

Researchers in the field of structural variants are relying on the accepted wisdom that structural variants are possibly more penetrant than smaller polymorphisms and that there are some low hanging fruits that can be easily harvested. Although this might be true for some CNVs, the few published studies usher a number of problems that need to be addressed, the chief of them being the claim of high penetrance. The best test for high penetrance is familial cosegregation of the purported mutations and the disease. This test should be conducted whenever familial information is available. Walsh and colleagues found that, where familial information was available in the sample of childhood onset schizophrenia, at least 18 structural variants were transmitted and 10 were nontransmitted. Adding transmitted and non-transmitted structural variants will lead to a very large proportion (more than 36%) of parents carrying structural vari- ants presumed to be highly penetrant. Unfortunately, no information on the disease status of these parents was reported (or provided after personal request), which precluded testing the claim of high penetrance. For recurrent mutations subject to very strong negative selection, the argument of cosegregation will be very hard to make owing to reduced fertility. In these cases, demonstrating causation (high penetrance) will require much more than statistical association, as outlined in Austin Bradford Hill’s well-known criteria of causation for epidemiologic research.

In addition to the fact that association studies do not prove causality, they are notoriously known for their sensitivity to various biases and confounders. It is notable that all CNV studies published to date were conducted with samples that were not collected for the specific purpose of testing the role of CNVs in mental disorders; rather, they were spin-offs of previous genetic studies designed for other purposes. This approach could be associated with potentially important biases or confounders. In our opinion, 1 major confounder that might have been operating in recent CNV studies is the lack of control for intelligence quotient (IQ) between patients and controls. Because chromosomal abnormalities and CNVs are very often associated with variable IQ deficits, the observed associations could be accounted for, at least in part, by differences in IQ between patients and controls. Consis-
whether these efforts will ultimately be fruitful or not and whether there is any “stopping rule” in this field. This remains an open question and it would require deep reflection and possible paradigmatic shifts before turning our backs on such a fundamental paradigm as genetics.

Competing interests: None declared.

Contributors: Both authors designed the study, acquired and analyzed the data, wrote and reviewed the article, and approved the final version for publication.

References

19. Walsh T, McClellan JM, McCarthy SE, et al. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. Science 2008;320:539-43.
35. Walsh T, McClellan JM, McCarthy SE, et al. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in large recurrent microdeletions associated with schizophrenia. Science 2008;320:539-43.