

# The 1000 Genomes Project: deep genomic sequencing waiting for deep psychiatric phenotyping

Ridha Joober, MD, PhD

Douglas Mental Health University Institute, Department of Psychiatry, McGill University, Montréal, Que.

Recently, the interim results of the 1000 Genomes Project were published and revealed that the genome of an apparently healthy individual contains hundreds of potentially harmful, rare variants occurring at various places among thousands of possible different loci.<sup>1</sup> Genetic variants are divided into 3 domains according to their frequency: very rare (< 1%), rare (1%–5%) and common variants (> 5%). This division is based on the assumption that the rarer a variant, the more strongly selected out of the gene pool. This selection pressure is due to a substantially reduced reproductive fitness that is associated with the phenotype caused by this variant.

Results from the last few decades of research in psychiatric genetics strongly suggest that genetic variants at the 2 extremes of the frequency spectrum (< 1% and > 5%) seem not to play a meaningful role in psychiatric disorders. Very rare mutations, which are most efficiently detected through reverse genetic mapping, have not been uncovered despite very serious efforts. Similarly, common variants that were the focus of recent genome-wide association studies appear not to greatly increase the risk for psychiatric disorders, although more work is needed in this field.<sup>2</sup> Even in somatic disorders, most of the heritability (> 95%) remains unexplained by common variants.<sup>3</sup> Thus it is hoped that many rare genetic variants with relatively high penetrance may capture most of the heritability, which is often greater than 40%, of psychiatric disorders.

## The 1000 Genomes Project

The primary purpose of the 1000 Genomes Project is to identify the largest number of rare genetic variants. In its interim analysis, the 1000 Genomes Project reported sequence results on the genomes of 2 families (1 Yoruba family from Ibadan, Nigeria, and 1 family from Utah, United States, with European ancestry), each comprising 2 parents and 1 daughter, and those of 60 nonrelated individuals from different races. In addition, 8104 exons from 906 randomly selected genes were entirely sequenced in 697 unrelated individuals. To date this

is the most intensive deep sequencing effort in so many individuals, revealing 14.4 million single-nucleotide polymorphisms (SNPs), 1.3 million short insertions/deletions (indels) and 20 000 large structural variants. Even more impressive is the number of genetic loci with variants potentially harmful to gene integrity in this otherwise unaffected sample: 714 small indels, 77 stop losses, 1057 stop-introducing SNPs, 517 splice site-disturbing SNPs, 954 small frameshift indels and 147 genes disturbed by large deletions. It was estimated that an individual genome differs from the reference human genome by 190–210 in-frame indels, 80–100 variants that lead to a premature stop codon (hence truncated proteins), 40–50 splice variant mutations and 220–250 frame shift variants. It was also estimated that each genome is heterozygous for 50–100 variants that could cause 1 or more inherited disorders from among those listed in the Human Gene Mutation database. These 530–610 genetic variations affecting the structure of proteins can, alone or through interacting effects, be harmful and cause various behavioural phenotypes. These estimates are bound to increase as more genomes are deeply sequenced and analyzed. How can this dormant load of harm inform the genetics of psychiatric disorders?

## Psychiatric genetic research

The response of the research community may be that every investigator who has studied patients with a specific disorder will explore a number of these potentially harmful variants; many variants will be discovered in a few patients, and results will be published. Of course, investigators will also search for these variants in other samples of patients with a variety of psychiatric disorders, and they will likely find variants and publish their results. Associations between rare variants and psychiatric phenotypes that have been published so far may reflect this trend. For example, 22q11 deletion was initially found to be overrepresented in patients with schizophrenia,<sup>4</sup> mood disorders,<sup>5</sup> anxiety disorders and attention-deficit/hyperactivity disorder,<sup>6</sup> pervasive developmental disorders<sup>7</sup>

**Correspondence to:** Dr. R. Joober, Douglas Hospital Research Centre, Frank B. Common Pavilion, 6875 Blvd. LaSalle, Verdun QC H4H 1R3; ridha.joober@douglas.mcgill.ca

*J Psychiatry Neurosci* 2011;36(3):147-9.

DOI: 10.1503/jpn.110026

and various levels of intellectual deficit.<sup>5,8,9</sup> The same pattern of indiscriminate association of many psychiatric phenotypes (e.g., mental retardation, schizophrenia, autism) and neurologic disorders (e.g., epilepsy) with chromosomal variants (e.g., translocation disrupting *DISC1*<sup>10–13</sup>) and copy number variations (CNVs)<sup>14–17</sup> is increasingly reported as investigators search for these rare variants in patients with various disorders or in carriers of a specific rare variant. Also, rare single point mutations that potentially disturb gene functions have been reported in several psychiatric phenotypes. For example, mutations in *SHANK3*, a gene important for synaptic integrity, were reported in patients with autism<sup>18</sup> and schizophrenia.<sup>19</sup> Given the large number of rare mutations with potential impact on brain development, there is no doubt that this field of research will result in a plethora of case reports or small series of case-control studies, each with a mutation associated with various phenotypes. This will amplify the already plethoric and controversial genetic association literature. However, because these genetic variants are rare, most of the studies, even those with large samples, will identify only a few participants (patients and controls) carrying these mutations, and results will be difficult to reproduce. One of the major problems in interpreting these associations is the question of whether they arise from an increased frequency of these genetic events in patients or from the underrepresentation of genetic events in controls. For example, given that intellectual deficiencies seem to be a common phenotype associated with rare variants, particularly CNVs,<sup>20</sup> any sample of controls that does not include participants with intellectual deficiencies will result in spurious apparent overrepresentation of these rare events in patients with psychiatric disorders. A testimony to this problem of recruiting appropriate participants is that the very large Icelandic control sample (33 250 participants) used in many recent CNV studies did not include participants with any 22q11 microdeletions,<sup>21</sup> despite a prevalence of this mutation of about 1 in 4000 in the general population. Thus, it is possible that published associations between various mental disorders and increased rates of CNVs might be driven by the confounding effect of IQ, which is often slightly but significantly reduced in many patients with major psychiatric disorders.<sup>22</sup> Supporting this hypothesis is the fact that bipolar disorder, which is not associated with a reduction in cognition,<sup>22</sup> has not been associated with an increased frequency of CNVs.<sup>23</sup>

## Genotypic and phenotypic mapping approaches

The question is how to take advantage of the incredibly rich knowledge gleaned from the 1000 Genomes Project without falling into the opportunistic “publish or perish” approach, which may mar the psychiatric genetics literature. I believe that particular attention to epidemiologic questions of sampling and phenotypic characterization is critical if this field is to advance. A systematic reverse phenotypic approach might be an adequate response.

Reverse phenotypic mapping is a concerted effort to collect a large, representative sample of the general population, regardless of psychiatric phenotype (Ph). Subsequently, the fre-

quency of affected and nonaffected participants is determined for specific genotype variants ( $G_i$ ) in each of the loci showing rare variations in a potentially harmful site; a significant association is identified when the probability of the phenotype given a specific genotype at the locus ( $Ph | G_i$ ) surpasses a statistically significant threshold. In other words, I propose to divide a population of people into 2 groups, 1 with a particular rare genetic variant and 1 without, and then determine the phenotypic differences between the 2 groups. In such an analysis, phenotypic differences would be defined in various ways using different approaches. For example, if an association were identified with a conglomerate psychiatric phenotype (i.e., all psychiatric disorders grouped together), a second round of fine phenotypic mapping would be conducted by excluding specific phenotypes, 1 at a time, until the best signal of association is identified. Of course, many sophisticated statistical approaches can be implemented to help in this reverse phenotypic mapping.

The opposite has been happening in traditional reverse genetic mapping approaches, which collect patients with specific phenotypes and use them to search the genotype space for genetic variants, such that the probability of  $G_i | Ph$  surpasses a certain threshold. Subsequently several rounds of fine genetic mapping are performed to identify the genetic variant that gives the highest signal. In this form of reverse genetic mapping, Mendelian phenotypes (those that segregate in families according to the laws of Mendel) represent robust phenotypic markers that are critical to reliably chart the genomic space and identify causative mutations. This approach failed in psychiatric genetics because psychiatric phenotypes are not appropriate phenotypic markers. Consequently, I propose that a reverse phenotypic mapping approach will take better advantage of our currently advanced genetic knowledge to chart the elusive psychiatric phenotypic space.

Of course, the question of the depth of phenotype characterization of randomly recruited participants is critical. Without deep phenotyping, deep sequencing may only lead to indiscriminate association of a large number of rare genetic variants with few psychiatric syndromes. Depth of phenotype characterization refers to the richness of charting the behavioural phenotype space using genetically pertinent phenotypic markers. However, current psychiatric classifications (that include only a handful of major psychiatric disorders) are probably not an adequate match to the thousands of rare genetic variants, nor do these psychiatric disorders show robustness as phenotypic markers (e.g., they do not show Mendelian segregation). Thus, it may be necessary to use other behavioural phenotypic markers to chart the rare variants space onto the behavioural space. Unfortunately, contrary to genetic markers that are naturally defined by their molecular structure, reliably assayed and stable over time, behavioural phenotypes are neither naturally segmented, easily and reliably defined nor stable over time. This will make this reverse phenotypic mapping much less obvious than reverse genetic mapping. Nevertheless, a reverse phenotypic mapping program may start using the current classifications of mental disorders (DSM or ICD) supplemented by neuropsychologic, neurophysiologic and behavioural traits as well as symptom dimensions and functional

outcomes that are amenable to high throughput phenotyping approaches. Ultimately, other phenotypes that are now being generated from various phenotypic initiatives<sup>24,25</sup> could be used and refined in subsequent reverse phenotypic mapping iterations. In doing so, the behavioural phenotype space can be reshaped using biologic anchors, namely rare genetic variants, which may turn out to be the “gold standard” needed to build a biologically rooted psychiatric classification. This approach contrasts sharply with reverse genetic mapping, which tries to find genes for psychiatric phenotypes derived on the basis of various criteria with little biologic relevance.

## Implications for future research

Assuming that there are about 3000 loci with rare variants in the human genome, that the average frequency of these rare variants is 2.5% and that the average relative risk associated with a risk variant is 2, a crude estimate of a sample size allowing detection of a risk variant associated with a conglomerate psychiatric phenotype with a 90% statistical power and a type-I error of  $p < 0.0001$  (allowing correction for multiple testing) is between 1200 and 1300 affected participants randomly sampled from the general population. Assuming that the prevalence of the aggregate psychiatric phenotype is 2.5%, we will need to identify 45 000 participants from the general population and characterize them with a relatively well-conceived battery of psychiatric phenotypes to answer systemically the question of the association of rare variants with behavioural phenotypes. For fine behavioural mapping, it is possible that 2 or 3 times this number will be needed to reshape these behavioural phenotypes. It is also possible to restrict deep phenotyping to subgroups of the sample whose members share rare variants that are believed to be highly pathogenic. This opportunistic approach may help identify some phenotypic regularities in some of these subgroups without the need for deep phenotyping the entire sample, thus making reverse phenotypic mapping more feasible, at least in its earliest stages.

These numbers may appear to be very high and should be more accurately calculated as a function of various parameters. The procedures proposed to perform deep phenotyping may seem quite complex and will certainly be very costly. However, we need only remember that billions of dollars and decades of concerted efforts have been dedicated to chart the human genome. Given the complexity of the human behavioural phenome, at least commensurate efforts and expenses are probably needed to accomplish this effort.

**Competing interests:** Dr. Joobar is on the advisory boards and speakers' bureaus of Pfizer Canada and Janssen Ortho Canada; he has received grant funding from them and from AstraZeneca. He has received honoraria from Janssen Ortho Canada for CME presentations and royalties for Henry Stewart talks.

## References

- Durbin RM, Abecasis GR, Altshuler DL, et al. A map of human genome variation from population-scale sequencing. *Nature* 2010;467:1061-73.
- Gershon ES, Alliey-Rodriguez N, Liu C. After GWAS: searching for genetic risk for schizophrenia and bipolar disorder. *Am J Psychiatry* 2011;168:253-6.
- Vineis P, Pearce N. Missing heritability in genome-wide association study research. *Nat Rev Genet* 2010;11:589.
- Lindsay EA, Morris MA, Gos A, et al. Schizophrenia and chromosomal deletions within 22q11.2. *Am J Hum Genet* 1995;56:1502-3.
- Green T, Gothelf D, Glaser B, et al. Psychiatric disorders and intellectual functioning throughout development in velocardiofacial (22q11.2 deletion) syndrome. *J Am Acad Child Adolesc Psychiatry* 2009;48:1060-8.
- Jolin EM, Weller RA, Weller EB. Occurrence of affective disorders compared to other psychiatric disorders in children and adolescents with 22q11.2 deletion syndrome. *J Affect Disord* 2011 Jan. 5. [Epub ahead of print]
- Niklasson L, Rasmussen P, Oskarsdottir S, et al. Autism, ADHD, mental retardation and behavior problems in 100 individuals with 22q11 deletion syndrome. *Res Dev Disabil* 2009;30:763-73.
- Jacobson C, Shearer J, Habel A, et al. Core neuropsychological characteristics of children and adolescents with 22q11.2 deletion. *J Intellect Disabil Res* 2010;54:701-13.
- Simon TJ. A new account of the neurocognitive foundations of impairments in space, time and number processing in children with chromosome 22q11.2 deletion syndrome. *Dev Disabil Res Rev* 2008;14:52-8.
- Mouaffak F, Kebir O, Chayet M et al. Association of Disrupted in Schizophrenia 1 (Drosoph Inf ServC1) missense variants with ultra-resistant schizophrenia. *Pharmacogenomics J* 2010 June 8. [Epub ahead of print]
- Crepel A, Breckpot J, Fryns JP, et al. Drosoph Inf ServC1 duplication in two brothers with autism and mild mental retardation. *Clin Genet* 2010;77:389-94.
- Hodgkinson CA, Goldman D, Jaeger J, et al. Disrupted in schizophrenia 1 (Drosoph Inf ServC1): association with schizophrenia, schizoaffective disorder, and bipolar disorder. *Am J Hum Genet* 2004;75:862-72.
- Harris SE, Hennah W, Thomson PA, et al. Variation in Drosoph Inf ServC1 is associated with anxiety, depression and emotional stability in elderly women. *Mol Psychiatry* 2010;15:232-4.
- Vassos E, Collier DA, Holden S, et al. Penetrance for copy number variants associated with schizophrenia. *Hum Mol Genet* 2010;19:3477-81.
- Sisodiya SM, Mefford HC. Genetic contribution to common epilepsies. *Curr Opin Neurol* 2011;24:140-5.
- Endris V, Hackmann K, Neuhann TM, et al. Homozygous loss of CHRNA7 on chromosome 15q13.3 causes severe encephalopathy with seizures and hypotonia. *Am J Med Genet A* 2010;152A:2908-11.
- Coon H, Villalobos ME, Robison RJ, et al. Genome-wide linkage using the Social Responsiveness Scale in Utah autism pedigrees. *Mol Autism* 2010;1:8.
- Durand CM, Betancur C, Boeckers TM, et al. Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. *Nat Genet* 2007;39:25-7.
- Gauthier J, Champagne N, Lafreniere RG, et al. De novo mutations in the gene encoding the synaptic scaffolding protein SHANK3 in patients ascertained for schizophrenia. *Proc Natl Acad Sci U S A* 2010;107:7863-8.
- McMullan DJ, Bonin M, Hehir-Kwa JY, et al. Molecular karyotyping of patients with unexplained mental retardation by SNP arrays: a multicenter study. *Hum Mutat* 2009;30:1082-92.
- Stefansson H, Rujescu D, Cichon S, et al. Large recurrent microdeletions associated with schizophrenia. *Nature* 2008;455:232-6.
- Koenen KC, Moffitt TE, Roberts AL, et al. Childhood IQ and adult mental disorders: a test of the cognitive reserve hypothesis. *Am J Psychiatry* 2009;166:50-7.
- Grozeva D, Kirov G, Ivanov D, et al. Rare copy number variants: a point of rarity in genetic risk for bipolar disorder and schizophrenia. *Arch Gen Psychiatry* 2010;67:318-27.
- Fangerau H, Ohlraun S, Granath RO, et al. Computer-assisted phenotype characterization for genetic research in psychiatry. *Hum Hered* 2004;58:122-30.
- Calkins ME, Dobie DJ, Cadenhead KS, et al. The Consortium on the Genetics of Endophenotypes in Schizophrenia: model recruitment, assessment, and endophenotyping methods for a multisite collaboration. *Schizophr Bull* 2007;33:33-48.