

Rare Variants in Complex Traits: Novel Identification Strategies and the Role of de novo Mutations

Loubna Jouan^{a, b} Julie Gauthier^{a, b} Patrick A. Dion^{a, c} Guy A. Rouleau^{a, b, d}

^aCenter of Excellence in Neuroscience and Department of Medicine, Université de Montréal, ^bCentre de Recherche du Centre Hospitalier de l'Université de Montréal (CRCHUM), ^cDepartment of Pathology and Cell Biology, Faculty of Medicine, Université de Montréal, and ^dResearch Center, CHU Sainte-Justine, Montreal, Que., Canada

Key Words

De novo · Autism · Schizophrenia · Rare variants

Abstract

Following the limited success of linkage and association studies aimed at identifying the genetic causes of common neurodevelopmental syndromes like autism and schizophrenia, complex traits such as these have recently been considered under the 'common disease-rare variant' hypothesis. Prior to this hypothesis, the study of candidate genes has enabled the discovery of rare variants in complex disorders, and in turn some of these variants have highlighted the genetic contribution of de novo variants. De novo variants belong to a subcategory of spontaneous rare variants that are largely associated with sporadic diseases, which include some complex psychiatric disorders where the affected individuals do not transmit the genetic defects they carry because of their reduced reproductive fitness. Interestingly, recent studies have demonstrated the rate of germline de novo mutations to be higher in individuals with complex psychiatric disorders by comparison to what is seen in unaffected control individuals; moreover, de novo mutations carried by affected individuals have generally been more deleterious than those observed in control individuals. Advanced

sequencing technologies have recently enabled the undertaking of massive parallel sequencing projects that can cover the entire coding sequences (exome) or genome of several individuals at once. Such advances have thus fostered the emergence of novel genetic hypotheses and ideas to investigate disease-causative genetic variations. The genetic underpinnings of a number of sporadic complex diseases is now becoming partly explained and more major breakthroughs for complex traits genomics should be expected in the near future.

Copyright © 2013 S. Karger AG, Basel

Introduction

The past decade saw a large body of linkage and association studies aimed at identifying the genetics underlying common neurodevelopmental syndromes like autism (ASD) and schizophrenia (SCZ). These studies have met with limited success but altogether they contributed to the development of novel hypotheses and the reconsideration of earlier ones. Complex traits like ASD and SCZ are believed to result from the combined effects of environmental and genetic factors and more recently they have been explored under the 'common disease-rare vari-

ant' hypothesis. This hypothesis supposes that rare highly penetrant mutations predispose to these disorders, by opposition to the 'common disease-common variant' hypothesis which proposes that a unique variant or a combination of highly frequent variants could be responsible for these disorders. The emergent notion that rare variants could cause common complex traits has been concomitant with two major advances: a growing awareness of the recent rapid population expansion and the consequent excess of rare variants as well as a revolution in the genomic tools to find these variants.

In the meantime, the concept of *de novo* mutation which defines a variation detected in the genome of a proband that is not present in the genome of both of their biological parents' has become increasingly evoked across common neurodevelopmental syndromes. *De novo* mutations belong to a subcategory of spontaneous rare variants and regroup single nucleotide variants, small insertions or deletions and even larger copy-number variants (CNVs) and structural rearrangements. Among the observations supporting a role for *de novo* mutations in common diseases like ASD and SCZ are their maintained relatively high incidence, despite a reduction of the reproductive fitness, and the limited success of classical linkage or association studies to explain the genetics behind these diseases. Considering the various genetic hypotheses and their inherent limits and issues, it is a challenge to elaborate the best strategies to discover these diseases-causing variants. Fortunately, new technologies such as massive parallel sequencing are now making it possible to seek and identify all types of variants, from the very common ones to the most private ones at a complete genomic scale. Here, we will address various analysis strategies and some technical issues as well as promising emerging concepts that progressively modify the picture of complex trait genomics.

Genomic DNA Variant Discovery Strategies: From Linkage Analyses to Massive Parallel Sequencing

The concept of a DNA variation (a variant) cannot be considered without the technology associated to its discovery. Following a number of studies based on information from the International HapMap Project that sought to identify common risk alleles in common diseases, a new avenue aimed at identifying more rare but penetrant variants has emerged with the recent approaches able to identify disease-causing variants. Working on case-parent trios at the whole-exome or at the whole-genome

scale is now a promising strategy to address complex diseases because of the technological progresses recently made.

Linkage Analyses and Association Studies: Searching for Common Variants in Common Complex Diseases

Classical family-based linkage studies are based on the fact that, during meiosis, genes that are physically close will remain linked and this consequently allows estimations about the chromosomal locations of disease genes [1]. Linkage studies have been successful in the identification of a number of Mendelian causal genes [2, 3], yet they have inherent limitations like low power and a weak resolution for variants of modest effect [4, 5]. Genome-wide association studies (GWAS) are based on the assumption that the genetic cause of a disease is the result of common DNA variants (present in more than 1–5% of the population) and may involve cumulative effects of a large number of these variants, each of them playing a modest role in the disease. According to this, each risk variant is postulated to confer a small degree of risk, but no variant alone is sufficient to cause the disorder. This gave rise to the International HapMap Project, a consortium among worldwide researchers aimed at building a haplotype map of the human genome. Around 1,000,000 single nucleotide polymorphisms (SNPs) were genotyped in phase I of this project, followed by a phase II that revealed the pattern of association among SNPs in the genomes and their variations across populations. Numerous studies have therefore used the frequencies and patterns of association between these SNPs and the specific phenotypes observed for a given disease [6–9]. The HapMap Project has had an undeniable impact on the studies of common diseases since it has enabled a new generation of genetic association studies based on genome-wide genotyping to examine cohorts of individuals with and without a particular disease. One important advantage of the GWAS approach using the HapMap-based genotyping platform is its independence from prior knowledge about the genes likely to be involved in a disease, hence providing unbiased insights into pathophysiological mechanisms. Using this strategy, variants associated with various complex diseases such as macular degeneration, inflammatory bowel disease, or cardiovascular disease have been discovered and replicated in several studies [10, 11]; this has revealed unanticipated molecular pathways to be involved in these diseases. However, GWAS of complex traits have shown that common variants have modest effects and that very large cohorts are needed for their detection [12]. Numerous

additional factors such as the heterogeneity of studies and probands, incomplete clinical information and the existence of endophenotypes represent serious methodological caveats. Additionally, GWAS approaches have only been able to explain 5–10% of the heritable content of most diseases [13, 14]. The remaining so-called ‘missing heritability’ might therefore be from structural variants or less common variants with larger effect [9, 15]. Indeed, most common variants individually or in combination do confer relatively small increments in risk (1.1–1.5-fold). It has been demonstrated that odds ratios for common versus rare variants largely differ, the second one generally reaching values of 3.5 or higher. Additionally, the fact that GWAS designs are not suited to detect a minor allele frequency (MAF) <5% has led to the hypothesis that we could find a part of this missing heritability with variants with a lower MAF. This could also explain why there is a limited success of GWAS to explain the genetic basis of many complex traits [16–19]. On the other hand, the use of genetic markers to predict health outcomes has still to be further explored and improved at multiple levels by integrating additional information such as the impact of structural variations, e.g. CNVs, gene-to-gene interactions, as well as the role of the environment. More particularly, attempts are made to modify and improve our genetic architecture concepts and the statistical methods used to assess the role of SNPs in association studies. Methods such as whole-genome marker enabled prediction of genetic values have been developed and could help to understand the hidden heritability of common variants [20, 21].

Sanger Sequencing of Candidate Genes: Towards Rare Variant Screening and de novo Variant Discovery

Considering the limitations of GWAS and the fact that complex traits are characterized by a wide genetic heterogeneity, direct sequencing of complete coding regions of candidate genes to identify rare highly penetrant mutations has been explored in recent studies. The hypothesis was that disease-causing mutations are more likely located on specific genes that are biologically relevant for the disease. Additionally, these mutations are likely less tolerated or more susceptible to have an impact on the functionality of the resulting protein. Several strategies have been used to choose candidate genes for Sanger sequencing. For example, Piton et al. [22] focused on 111 X-linked synaptic genes in a well-defined cohort of patients with ASD and SCZ, based on the fact that these two diseases are likely to be synaptic disorders and that genes involved in these diseases should be related to

the formation, regulation or function of the synapse. Candidate genes were chosen according to a scoring system built on several criteria, including synaptic localization, tissue expression pattern data, their effect on cognition in animal models or any involvement in a relevant pathway for disease or learning. Of interest, one de novo frameshift was found in gene *ILIRAPL1* which is involved in calcium-regulated vesicle release and dendrite differentiation in one patient with ASD [22]. In another candidate gene study, the hypothesis was based on the fact that glutamatergic synapses could play a major role in common neurodevelopmental disorders such as intellectual deficiency (ID), ASD and SCZ [23–25]. The screening was done on 197 genes encoding for various synaptic glutamate receptor complexes (enriched in proteins involved in synaptic plasticity, learning and memory) and their known interacting proteins in sporadic cases of non-syndromic ID (NSID). Interestingly, de novo mutations were found in 11% of the patients, the majority of them being predicted to be pathogenic and affecting the protein function. Similarly, rare or de novo point mutations were also evaluated in the neurexin gene family (*NRXN1*, *NRXN2* and *NRXN3*) in 379 subjects affected by ASD, SCZ or NSID and their parents by coding exons and flanking splice junction Sanger sequencing [26]. A de novo truncating mutation in *NRXN1* and a truncating mutation in *NRXN2* were found in a patient with SCZ and ASD, respectively, which underscores the role of rare variants in these diseases as well as the important role of synaptic organizing proteins and the impact of neurexin molecular pathways in the pathogenesis of these neurodevelopmental disorders. Rare mutations in multiple neuroligins (*NLGN1*, *NLGN3* and *NLGN4*) have also been reported in ASD and ID [27–32] as well as variants in *LRRTM1* and *LRRTM4* in SCZ and ASD [33, 34]. Another successful candidate gene analysis identified 2 de novo mutations in the coding region of *SHANK3* following the screening of 185 unrelated SCZ patients [35]. One of them, a nonsense de novo R1117X mutation, was identified in a proband and his 2 affected brothers, revealing a germline mosaicism inferred from the haplotype analysis. Interestingly, *SHANK3* was shown to be gene-dosage sensitive [36], and numerous studies previously identified a linkage in SCZ families at locus 22q11–13 which includes *SHANK3* [37–41]. Hence, candidate gene studies have had a great impact on the discovery of important rare variants in complex diseases and have also confirmed that de novo variants could explain a part of the missing heritability in these complex traits.

Massive Parallel Sequencing – A Genome-Wide Analysis Tool to Detect Disease-Causing Variants at the Very Single-Base Level

For several years now, revolutionary technologies have catalyzed the development of massive parallel sequencing and allowed the opening of new avenues to address the impact of genetic variations on human health [42]. This now offers a reliable way at an unprecedented speed and the resolution to process millions of sequence reads in parallel, hence providing a comprehensive picture of the human genome at a very lower cost. Whole-genome sequencing is the ultimate approach for the detection of genomic variations in a patient's genome, whereas exome sequencing focuses on the protein-coding regions [43]. By quickly detecting sequence variations at the exome or whole-genome scales, it is now possible to rapidly search for the underlying genetic components of a complex disease and to identify disease-causing variations in affected individuals [44, 45]. With these new technologies, we can now address an increasing number of questions and issues, from the identification of a novel entity (bacteria, parasites) to the analysis of DNA obtained from a tumor or from an affected patient with a known or unknown disease, all within a reasonable timeline. One of the strength of next-generation sequencing (NGS) is its ability to identify in an unbiased manner different mutations in many different genes, a strategy that facilitates the study of heterogeneity in a given complex trait. As a proof of concept, Toydemir et al. [46] successfully used exome sequencing to evaluate whether it would be possible to retrieve the causative gene in 4 unrelated individuals with Freeman-Sheldon syndrome, a rare autosomal dominant disorder caused by a known mutation in *MYH3* gene. They successfully identified the expected gene that had various mutations shared by the 4 affected individuals. This was the first demonstration that NGS is a reliable strategy to identify disease-causing variants even in a small number of unrelated patients. Ng et al. [47, 48] later used exome sequencing to identify the unknown causative gene of Miller syndrome, a rare Mendelian disease. By sequencing only 4 patients and 8 controls, they identified multiple coding insertions and deletions (indels), non-synonymous and splice site variants in a single candidate gene, *DHODH*, in the affected cases (none of which were present in the controls). In another study, Choi et al. [49] identified a variant in a highly conserved position in an individual with unclear diagnosis; the homozygous missense (D652N) was in a gene causing congenital chloride-losing diarrhea (*SLC26A3*) and was consistent with the patient's symptoms, hence proving for the first time that NGS is

also a reliable tool for medical diagnosis. NGS can also be used in combination with another analysis strategy such as linkage analysis. Sobreira et al. [50] used a combined analysis strategy, the first one aimed at focusing on a region of interest using linkage analysis and the second one using NGS to sequence the whole genome of 1 affected individual and 8 controls and then analyze all rare variants located in the region of interest; this led to the discovery of a rare deletion and a nonsense variant in gene *PTPN11* in 2 patients with metachondromatosis, a dominant disease with incomplete penetrance [50, 51]. The use of exome or whole-genome sequencing to identify rare variants in complex traits has now proved to be a viable and efficient approach and the list of successful studies keeps growing. Various steps could nonetheless be improved, from the selection of ideal individuals to be sequenced to the choice of variant-filtering criteria, in order to build more refined lists of variants and candidate genes. While recessive disorders are easier to address by selecting rare homozygous or compound heterozygous variants, dominant disorders still remain a challenge [43]. Cirulli and Goldstein [51] suggest that by choosing the most distantly related individuals in a pedigree or by selecting individuals with extreme-trait phenotype, we could improve our results and get the most out of these technologies.

Rare Variants: The Most Recent Advances

Rare Variants Are 'Common'

Across the spectrum of allele frequencies found in a given population, 3 major classes of variants can be observed according to their respective MAF: common variants (MAF >5%), rare variants (MAF <0.5 or 1%) and intermediate allele-frequency variants with a MAF between these 2 values. The strategy to search for rare disease-causing variants is based on the fact that the disease being studied is deleterious in regard to reproductive fitness; so a variant promoting this disease should be selected against and therefore it could not be common. It is believed that low-frequency and rare variants conferring susceptibility to common complex diseases may have larger effect sizes compared to the common variants identified to date. Such rare variants are subjected to natural selection and are therefore either removed or maintained at a low frequency [52–54]. Interestingly, rare genetic variants are predicted to largely outnumber common variants in the human genome [55, 56]. In the context of the resequencing of 408 brain-expressed genes in a cohort of ASD and SCZ patients, it was reported that rare mis-

sense variants showed an increase of predicted detrimental effects and that some specific genes had an excess of rare missense variants, as observed in genes *GRIN2B*, *MAP1A* and *CACNAF1* [53]. Moreover, it has recently been reported that the distribution of MAFs is skewed towards an excess of rare variants: over one-third of all polymorphisms have frequencies <5% [57]. Two different groups have recently published the results of NGS of large populations and specifically characterized the rare variant events [58, 59]. Interestingly, they found that the majority of coding variations are rare, novel and deleterious. Following exome sequencing of 15,585 human protein-coding genes in 2,440 individuals from both European and African ancestry, Tennessen et al. [59] found that among more than 500,000 rare variants, 86% had a MAF <0.5%. Interestingly, among the point mutations detected, 38% were synonymous, whereas 58% were non-synonymous (of them, 97.8% were missenses and 2.11% non-senses), and the remaining 4% were targeting splice sites or non-coding regions. Moreover, based on several prediction softwares, at least 16.9% were predicted to be deleterious. They also demonstrated that the proportion of rare variants per base pair in each gene is higher than that of common variants, and the average ratio of rare to common alleles per base pair is 6:1. Nelson et al. [58] sequenced 202 candidate genes (known or supposed drug targets) in 14,002 individuals, and also demonstrated a very high number of rare variants (MAF <0.5%) compared to common variants, an observation that the authors have linked to the recent exponential growth of the human population. These variants are believed to have appeared very recently in the human evolution history. In fact, Nelson et al. [50] showed that a study based on fewer than 100 individuals will allow the discovery of rare variants with a frequency around 1:100, with a focus on rare variants dated from a long time ago. When the sample size gets larger, the amount of very rare variants (MAF <0.01%) becomes more detectable. These rare variants are also young variants, dated from less than 2,500 years ago. This implies that a rare variant found in a cohort of 100 individuals is likely to be older than a variant with a MAF of 0.005% in a cohort of 10,000 individuals. Common variants with a frequency >5% have likely been segregating for several tens of thousands of years [60]. Consequently, the larger the sample, the more recent the historical period it targets [61]. Therefore, rare variations (with a MAF <1%) which have resulted from the recent expansion of the human population likely contribute to human phenotypic variation and to disease susceptibility as they have been subjected to less selection.

De novo Variants – What, Where, and How Many?

De novo germline mutations have been extensively studied in the context of neurodevelopmental diseases such as psychiatric disorders, where a maintained occurrence is observed despite a decreased fertility in affected cases. In fact, this mutation type is clearly associated with sporadic diseases and is globally found to be more deleterious than inherited variation, based on the fact that it has been subjected to less stringent evolutionary selection [54, 62]. De novo chromosomal abnormalities and de novo CNVs detected by genomic microarrays have been largely studied in ASD, SCZ and ID [63–69]. It has been shown that large de novo CNVs occur in 10% of the patients with ASD and ID [30, 65, 68–71]. We have also discussed above that candidate gene screening using Sanger sequencing allowed for the fine tuning detection of small indels and de novo point mutations.

De novo mutations can occur in the germline, during embryogenesis or somatically as revealed in the case of Proteus syndrome [72]. Of note, detecting somatic de novo mutations is more complex than calling de novo germline mutations, since in this case somatic mutations are not expressed in all tissues; in fact, to be more precisely defined, it would require NGS of a single cell [73]. If both somatic and germline mosaicism exist, only the latter contributes to the transmission of what might appear to be a de novo mutation to the offspring. Interestingly, the rate of germline de novo mutations is higher in individuals with genetic diseases. This implies that there would be an excess of de novo mutations in individuals affected by a complex trait. This point has been introduced by a reported excess of de novo CNVs in ASD and SCZ compared to controls [69, 74, 75]. More recently, Awadalla et al. [76] have demonstrated that the sequencing of 401 genes in families with affected individuals revealed an excess of missense relative to neutral de novo mutations and consequently potentially disease-predisposing de novo mutations in the SCZ and ASD cohort, and that these mutations are therefore located in candidate genes for SCZ and ASD. They identified 15 validated de novo mutations which were either germline-derived mutations or had arisen as somatic mutations in blood tissue; among them, 14 were detected in the ASD and SCZ cohort. This corresponds to 5% of the affected individuals, and the fact that the rate of missense or nonsense to neutral or genome-wide mutations is 5 times higher supports that the mutations found could be causative for ASD or SCZ.

There are several lines of evidence supporting the role of de novo mutations as a mechanism causing common

complex traits such as SCZ and ASD. Mutations in genes leading to severe diseases often cause lethality at embryonic stages or a reduced reproductive fitness. Therefore, such mutations are not transmitted to multiple family members and thus cannot be discovered through the classical technique of linkage gene mapping or association studies. In addition, in diseases such as SCZ and ASD, there is a large difference in the prevalence in monozygotic and dizygotic twins, which could also be explained by a spontaneous novel mutation and supports a de novo model [77, 78]. The incidence rate of these complex traits is maintained through time and generation at a relatively high and constant level (1% worldwide for ASD and SZC). Moreover, there is a significant increased risk of these diseases with increasing paternal age, which is parallel to an age-related increase in paternal de novo mutations [79–81]. Indeed, the male-to-female ratio of the de novo mutation rate is estimated to be 4–6:1 due to a higher number of germ-cell divisions in males; so one would predict that de novo mutations would be more frequent in males and even more in older males. This male bias is thought to be linked to the increased number of cell divisions in the male germline and the consequent increased chance of replication mistakes, whereas other replication mistakes are due to a selective advantage conferred by de novo mutations during the spermatogenesis [82, 83]. The importance of the father's age for the rate of de novo mutations has been reinforced by a very recent study showing that the mutation rate in a given population is highly linked to the father's age in the group. This last one is now an unavoidable parameter when addressing complex diseases such as SCZ or ASD [80]. Several hypotheses could explain the apparition of a de novo mutation, among them the DNA environment (chromosomal location, surrounding DNA sequence types, repetition), the nucleotide composition and any DNA repair mechanism impairment [84, 85].

Recently, the de novo germline mutation rate has been estimated at around 1.18×10^{-8} per position for point mutations, corresponding to 74 novel variants per genome per generation [86]. For small indels, it has been estimated at around 4×10^{-10} per position, resulting in 3 novel indels per genome per generation [87], and CNVs larger than 100 kb are estimated to occur de novo in 1 out of 50 individuals [88]. Comparative genomic methods estimate that in any single conceptus there are 1–3 new deleterious mutations that lead to an altered amino acid per genome, hence corresponding to 1 new potentially damaging mutation in 10,000 genes/zygote [89]. The de novo hypothesis has had a major impact on autosomal NSID.

Given the lower reproductive fitness in this disease and the fact that de novo chromosomal rearrangements represent the most recognized cause of ID has raised the hypothesis that de novo point mutations could be a possible avenue to explore. A candidate gene approach revealed 3 de novo protein-truncating mutations in 3 out of 94 patients with non-syndromic mental retardation in *SYNGAP1*, coding for a GTPase-activating protein [90]. The resulting mutated protein lacked key-interacting domains (RASGAP and QTRV) which are involved in synaptic plasticity and spine morphogenesis important for learning and memory [91, 92]. So it seems that disruption of *SYNGAP1* by truncating de novo mutations is associated with a homogeneous clinical phenotype characterized by moderate-to-severe mental retardation and severe language impairment. Hamdan et al. [93] also reported a significant enrichment of functional versus neutral de novo mutations in a NSID cohort in genes *KIF1A*, *GRIN1*, *EPB41L1* and *CACNG2*. These de novo mutations were located at well-conserved residues, were predicted to be damaging and affected protein function in cell culture [93]. Indeed, since de novo mutations occur spontaneously, they are less subjected to evolutionary selection [54, 62] and therefore may be more deleterious. This is one of the reasons why this type of mutation is believed to cause sporadic genetic diseases by affecting a key protein structure or its important functional domains.

Massive Parallel Sequencing and Screening for de novo Variants – A Winning Strategy to Address Complex Traits

A complex trait results from the confluence of various non-genetic and genetic factors as various as SNPs, CNVs, microRNA, splice variants, posttranslation modifications, DNA modifications, external environment, or chromatin modifications. By influencing gene expression, protein structure and function, a deleterious variant can have a profound impact on interacting networks that together influence susceptibility to a complex phenotype and account for the genetic component of a complex trait's etiology. The neural circuitry is a highly complex platform which offers a large number of deleterious mutational targets that could predispose to neurodevelopmental diseases. As we have previously introduced, rare variants are actively contributing to several complex psychiatric diseases. The availability of whole-genome or exome sequencing now permits the discovery of de novo

mutations, including point mutations and small indels at the genome-wide scale [51, 67, 94]. An increased number of studies aimed at characterizing the mutation rate and the fine-mapping of de novo mutations as well as their association with diseases has therefore progressively emerged [67, 95].

Examples of Successful Studies

The discovery of de novo mutations by exome sequencing was first achieved with the rare syndromic forms of dominant sporadic diseases, i.e. Schinzel-Giedion syndrome and Kabuki syndrome [96, 97], both characterized by ID and typical facial features. Hoischen et al. [96] prioritized heterozygous de novo mutations in a single gene as Schinzel-Giedion syndrome occurs sporadically. Exome sequencing was done on 4 unrelated individuals to identify potentially causing variants. De novo mutations in all 4 patients were found in the *SETBP1* (SET-binding protein 1) gene, and were further confirmed in 8 out of 9 additional individuals by Sanger sequencing screening [96]. In the case of Kabuki syndrome, a heterogeneous disease, mutations in the *MLL2* gene were found to be de novo in one of the cases and transmitted from parents to children in other cases [97]. In Bohring-Opitz syndrome, a sporadic disease characterized by a severe ID, exome sequencing allowed for the identification of a number of de novo mutations resulting in an heterozygous stop codon in one gene, *ASXL1* (additional sex combs-like 1), in the 3 affected individuals that were sequenced, a finding that was further confirmed by Sanger sequencing in 4 additional patients [98]. Visers et al. [67] used a family-based exome sequencing approach to test the de novo mutation hypothesis in ID and sequenced 10 case-parent trios. Nine de novo mutations were detected in 7 different patients; 8 of these mutations were heterozygous on the autosomes and 1 was hemizygous on the X chromosome, all in different genes. Among them, some deleterious de novo mutations were found in different genes such as *RAB39B* or *SYNGAP1*, two important candidate genes involved in brain structure and function. Sanders et al. [99] performed whole-exome sequencing in 238 families from the Simons Simplex Collection (SSC), a major research project that has led to the collection of genetic samples from 2,700 families with one ASD child, unaffected parents and additional siblings. They clearly showed that non-synonymous de novo mutations and more particularly highly damaging ones, such as nonsense and splicing variants, were strongly associated with ASD. An additional study using families from the SSC also revealed that highly disrupting mutations are

statistically more frequent in children with ASD than in unaffected siblings, in contrast to missense or synonymous mutations, and that the paternal line is the main source for these de novo variations, the majority of them taking place before conception [100]. Finally, recent studies that combined exome sequencing data and protein-protein interaction network and pathway analysis data revealed that de novo mutations were targeting genes belonging to common functional pathways [101, 102].

SCZ – Step by Step

High heritability (80%) has been demonstrated in SCZ, and the predominant view was that the genetic basis for this disease involves a combination of multiple common variants which interact with various environmental factors. However, after different genetic approaches, such as segregation analysis, cytogenetics studies, genome-wide linkage, GWAS and CNVs, a few causative variants were identified and replicated [103–106]. The hypothesis that de novo mutations could cause SCZ is attractive, since this model is particularly fitted for such a largely sporadic disease. Indeed, an enrichment of de novo CNVs was found in patients with SCZ [69, 107, 108]. Additionally, some structural variants with a frequency <0.5% and a genotypic relative risk of 5–20 have been demonstrated to play a role in SCZ, even if none of them is fully penetrant and sometimes shared with other phenotypes such as ASD or developmental delay [109, 110]. Rare structural variants have been found in *NRXN1* and *VIPR2*, and some studies have reported rare structural variants and sometimes de novo structural variants in genes belonging to common functional pathways, suggesting potential enrichment for disruption in genes involved in neuronal pathways such as postsynaptic signaling [108, 111–114]. In parallel, de novo CNVs have also emerged as an important genomic cause of SCZ [115], the variant with the strongest association being the 22q11.2 microdeletion, a recurrent and de novo mutation [69, 116]. Based on increasing evidence suggesting that defective synaptic function could account for this type of neurodevelopmental diseases [117–119], several genes were screened using the candidate gene approach. *KIF17*, a dendrite-specific motor protein, was screened in 188 patients with sporadic SCZ and a de novo truncating mutation was reported in 1 SCZ patient [120]. This deleterious mutation was further confirmed in a zebrafish model, where knocking down of its expression resulted in a developmental defect. Another gene, *SHANK3*, was also sequenced in 185 SCZ trios (proband and parents) and 285 controls, since its implication in sporadic SCZ cases was also reinforced by

a model of heterozygous *SHANK3* knock-out mice that displayed deficits in hippocampal long-term potentiation, supporting a role for this gene in synaptic transmission and plasticity [121]. Interestingly, one nonsense and one missense de novo variants were found in this gene in a SCZ cohort [35]. Girard et al. [122] have recently sequenced the exomes of 14 individuals with SCZ and their parents (trios). Interestingly, they found 15 de novo mutations in 8 SCZ individuals. Among the 15 mutations, 4 were nonsense mutations in *LRP1*, *ALS2CL*, *ZNF480* and *KPNA1* and therefore caused a premature stop codon. A major observation in this study was that the ratio of nonsense to missense mutations is 4:15, which is significantly higher than the expected 1:20 ratio, as evaluated in a previous study [122]. This study strongly suggests that there is an enrichment of de novo mutations within coding sequences in patients with SCZ. This highlights the fact that the de novo mutation rate is significantly increased in SCZ and that a portion of sporadic complex diseases is very likely caused by de novo mutations. In a similar study, exome sequencing of 53 family trios of subjects diagnosed with SCZ or schizoaffective disorder as well as 22 unrelated healthy controls recruited from the genetically homogeneous Afrikaner population of European descent in South Africa also supports a de novo mutational paradigm for SCZ [123]. Interestingly, 34 de novo point mutations (32 being non-synonymous missense mutations) and 4 de novo indels were observed. Among the 34 missenses, 19 were predicted to affect the protein function. Globally, 27 out of 53 cases carry at least one de novo mutational event corresponding to 51%, a rate which is comparable to a rate of 51% previously found in 20 parent-child trios with ASD [94]. In the control cohort, 7 out of 22 individuals carried de novo events. A major observation in this study is that the ratio of non-synonymous missense to synonymous de novo changes is

16:1, which is higher than the expected 2.85:1 and the observed 1.33:1 in the control cohort. Hence, de novo variants are 10 times more likely to harbor non-synonymous changes than inherited rare variants in SCZ sporadic cases. Overall, these 2 analyses using exome sequencing demonstrate a significant enrichment of potentially deleterious variants among de novo events in SCZ cohorts.

Conclusion

NGS has already profoundly modified the genetic analysis of complex traits, and greatly enhanced the detection of de novo mutations. Using these rapidly evolving high-throughput sequencing tools to search for de novo mutations at the whole-genome scale was found to explain in part the genetic underpinnings of a number of sporadic complex diseases, which despite a reduced fitness still remain frequent in the global population. It is now more convincing than ever that those rare variants play a central role in neurodevelopmental diseases and this supports an old hypothesis that rare variants could be the primary drivers of complex diseases. More specifically, de novo variants explain a part of the missing heritability for sporadic diseases. Now that sequencing platforms to rapidly screen the entire genome of an individual are widely available and that variant-prioritizing analytical software enables the testing of more and more hypotheses, one of the next challenges will be to find a way to predict the effects of mutations at the pathophysiological level. Building an interactome map of the neuronal system through our growing knowledge of its molecular pathways, its dysfunctions and future discoveries will probably be an important avenue for a better understanding of neurodevelopmental diseases.

References

- 1 Pulst SM: Genetic linkage analysis. *Arch Neurol* 1999;56:667–672.
- 2 Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, Roses AD, Haines JL, Pericak-Vance MA: Gene dose of apolipoprotein e type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* 1993;261:921–923.
- 3 Lifton RP: Genetic dissection of human blood pressure variation: common pathways from rare phenotypes. *Harvey Lect* 2004;100:71–101.
- 4 Altmuller J, Palmer LJ, Fischer G, Scherb H, Wjst M: Genomewide scans of complex human diseases: true linkage is hard to find. *Am J Hum Genet* 2001;69:936–950.
- 5 Risch NJ: Searching for genetic determinants in the new millennium. *Nature* 2000;405:847–856.
- 6 International HapMap Consortium: The International HapMap Project. *Nature* 2003;426:789–796.
- 7 International HapMap Consortium: A haplotype map of the human genome. *Nature* 2005;437:1299–1320.
- 8 International HapMap Consortium, Frazer KA, Ballinger DG, Cox DR, et al: A second generation human haplotype map of over 3.1 million SNPs. *Nature* 2007;449:851–861.
- 9 Manolio TA, Brooks LD, Collins FS: A HapMap harvest of insights into the genetics of common disease. *J Clin Invest* 2008;118:1590–1605.
- 10 Rioux JD, Xavier RJ, Taylor KD, et al: Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nat Genet* 2007;39:596–604.

- 11 Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST, Bracken MB, Ferris FL, Ott J, Barnstable C, Hoh J: Complement factor H polymorphism in age-related macular degeneration. *Science* 2005;308:385–389.
- 12 Kiezun A, Garimella K, Do R, Stitzel NO, Neale BM, McLaren PJ, Gupta N, Sklar P, Sullivan PF, Moran JL, Hultman CM, Lichtenstein P, Magnusson P, Lehner T, Shugart YY, Price AL, de Bakker PI, Purcell SM, Sunyaev SR: Exome sequencing and the genetic basis of complex traits. *Nat Genet* 2012;44:623–630.
- 13 Asimit J, Zeggini E: Rare variant association analysis methods for complex traits. *Annu Rev Genet* 2010;44:293–308.
- 14 Schork NJ, Murray SS, Frazer KA, Topol EJ: Common vs. rare allele hypotheses for complex diseases. *Curr Opin Genet Dev* 2009;19:212–219.
- 15 Scott LJ, Mohlke KL, Bonnycastle LL, et al: A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 2007;316:1341–1345.
- 16 Manolio TA, Collins FS, Cox NJ, et al: Finding the missing heritability of complex diseases. *Nature* 2009;461:747–753.
- 17 Maher B: Personal genomes: the case of the missing heritability. *Nature* 2008;456:18–21.
- 18 Frazer KA, Murray SS, Schork NJ, Topol EJ: Human genetic variation and its contribution to complex traits. *Nat Rev Genet* 2009;10:241–251.
- 19 Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, Nadeau JH: Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet* 2010;11:446–450.
- 20 de los Campos G, Gianola D, Allison DB: Predicting genetic predisposition in humans: the promise of whole-genome markers. *Nat Rev Genet* 2010;11:880–886.
- 21 Meuwissen TH, Hayes BJ, Goddard ME: Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 2001;157:1819–1829.
- 22 Piton A, Gauthier J, Hamdan FF, et al: Systematic resequencing of x-chromosome synaptic genes in autism spectrum disorder and schizophrenia. *Mol Psychiatry* 2011;16:867–880.
- 23 Kaufman L, Ayub M, Vincent JB: The genetic basis of non-syndromic intellectual disability: a review. *J Neurodev Disord* 2010;2:182–209.
- 24 Laumonnier F, Cuthbert PC, Grant SG: The role of neuronal complexes in human x-linked brain diseases. *Am J Hum Genet* 2007;80:205–220.
- 25 Pocklington AJ, Cumiskey M, Armstrong JD, Grant SG: The proteomes of neurotransmitter receptor complexes form modular networks with distributed functionality underlying plasticity and behaviour. *Mol Syst Biol* 2006;2:2006.0023.
- 26 Gauthier J, Siddiqui TJ, Huashan P, et al: Truncating mutations in NRXN2 and NRXN1 in autism spectrum disorders and schizophrenia. *Hum Genet* 2011;130:563–573.
- 27 Glessner JT, Wang K, Cai G, et al: Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature* 2009;459:569–573.
- 28 Jamain S, Quach H, Betancur C, Rastam M, Colineaux C, Gillberg IC, Soderstrom H, Giron B, Leboyer M, Gillberg C, Bourgeron T; Paris Autism Research International Sibpair Study: Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nat Genet* 2003;34:27–29.
- 29 Laumonnier F, Bonnet-Brilhault F, Gomot M, Blanc R, David A, Moizard MP, Raynaud M, Ronce N, Lecomte E, Calvas P, Laudier B, Chelly J, Fryns JP, Ropers HH, Hamel BC, Andres C, Barthelemy C, Moraine C, Briault S: X-linked mental retardation and autism are associated with a mutation in the NLGN4 gene, a member of the neuroligin family. *Am J Hum Genet* 2004;74:552–557.
- 30 Marshall CR, Noor A, Vincent JB, et al: Structural variation of chromosomes in autism spectrum disorder. *Am J Hum Genet* 2008;82:477–488.
- 31 Sudhof TC: Neuroligins and neuroligins link synaptic function to cognitive disease. *Nature* 2008;455:903–911.
- 32 Zhang C, Milunsky JM, Newton S, Ko J, Zhao G, Maher TA, Tager-Flusberg H, Bolliger MF, Carter AS, Boucard AA, Powell CM, Sudhof TC: A neuroligin-4 missense mutation associated with autism impairs neuroligin-4 folding and endoplasmic reticulum export. *J Neurosci* 2009;29:10843–10854.
- 33 Pinto D, Pagnamenta AT, Klei L, et al: Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* 2010;466:368–372.
- 34 Francks C, Maegawa S, Lauren J, et al: LRRMT1 on chromosome 2p12 is a maternally suppressed gene that is associated paternally with handedness and schizophrenia. *Mol Psychiatry* 2007;12:1129–1139, 1057.
- 35 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 USA* 2010;107:7863–7868.
- 36 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;39:25–27.
- 37 Condra JA, Neibergs H, Wei W, Brennan MD: Evidence for two schizophrenia susceptibility genes on chromosome 22q13. *Psychiatr Genet* 2007;17:292–298.
- 38 DeLisi LE, Shaw SH, Crow TJ, Shields G, Smith AB, Larach VW, Wellman N, Loftus J, Nanthakumar B, Razi K, Stewart J, Comazzi M, Vita A, Heffner T, Sherrington R: A genome-wide scan for linkage to chromosomal regions in 382 sibling pairs with schizophrenia or schizoaffective disorder. *Am J Psychiatry* 2002;159:803–812.
- 39 Jorgensen TH, Borglum AD, Mors O, Wang AG, Pinaud M, Flint TJ, Dahl HA, Vang M, Kruse TA, Ewald H: Search for common haplotypes on chromosome 22q in patients with schizophrenia or bipolar disorder from the Faroe Islands. *Am J Med Genet* 2002;114:245–252.
- 40 Mowry BJ, Holmans PA, Pulver AE, et al: Multicenter linkage study of schizophrenia loci on chromosome 22q. *Mol Psychiatry* 2004;9:784–795.
- 41 Vallada H, Curtis D, Sham PC, Murray RM, McGuffin P, Nanko S, Gill M, Owen M, Collier DA: Chromosome 22 markers demonstrate transmission disequilibrium with schizophrenia. *Psychiatr Genet* 1995;5:127–130.
- 42 Metzker ML: Sequencing technologies – the next generation. *Nat Rev Genet* 2010;11:31–46.
- 43 Gilissen C, Hoischen A, Brunner HG, Veltman JA: Unlocking Mendelian disease using exome sequencing. *Genome Biol* 2011;12:228.
- 44 Mardis ER: The impact of next-generation sequencing technology on genetics. *Trends Genet* 2008;24:133–141.
- 45 Mardis ER: A decade's perspective on DNA sequencing technology. *Nature* 2011;470:198–203.
- 46 Toydemir RM, Rutherford A, Whitby FG, Jorde LB, Carey JC, Bamshad MJ: Mutations in embryonic myosin heavy chain (MYH3) cause Freeman-Sheldon syndrome and Sheldon-Hall syndrome. *Nat Genet* 2006;38:561–565.
- 47 Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacherjee A, Eichler EE, Bamshad M, Nickerson DA, Shendure J: Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 2009;461:272–276.
- 48 Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ: Exome sequencing identifies the cause of a Mendelian disorder. *Nat Genet* 2010;42:30–35.
- 49 Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, Nayir A, Bakkaloglu A, Ozen S, Sanjad S, Nelson-Williams C, Farhi A, Mane S, Lifton RP: Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci USA* 2009;106:19096–19101.

- 50 Sobreira NL, Cirulli ET, Avramopoulos D, Wohler E, Oswald GL, Stevens EL, Ge D, Shianna KV, Smith JP, Maia JM, Gumbs CE, Pevsner J, Thomas G, Valle D, Hoover-Fong JE, Goldstein DB: Whole-genome sequencing of a single proband together with linkage analysis identifies a Mendelian disease gene. *PLoS Genet* 2010;6:e1000991.
- 51 Cirulli ET, Goldstein DB: Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet* 2010;11:415–425.
- 52 Pritchard JK, Cox NJ: The allelic architecture of human disease genes: common disease-common variant...or not? *Hum Mol Genet* 2002;11:2417–2423.
- 53 Myers RA, Casals F, Gauthier J, Hamdan FF, Keebler J, Boyko AR, Bustamante CD, Piton AM, Spiegelman D, Henrion E, Zilversmit M, Hussin J, Quinlan J, Yang Y, Lafreniere RG, Griffing AR, Stone EA, Rouleau GA, Awadalla P: A population genetic approach to mapping neurological disorder genes using deep resequencing. *PLoS Genet* 2011;7:e1001318.
- 54 Crow JF: The origins, patterns and implications of human spontaneous mutation. *Nat Rev Genet* 2000;1:40–47.
- 55 1000 Genomes Project Consortium, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA: A map of human genome variation from population-scale sequencing. *Nature* 2010;467:1061–1073.
- 56 Fu YX: Statistical properties of segregating sites. *Theor Popul Biol* 1995;48:172–197.
- 57 Gibson G: Rare and common variants: twenty arguments. *Nat Rev Genet* 2012;13:135–145.
- 58 Nelson MR, Wegmann D, Ehm MG, et al: An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. *Science* 2012;337:100–104.
- 59 Tennessen JA, Bigham AW, O'Connor TD, et al: Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* 2012;337:64–69.
- 60 Coventry A, Bull-Otterson LM, Liu X, et al: Deep resequencing reveals excess rare recent variants consistent with explosive population growth. *Nat Commun* 2010;1:131.
- 61 Keinan A, Clark AG: Recent explosive human population growth has resulted in an excess of rare genetic variants. *Science* 2012;336:740–743.
- 62 Eyre-Walker A, Keightley PD: The distribution of fitness effects of new mutations. *Nat Rev Genet* 2007;8:610–618.
- 63 Cook EH Jr, Scherer SW: Copy-number variations associated with neuropsychiatric conditions. *Nature* 2008;455:919–923.
- 64 Sanders SJ, Ercan-Sencicek AG, Hus V, et al: Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron* 2011;70:863–885.
- 65 Sebat J, Lakshmi B, Malhotra D, et al: Strong association of de novo copy number mutations with autism. *Science* 2007;316:445–449.
- 66 Veltman JA, Brunner HG: De novo mutations in human genetic disease. *Nat Rev Genet* 2012;13:565–575.
- 67 Vissers LE, de Ligt J, Gilissen C, Janssen I, Stehouwer M, de Vries P, van Lier B, Arts P, Wiskamp N, del Rosario M, van Bon BW, Hoischen A, de Vries BB, Brunner HG, Veltman JA: A de novo paradigm for mental retardation. *Nat Genet* 2010;42:1109–1112.
- 68 Vissers LE, de Vries BB, Veltman JA: Genomic microarrays in mental retardation: from copy number variation to gene, from research to diagnosis. *J Med Genet* 2010;47:289–297.
- 69 Xu B, Roos JL, Levy S, van Rensburg EJ, Gogos JA, Karayiorgou M: Strong association of de novo copy number mutations with sporadic schizophrenia. *Nat Genet* 2008;40:880–885.
- 70 Levy D, Ronemus M, Yamrom B, Lee YH, Lexton A, Kendall J, Marks S, Lakshmi B, Pai D, Ye K, Buja A, Krieger A, Yoon S, Troge J, Rodgers L, Iossifov I, Wigler M: Rare de novo and transmitted copy-number variation in autistic spectrum disorders. *Neuron* 2011;70:886–897.
- 71 Koolen DA, Pfundt R, de Leeuw N, Hehir-Kwa JY, Nillesen WM, Neefs I, Scheltinga I, Siermans E, Smeets D, Brunner HG, van Kessel AG, Veltman JA, de Vries BB: Genomic microarrays in mental retardation: a practical workflow for diagnostic applications. *Hum Mutat* 2009;30:283–292.
- 72 Lindhurst MJ, Sapp JC, Teer JK, et al: A mosaic activating mutation in AKT1 associated with the Proteus syndrome. *N Engl J Med* 2011;365:611–619.
- 73 Carlson CA, Kas A, Kirkwood R, Hays LE, Preston BD, Salipante SJ, Horwitz MS: Decoding cell lineage from acquired mutations using arbitrary deep sequencing. *Nat Methods* 2012;9:78–80.
- 74 Piton A, Michaud JL, Peng H, et al: Mutations in the calcium-related gene IL1RAPL1 are associated with autism. *Hum Mol Genet* 2008;17:3965–3974.
- 75 Stefansson H, Rujescu D, Cichon S, et al: Large recurrent microdeletions associated with schizophrenia. *Nature* 2008;455:232–236.
- 76 Awadalla P, Gauthier J, Myers RA, et al: Direct measure of the de novo mutation rate in autism and schizophrenia cohorts. *Am J Hum Genet* 2010;87:316–324.
- 77 Bailey A, Le Couteur A, Gottesman I, Bolton P, Simonoff E, Yuzda E, Rutter M: Autism as a strongly genetic disorder: evidence from a British twin study. *Psychol Med* 1995;25:63–77.
- 78 Cardno AG, Gottesman II: Twin studies of schizophrenia: from bow-and-arrow concordances to star wars Mx and functional genomics. *Am J Med Genet* 2000;97:12–17.
- 79 Malaspina D, Brown A, Goetz D, Alia-Klein N, Harkavy-Friedman J, Harlap S, Fennig S: Schizophrenia risk and paternal age: a potential role for de novo mutations in schizophrenia vulnerability genes. *CNS Spectr* 2002;7:26–29.
- 80 Kong A, Frigge ML, Masson G, et al: Rate of de novo mutations and the importance of father's age to disease risk. *Nature* 2012;488:471–475.
- 81 Croen LA, Najjar DV, Fireman B, Grether JK: Maternal and paternal age and risk of autism spectrum disorders. *Arch Pediatr Adolesc Med* 2007;161:334–340.
- 82 Goriely A, Wilkie AO: Paternal age effect mutations and selfish spermatogonial selection: causes and consequences for human disease. *Am J Hum Genet* 2012;90:175–200.
- 83 Arnheim N, Calabrese P: Understanding what determines the frequency and pattern of human germline mutations. *Nat Rev Genet* 2009;10:478–488.
- 84 Ku CS, Polychronakos C, Tan EK, Naidoo N, Pawitan Y, Roukos DH, Mort M, Cooper DN: A new paradigm emerges from the study of de novo mutations in the context of neurodevelopmental disease. *Mol Psychiatry* 2012;18:141–153.
- 85 Cooper DN, Bacolla A, Ferec C, Vasquez KM, Kehrer-Sawatzki H, Chen JM: On the sequence-directed nature of human gene mutation: the role of genomic architecture and the local DNA sequence environment in mediating gene mutations underlying human inherited disease. *Hum Mutat* 2011;32:1075–1099.
- 86 Conrad DF, Keebler JE, DePristo MA, Lindsay SJ, Zhang Y, Casals F, Idaghdour Y, Hartl CL, Torroja C, Garimella KV, Zilversmit M, Cartwright R, Rouleau GA, Daly M, Stone EA, Hurles ME, Awadalla P, Genomes P: Variation in genome-wide mutation rates within and between human families. *Nat Genet* 2011;43:712–714.
- 87 Lynch M: Rate, molecular spectrum, and consequences of human mutation. *Proc Natl Acad Sci USA* 2010;107:961–968.
- 88 Itsara A, Wu H, Smith JD, Nickerson DA, Romieu I, London SJ, Eichler EE: De novo rates and selection of large copy number variation. *Genome Res* 2010;20:1469–1481.
- 89 Eyre-Walker A, Keightley PD: High genomic deleterious mutation rates in hominids. *Nature* 1999;397:344–347.
- 90 Hamdan FF, Gauthier J, Spiegelman D, et al: Mutations in SYNGAP1 in autosomal non-syndromic mental retardation. *N Engl J Med* 2009;360:599–605.
- 91 Vazquez LE, Chen HJ, Sokolova I, Knuesel I, Kennedy MB: SynGAP regulates spine formation. *J Neurosci* 2004;24:8862–8872.
- 92 Komiyama NH, Watabe AM, Carlisle HJ, Porter K, Charlesworth P, Monti J, Strathdee DJ, O'Carroll CM, Martin SJ, Morris RG, O'Dell TJ, Grant SG: SynGAP regulates ERK/MAPK signaling, synaptic plasticity, and learning in the complex with postsynaptic density 95 and NMDA receptor. *J Neurosci* 2002;22:9721–9732.
- 93 Hamdan FF, Gauthier J, Araki Y, et al: Excess of de novo deleterious mutations in genes associated with glutamatergic systems in non-syndromic intellectual disability. *Am J Hum Genet* 2011;88:306–316.

- 94 O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, Girirajan S, Karakoc E, Mackenzie AP, Ng SB, Baker C, Rieder MJ, Nickerson DA, Bernier R, Fisher SE, Shendure J, Eichler EE: Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nat Genet* 2011;43:585–589.
- 95 Roach JC, Glusman G, Smit AF, Huff CD, Hubley R, Shannon PT, Rowen L, Pant KP, Goodman N, Bamshad M, Shendure J, Drmanac R, Jorde LB, Hood L, Galas DJ: Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* 2010;328:636–639.
- 96 Hoischen A, van Bon BW, Gilissen C, et al: De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. *Nat Genet* 2010;42:483–485.
- 97 Ng SB, Bigam AW, Buckingham KJ, et al: Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet* 2010;42:790–793.
- 98 Hoischen A, van Bon BW, Rodriguez-Santiago B, et al: De novo nonsense mutations in ASXL1 cause Bohring-Opitz syndrome. *Nat Genet* 2011;43:729–731.
- 99 Sanders SJ, Murtha MT, Gupta AR, et al: De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* 2012;485:237–241.
- 100 Iossifov I, Ronemus M, Levy D, et al: De novo gene disruptions in children on the autistic spectrum. *Neuron* 2012;74:285–299.
- 101 Neale BM, Kou Y, Liu L, et al: Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* 2012;485:242–245.
- 102 O'Roak BJ, Vives L, Girirajan S, et al: Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* 2012;485:246–250.
- 103 Kim Y, Zerwas S, Trace SE, Sullivan PF: Schizophrenia genetics: where next? *Schizophr Bull* 2011;37:456–463.
- 104 Moskvina V, Craddock N, Holmans P, Nikolov I, Pahwa JS, Green E, Wellcome Trust Case Control Consortium, Owen MJ, O'Donovan MC: Gene-wide analyses of genome-wide association data sets: evidence for multiple common risk alleles for schizophrenia and bipolar disorder and for overlap in genetic risk. *Mol Psychiatry* 2009;14:252–260.
- 105 O'Donovan MC, Kirov G, Owen MJ: Phenotypic variations on the theme of CNVs. *Nat Genet* 2008;40:1392–1393.
- 106 Stefansson H, Ophoff RA, Steinberg S, et al: Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 2009;460:744–747.
- 107 International Schizophrenia Consortium, Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, Sklar P: Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 2009;460:748–752.
- 108 Walsh T, McClellan JM, McCarthy SE, et al: Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* 2008;320:539–543.
- 109 Itsara A, Cooper GM, Baker C, Girirajan S, Li J, Absher D, Krauss RM, Myers RM, Ridker PM, Chasman DI, Mefford H, Ying P, Nickerson DA, Eichler EE: Population analysis of large copy number variants and hotspots of human genetic disease. *Am J Hum Genet* 2009;84:148–161.
- 110 Sullivan PF, Daly MJ, O'Donovan M: Genetic architectures of psychiatric disorders: the emerging picture and its implications. *Nat Rev Genet* 2012;13:537–551.
- 111 Kirov G, Pocklington AJ, Holmans P, et al: De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Mol Psychiatry* 2012;17:142–153.
- 112 Raychaudhuri S, Korn JM, McCarroll SA, International Schizophrenia Consortium, Altshuler D, Sklar P, Purcell S, Daly MJ: Accurately assessing the risk of schizophrenia conferred by rare copy-number variation affecting genes with brain function. *PLoS Genet* 2010;6:e1001097.
- 113 Rujescu D, Ingason A, Cichon S, et al: Disruption of the neurexin 1 gene is associated with schizophrenia. *Hum Mol Genet* 2009;18:988–996.
- 114 Vacic V, McCarthy S, Malhotra D, et al: Duplications of the neuropeptide receptor gene VIPR2 confer significant risk for schizophrenia. *Nature* 2011;471:499–503.
- 115 Lupski JR: Genomic rearrangements and sporadic disease. *Nat Genet* 2007;39:S43–S47.
- 116 Karayiorgou M, Morris MA, Morrow B, Shprintzen RJ, Goldberg R, Borrow J, Gos A, Nestadt G, Wolynec PS, Lasseter VK, et al: Schizophrenia susceptibility associated with interstitial deletions of chromosome 22q11. *Proc Natl Acad Sci USA* 1995;92:7612–7616.
- 117 Maycox PR, Kelly F, Taylor A, Bates S, Reid J, Logendra R, Barnes MR, Larminie C, Jones N, Lennon M, Davies C, Hagan JJ, Scorer CA, Angelinetta C, Akbar MT, Hirsch S, Mortimer AM, Barnes TR, de Belleruche J: Analysis of gene expression in two large schizophrenia cohorts identifies multiple changes associated with nerve terminal function. *Mol Psychiatry* 2009;14:1083–1094.
- 118 Mei L, Xiong WC: Neuregulin 1 in neural development, synaptic plasticity and schizophrenia. *Nat Rev Neurosci* 2008;9:437–452.
- 119 Stephan KE, Friston KJ, Frith CD: Dysconnection in schizophrenia: from abnormal synaptic plasticity to failures of self-monitoring. *Schizophr Bull* 2009;35:509–527.
- 120 Tarabeux J, Champagne N, Brustein E, et al: De novo truncating mutation in Kinesin 17 associated with schizophrenia. *Biol Psychiatry* 2010;68:649–656.
- 121 Betancur C, Sakurai T, Buxbaum JD: The emerging role of synaptic cell-adhesion pathways in the pathogenesis of autism spectrum disorders. *Trends Neurosci* 2009;32:402–412.
- 122 Girard SL, Gauthier J, Noreau A, Xiong L, Zhou S, Jouan L, Dionne-Laporte A, Spiegelman D, Henrion E, Diallo O, Thibodeau P, Bachand I, Bao JY, Tong AH, Lin CH, Millet B, Jaafari N, Joobar R, Dion PA, Lok S, Krebs MO, Rouleau GA: Increased exonic de novo mutation rate in individuals with schizophrenia. *Nat Genet* 2011;43:860–863.
- 123 Xu B, Roos JL, Dexheimer P, Boone B, Plummer B, Levy S, Gogos JA, Karayiorgou M: Exome sequencing supports a de novo mutational paradigm for schizophrenia. *Nat Genet* 2011;43:864–868.