Genome Biology

RESEARCH HIGHLIGHT

Whole-exome sequencing for finding *de novo* mutations in sporadic mental retardation

Peter N Robinson*[1,2,3]

Abstract

Recent work has used a family-based approach and whole-exome sequencing to identify *de novo* mutations in sporadic cases of mental retardation.

Sarah Ng and colleagues first showed that whole-exome sequencing (WES - sequencing of all the exons in the genome) can be used to identify disease genes in 2009 [1]. Clearly, WES represents a revolutionary technology. In light of the ever increasing number of novel disease genes that have been discovered with WES, it has almost come to seem routine, and in fact the first descriptions of applications of WES to make clinical decisions in the care of patients have begun to appear [2]. The article by Lisenka Vissers and colleagues in the December 2010 issue of *Nature Genetics* [3] demonstrates that the revolution is far from over. By providing convincing evidence that WES can be used to detect *de novo* mutations in patients with sporadic cases of non-syndromic mental retardation, the authors [3] have extended the range of clinical situations in which WES can be used and given suggestive evidence that *de novo* mutations are a common cause of mental retardation.

Analysis strategies for disease-gene discovery with WES data

One of the main challenges in WES is the sheer number of variants that are found in an exome. It is likely that each one of us - healthy or not - carries multiple sequence variants, such as nonsense mutations, that most geneticists would have assessed as disease-causing before whole-exome and whole-genome sequencing showed us how common such changes are in human genomes. The challenge in molecular genetic diagnostics is thus quickly shifting from the identification of sequence variants to the interpretation of the variants: which of the tens to hundreds of rare variants in an exome sequence with apparent pathogenic potential is the disease-causing mutation?

Initial analysis strategies in exome sequencing sought to identify genes with rare, potentially pathogenic variants in all or most sequenced individuals with a given disease [1,4-6]. This approach, which we will call intersection filtering, narrows down the candidate set of genes by removing variants also found in databases of common polymorphisms such as dbSNP and also removing variants that are deemed to be unlikely to cause disease, such as synonymous or intergenic nucleotide substitutions. Several studies used linkage analysis in addition to WES to narrow down the region of the exome in which to search for candidate mutations. For autosomal recessive disorders, it is possible to infer linkage intervals directly from WES data in order to restrict the search space [7].

None of these analysis strategies is applicable to single cases of sporadic disease. The study of Vissers *et al.* [3] was motivated by recent estimations that put the per generation mutation rate at about 1 in 100 million positions in the haploid genome, which corresponds to 0.86 *de novo* amino acid altering mutations per newborn [8,9].

The authors [3] developed an analysis strategy that involves a family-based WES approach in which case-parent trios are sequenced in order to identify potentially pathogenic *de novo* changes in the exome sequences of the affected children. *De novo* copy number variations (CNVs) are a known cause of mental retardation, so array comparative genome hybridization (CGH) investigations were used to rule out such CNVs before WES was performed. The exome sequences of the affected child and the healthy parent of ten trios were captured in solution, and approximately 3 Gb of mappable sequence data was generated for each individual, with about 90% of the exons of the about 18,000 targeted genes being covered at least 10 times. There were nearly 22,000 high-confidence variant calls per individual, which were analyzed using a bioinformatic pipeline to exclude

*Correspondence: peter.robinson@charite.de
[1]Institute for Medical and Human Genetics, Charité-Universitätsmedizin Berlin, Augustenburger Platz 1, 13353 Berlin, Germany
Full list of author information is available at the end of the article

variants such as intergenic or synonymous variants that were unlikely to be disease-causing mutations. The remaining changes were filtered against databases of common variants, which reduced the search space to an average of 143 variants per patient. These variants were then compared with the exome sequences of the parents to remove inherited variants, leaving between two and seven candidate *de novo* mutations per trio.

WES data are still relatively noisy, with both false-positive and false-negative variant calls being relatively common. Therefore, the authors [3] set out to validate the WES results using targeted Sanger sequencing. For instance, in one trio with seven candidate mutations, five of the mutations could not be validated by Sanger sequencing in the index patient (and were thus false-positive WES calls in the index patient), and one of the mutations was found by Sanger sequencing in one of the parents (a false-negative WES call in the parent). In total, Sanger sequencing led to the validation of nine *de novo* mutations in seven of the ten trios (Figure 1).

De novo mutations are a common cause of sporadic cases of mental retardation

The *de novo* mutations were all identified in different genes. A *de novo* nonsense mutation was identified in a 2-year-old boy with mental retardation in *RAB39B*, a gene encoding a small GTPase; this mutation is a known cause of X linked mental retardation (also called MRX72; MIM ID 300271). Another *de novo* nonsense mutation was found in an 8-year-old girl with mental retardation in *SYNGAP1* (encoding synaptic Ras GTPase activating protein 1), a known cause of autosomal dominant mental retardation (also called MRD5; MIM ID 612621). No *de novo* mutation could be validated in the 10-year-old boy of trio 10, but a maternally inherited non-synonymous variant in *JARID1C* (also known as *KDM5C*, or lysine (K)-specific demethylase 5C) was identified. Mutations in this gene are a well known cause of syndromic X-linked mental retardation (MIM ID 300534). In four other individuals, *de novo* mutations were found in four genes that seem to be highly likely candidates based on model organism studies and sequence conservation of the affected amino acid residues, but that had not been previously proven to be related to mental retardation in humans. Strictly speaking, the proof that mutations in these genes cause mental retardation in humans will have to await the identification of additional mutations in more patients.

The authors [3] identified several additional *de novo* variants in their patients that probably have nothing to do with the mental retardation. These changes were in genes with no obvious connection to mental functioning, occurred in less highly conserved residues, and in two cases were found in addition to other *de novo* mutations

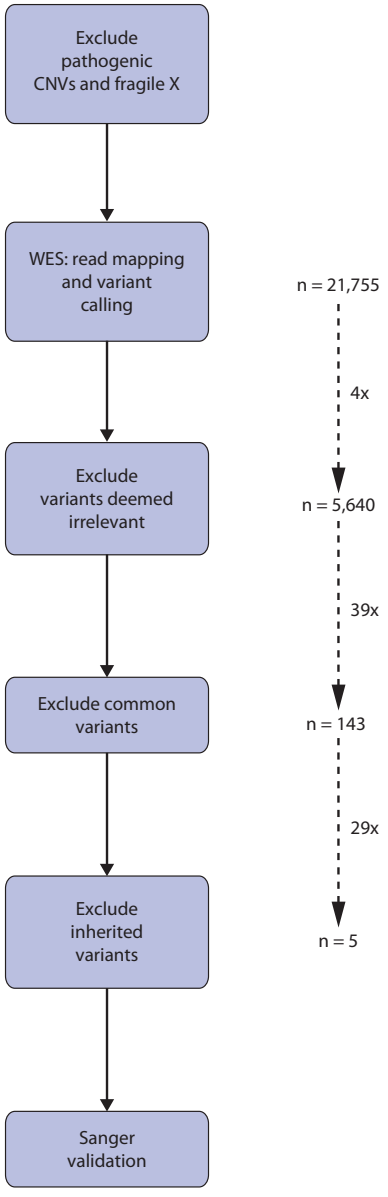

Figure 1. Family-based diagnostic strategy for sporadic cases of mental retardation [3]. After excluding *de novo* pathogenic copy number variations (CNVs) by array CGH analysis and ruling out fragile X syndrome, whole-exome sequencing is performed on the index case and the unaffected parents. A series of bioinformatic filters are used to narrow down the list of candidate mutations. First, mutations deemed to be irrelevant, such as intergenic and synonymous mutations, are removed; common variants are then removed by comparison with genetic variation databases and exome sequencing controls. Finally, inherited variants are removed. The average number of candidate variants after each step in the ten trios investigated by Vissers *et al.* [3] is shown on the right, and the number next to the dotted arrow indicates the fold reduction of candidates achieved by each step. The final analysis step is performed by means of Sanger sequencing to validate the candidates derived by computational analysis of the WES data.

in one of the candidate mental retardation genes. Thus, it seemed unlikely that these mutations are related to the cases of mental retardation. This leads to the obvious conclusion that the mere finding of a *de novo* missense mutation by WES cannot be taken as proof that the change is causally related to the disease being investigated.

It has been notoriously difficult to provide an etiological diagnosis in cases of sporadic mental retardation. Although detection rates vary widely between different reports, in general it has been possible to identify the cause of mental retardation in less than half of patients even after an extensive analysis [10]. Vissers *et al.* [3] identified the conclusive or probable cause of mental retardation in seven of the ten cases they examined. Their results suggest that *de novo* mutations are a major cause of mental retardation and that family-based WES is not only an excellent approach for diagnostics in sporadic mental retardation but also a method of choice for the discovery of novel disease genes in mental retardation. It is tempting to speculate that similar approaches might be fruitful for other classes of sporadic disease with an apparent genetic background. The work of Vissels *et al.* [3] provides yet another demonstration of the enormous potential that the revolutionary method of WES has to improve diagnostic routines in genetics clinics and beyond.

Competing interests
The author declares that he has no competing interests.

Acknowledgements
The author's laboratory is supported by grants from the Deutsche Forschungsgemeinschaft (DFG SFB 760 and RO 2005/4-1), the Bundesministerium für Bildung und Forschung (BMBF, project number 0313911), and the European Commission (FAD; HEALTH-F2-2008-200647).

Author details
[1]Institute for Medical and Human Genetics, Charité-Universitätsmedizin Berlin, Augustenburger Platz 1, 13353 Berlin, Germany. [2]Berlin-Brandenburg Center for Regenerative Therapies, Charité-Universitätsmedizin Berlin, Augustenburger Platz 1, 13353 Berlin, Germany. [3]Max Planck Institute for Molecular Genetics, Ihnestrasse 73, 14195 Berlin, Germany.

Published: 21 December 2010

References

1. Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE, Bamshad M, Nickerson DA, Shendure J: **Targeted capture and massively parallel sequencing of 12 human exomes.** *Nature* 2009, **461:**272-276.
2. Bonnefond A, Durand E, Sand O, De Graeve F, Gallina S, Busiah K, Lobbens S, Simon A, Bellanné-Chantelot C, Létourneau L, Scharfmann R, Delplanque J, Sladek R, Polak M, Vaxillaire M, Froguel P: **Molecular diagnosis of neonatal diabetes mellitus using next-generation sequencing of the whole exome.** *PLoS One* 2010, **5:**e13630.
3. Vissers LELM, de Ligt J, Gilissen C, Janssen I, Steehouwer M, de Vries P, van Lier B, Arts P, Wieskamp N, Rosario MD, van Bon BWM, Hoischen A, de Vries BBA, Brunner HG, Veltman JA: **A *de novo* paradigm for mental retardation.** *Nat Genet* 2010, **42:**1109-1112.
4. Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, Beck AE, Tabor HK, Cooper GM, Mefford HC, Lee C, Turner EH, Smith JD, Rieder MJ, Yoshiura KI, Matsumoto N, Ohta T, Niikawa N, Nickerson DA, Bamshad MJ, Shendure J: **Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome.** *Nat Genet* 2010, **42:**790-793.
5. 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.
6. Hoischen A, van Bon BWM, Gilissen C, Arts P, van Lier B, Steehouwer M, de Vries P, de Reuver R, Wieskamp N, Mortier G, Devriendt K, Amorim MZ, Revencu N, Kidd A, Barbosa M, Turner A, Smith J, Oley C, Henderson A, Hayes IM, Thompson EM, Brunner HG, de Vries BBA, Veltman JA: ***De novo* mutations of *SETBP1* cause Schinzel-Giedion syndrome.** *Nat Genet* 2010, **42:**483-485.
7. Krawitz PM, Schweiger MR, Rödelsperger C, Marcelis C, Kölsch U, Meisel C, Stephani F, Kinoshita T, Murakami Y, Bauer S, Isau M, Fischer A, Dahl A, Kerick M, Hecht J, Köhler S, Jäger M, Grünhagen J, de Condor BJ, Doelken S, Brunner HG, Meinecke P, Passarge E, Thompson MD, Cole DE, Horn D, Roscioli T, Mundlos S, Robinson PN: **Identity-by-descent filtering of exome sequence data identifies *PIGV* mutations in hyperphosphatasia mental retardation syndrome.** *Nat Genet* 2010, **42:**827-829.
8. Lynch M: **Rate, molecular spectrum, and consequences of human mutation.** *Proc Natl Acad Sci U S A* 2010, **107:**961-968.
9. 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.
10. Rauch A, Hoyer J, Guth S, Zweier C, Kraus C, Becker C, Zenker M, Hüffmeier U, Thiel C, Rüschendorf F, Nürnberg P, Reis A, Trautmann U: **Diagnostic yield of various genetic approaches in patients with unexplained developmental delay or mental retardation.** *Am J Med Genet A* 2006, **140:**2063-2074.

doi:10.1186/gb-2010-11-12-144
Cite this article as: Robinson PN: **Whole-exome sequencing for finding *de novo* mutations in sporadic mental retardation.** *Genome Biology* 2010, **11:**144.