Developmental disorders: deciphering exomes on a grand scale

Completion of the first reference human genomes, now nearly 15 years ago, was a mammoth achievement. Expectations were high and predictions of revolutionary effects on science, and medical practice in particular, justifiable. However, we had to wait another 5 years to read individual genomes affordably, and another 5 years before we started to use the information to address the genetics of rare human diseases. The past 5 years have been spectacular, with almost daily novel gene discoveries, not only for rare mendelian diseases but also for complex and multifactorial disorders. With the collected knowledge from the tens of thousands of individuals’ exomes and genomes available, and the thousands now being generated daily worldwide, we have come to realise the vastness of individually rare genetic variation in human genomes. We have learned much about the frequency of de-novo mutations and their relevance to disease. In particular, study of neurodevelopmental diseases such as intellectual disability, autism, epilepsy, and schizophrenia has benefitted, together with that of cancers. Several excellent how-to exome guides, most of which tackle the difficult tasks of sorting pathogenic from non-pathogenic DNA and protein variants, using disease inheritance models or a de-novo mutation hypothesis combined with an appropriate selection of bioinformatics tools and laboratory validation methods, have been proposed. With these approaches, genome-scale sequencing technologies are finally entering medical practice more broadly as unifying tests for diagnosis of genetic disorders.

In The Lancet, Caroline Wright and colleagues report a robust and scalable diagnostic whole exome sequencing workflow, and its practical use when applied to data for 1133 patients collected as part of the Deciphering Developmental Disorders (DDD) study in the UK. The report outlines the processes taken from recruitment, data management, and processing, the choices made to do both automatic and manual variant filtering of around 80,000 variants per individual, and the framework for reporting results. Great care has been taken at every
stage to ensure that accurate diagnoses are achieved, incidental findings minimised, and that there is a clear path for release of data relating to solved and unsolved cases for research purposes, with appropriate patient consent. An overall diagnostic yield of 27% was achieved. This involved manual review of genotype-phenotype correlations, taking advantage of a curated list of more than 1000 known developmental disease genes (Developmental Disorders Genotype-to-Phenotype database, DDG2P). The main focus was on de-novo (72%) and segregating variants. A major, ten-fold, reduction in the number of potentially causal variants needing examination in sporadic cases was achieved by inclusion of parents. This effectively reduced the number of potentially clinically relevant variants to one in most patients. The identification of multiple clinically relevant findings in 17 (2%) of 1133 cases is notable, because often the diagnostic odyssey stops with one major finding. The international community will be watching with interest to see whether and how the proportion of such cases with multiple clinically relevant variants will grow with growing numbers of DDD cases completed, regular re-analyses, and continuing novel disease gene discovery.

For investigators looking to set up their own diagnostic exome service for heterogeneous groups of developmental disorders, take heed—a diagnostic yield of 27% and the need for manual review is what you might conservatively expect because it is mirrored in the independent work of Lee and colleagues, achieving 26% (95% CI 23–29) with a similarly sized cohort (814) and using a similar approach.

The decision not to search actively for incidental findings, even if in clinically relevant genes as recommended by the American College of Medical Genetics and Genomics, although at this stage justified by the DDD team and explored further in their associated ethics study, will no doubt be a point of debate. The authors argue that although the analysis framework could be adapted to search for incidental findings in selected genes, the additional resources and time needed would be significant. Recent studies indicate that 1–2% of individuals sequenced will have an actionable incidental finding. The extra resources, expertise, and cost needed to detect and deal effectively with all medically actionable findings from diagnostic exomes, incidental or not, appears too high now, but isn’t it what we should ultimately aim for?

The 1133 patients represent the tip of the iceberg for the DDD project, which looks to eventually enrol 12 000 patients by April, 2015. A simple extrapolation predicts that at least 3240 families will be diagnosed from this study alone. It will not end there, however, because the flow from controlled public release of these data using the existing DECIPHER infrastructure will mean that variants of hitherto unknown function that are currently sitting in databases in many research and clinical diagnostic laboratories might suddenly have a disease to call home. For forward-thinking molecular and clinical geneticists looking for guidance on how to incorporate genomic technologies into standard practice, and for research laboratories looking for that elusive second family, this work and that to follow from the DDD study is a must read.

*Jozef Gecz, Mark Corbett
School of Paediatrics and Reproductive Health and Robinson Research Institute, The University of Adelaide at the Women’s and Children’s Hospital, North Adelaide, Adelaide SA 5006, Australia
jozef.gecz@adelaide.edu.au
We declare no competing interests.

Copyright © Gecz et al. Open Access article distributed under the terms of CC BY.