Another locus, a new method

The long slow trek of scientific progress is occasionally punctuated by great leaps and quick sprints, towards an obscure finish line. Much of this forward progress has been made possible by advances in technology or experimental methods. Such developments have allowed empirical testing of an extant, previously non-testable hypothesis, or have drastically enhanced investigators’ ability to address questions by speeding the research process. This is exquisitely apparent when viewing the history of human molecular genetics. Through the development of methods that allowed typing of variable markers across the genome, disease-causing mutations were localized; the invention and improvement of DNA sequencing methods allowed individual mutations to be found; and more recently highly parallel genotyping methods have afforded the ability to perform genome-wide association studies. These and other advances have greatly enhanced the ability of scientists to find genetic variants that impart risk for—and cause—disease.

The field of genetics is at the beginning of another period of rapid discovery, this time facilitated by the adoption of a novel method called exome sequencing. A handful of recent manuscripts (Choi et al., 2009; Hedges et al., 2009; Bilguvar et al., 2010; Ng et al., 2010a, b; Shoubridge et al., 2010; Walsh et al., 2010), including one in this issue of Brain (Wang et al., 2010), serve as declarative papers for the practical application of exome sequencing to find disease-causing mutations. Although there are several different exome-sequencing approaches, these all centre on a common methodological principle. First a genomic DNA sample is processed to enrich the sample for all the regions of the genome that are protein coding (the exons); second, this exon-enriched sample is analysed using massively parallel DNA sequencing that produces billions of base pairs of sequence data each day. In this way, researchers can screen the protein-coding portion of an individual genome for rare, disease-associated mutations over a short period of time.

In the current paper, Wang and colleagues use exome sequencing to link TGM6 mutations to a familial form of spinocerebellar ataxia. The authors performed exome sequencing on four patients from a large multigenerational pedigree. Producing just over 4 billion base pairs of sequence data for each, they were able to identify approximately 6000 genetic variants in each sample that resulted in a non-synonymous protein-coding change, a splice site change or an insertion/deletion of nucleotides within the protein-coding region. Because the authors hypothesized that the responsible mutation was likely to be very rare, they removed all the previously reported genetic variants present in public databases, reducing this list down to ~300 novel variants in each individual. Lastly, they looked to see which of these variants were shared amongst each of the four affected family members that had been exome sequenced. This process revealed a single variant, L517W, within TGM6, that was extremely rare and segregated with disease in the family.

While the power of this type of approach lies in its comprehensive nature, there are some methodological limitations. In this instance, the capture array included ~90% of known genes as targets for enrichment but this concentration of targeted exons is not completely efficient; thus not every base pair of the protein-coding regions of each gene was captured and sequenced. Second, this approach does not inform on each of the possible types of mutations; larger deletions, duplications and rearrangements, in addition to certain types of repeat mutations, may easily be missed. Because of these limitations, it is possible that other potential disease-causing mutations were over-looked. Therefore, the investigators used two additional approaches to increase their confidence that TGM6 mutations are indeed the cause of disease. First, they performed linkage to show that the region of the genome containing TGM6 is that part of the genome statistically most likely to contain disease-causing mutations—which it was; and second, they screened other families for TGM6 mutations. As a result of this work, the authors identified a second TGM6 mutation, D327G, linked to ataxia in a Chinese spinocerebellar ataxia family. Collectively, these data support the idea that TGM6 mutations are indeed a novel cause of ataxia. However, because of the limitations noted above, one cannot be completely certain that TGM6 mutations are pathogenic. There could still exist a different genetic cause of disease at this locus (and it is notable that PRNP is within the risk locus). The ultimate proof of pathogenicity will lie in the independent identification of additional families with disease linked to TGM6 mutations and, in publishing this article, the authors have ensured that this mutation will now be tested in families across the world.

Even with these caveats in mind, the approach of exome sequencing is still remarkable on several fronts: first, it promises to condense the time required for the identification of novel mutations from years to weeks; second, it allows geneticists to use...
families that would have previously been considered underpowered for gene identification; and third, using this as a screening tool promises to afford extremely rapid mutation detection and consequently broaden our understanding of how and when mutations manifest.

So now we view the past from a different perspective, one where 20 years of molecular genetics work aimed at finding disease-causing mutations, can now be achieved in exponentially less time and at a much lower cost. It is to be expected that the next year will see a landslide of exome-sequencing articles, much in the same way that there was a deluge of genome-wide association articles from 2006 onwards. These efforts will lead to the identification of novel mutations, a greater understanding of the mutation load each of us carries, and a better appreciation of the varied presentation of genetic diseases. Like almost all technological advances, we can expect this to be quite transient; soon the cost of DNA sequencing will be so low that it will make more sense to perform whole-genome sequencing, rather than targeted exome sequencing. This too, will be another opportunity for a quick sprint towards the finish line.

**Funding**

This work was supported by the Intramural Research Program of the National Institutes on Aging, National Institutes of Health, Department of Health and Human Services. Project number Z01 AG000957-08.

**Andrew B. Singleton and J. Raphael Gibbs**  
NIH - Molecular Genetics Section, Bethesda, MD 20892, USA

Correspondence to: Andrew B. Singleton,  
NIH - Molecular Genetics Section,

**References**


