HotSpots

Exome sequencing: expanding the genetic testing toolbox

CE Kobelka

The Kaiser Permanente Medical Group, San Francisco Genetics Department, San Francisco, CA 94115, USA.
e-mail: Christine.x.kobelka@kp.org

Exome sequencing identifies the cause of a Mendelian disorder
Ng et al. (2009)
Nature Genetics 41(1):30–36

Historically, genetic disorders were diagnosed on a clinical basis, with unifying physical or biochemical findings. Because of continuing advances in genetic testing, an increasing number of clinical diagnoses can now be confirmed with a genetic test result. Although there are many tools in the modern-day genetic testing toolbox, cost and limitations in current ability to interpret results completely place limits on the utility of some of these tests.

One promising tool in the discovery of the genetic underpinnings of rare Mendelian disorders is sequencing of the exome, which is made up of only the protein-coding sequences. In contrast to whole-genome sequencing, which is currently largely cost prohibitive, exome sequencing requires only approximately 5% as much sequencing, and is therefore far less expensive. Exome sequencing is likely to be a reasonable and effective tool to determine disease-causing variants underlying monogenic disorders for several reasons. Positional cloning of protein-coding sequences has proven to be successful in identifying causative variants, and the majority of variants underlying Mendelian disorders are found to be in protein-coding sequences or splice donor and acceptor sites (SS). Additionally, a large percentage of rare non-synonymous (NS) variants are predicted to be deleterious, as opposed to variants in non-coding sequences that usually have neutral or weak phenotypic effects.

In the January 2010 issue of Nature Genetics, Ng et al. described their use of exome sequencing to identify the genetic cause of the rare Mendelian disorder, Miller syndrome. Of the approximately 30 cases of Miller syndrome reported worldwide, only three pairs of affected siblings have been identified in unrelated, non-consanguineous families. As such, Miller syndrome has been suspected to be de novo autosomal dominant, with rare instances of affected siblings resulting from parental germline mosaicism.

To search for the genetic cause of Miller syndrome, the investigators sequenced the exons, and SS, of four individuals with Miller syndrome. Two of these individuals were siblings (kindred 1) and the other two were unrelated (kindreds 2 and 3). Anticipating synonymous variants as far less likely to be pathogenic, investigators focused only on NS, SS, and short coding insertions or deletions (I). A new variant was defined as one that did not exist in the databases compared, namely dbSNP129, exome data from the eight HapMap individuals, and both of these groups combined.

Using models of dominant and recessive inheritance, the investigators attempted to narrow down the candidate genes for Miller syndrome (Fig. 1). Using the siblings, by identifying at least one NS/SS/I variant in the same gene for the dominant model and at least two new NS/SS/I variants for the recessive model, the pool of candidate genes was reduced. To further narrow the search for the causative gene, candidate genes from both siblings were compared with those in the two unrelated individuals with Miller syndrome. Using this approach, the recessive model revealed a single candidate gene, DHODH.

Investigators then combined variant filtering with predictions of whether the NS/SS/I variants were damaging. The recessive model was favored over the dominant model because each affected person was a compound heterozygote for new DHODH mutations, and five of the six mutations were predicted to be damaging. To confirm that the mutations in DHODH were responsible for Miller syndrome, Sanger sequencing of DHODH was used for three additional (unrelated) simplex cases of Miller syndrome, and a previously untested, affected sibling in kindred 2. All four people were found to be heterozygous for missense mutations predicted to be deleterious in DHODH. None of the tested, unaffected siblings were compound heterozygous carriers, and none of the mutations appeared to be de novo, as all of the parents of affected individuals who were tested
1) Sequencing of exons and splice donor and acceptor sites

2) Comparison of case variants to variant databases

3) Filtering of candidate genes in kindred 1 against other kindred 2 and 3

**Fig. 1.** Filtering exome sequencing variation to narrow down candidate genes.

were heterozygous carriers. Additionally, genotyping failed to identify any of these mutations in any of 200 chromosomes from geographically matched, unaffected individuals.

The investigators clearly demonstrate the effectiveness of exome sequencing to identify the candidate gene underlying a rare monogenic disorder. However, they also identify several limitations to this method. First, genetic heterogeneity may underlie a clinical diagnosis, which could decrease the power of an exome scan. Second, if a sequence has more than one location in the genome (such as duplications), and if duplicated sequences are not accounted for in the reference sequence, variants in all locations would erroneously seem to be contained within the same gene. Third, potential variation can be missed if coverage is inadequate. Fourth, it would be ideal to rely on a more consistently ascertained catalog of common variants rather than the uneven ascertainment in publicly available databases such as dbSNP. Along these lines, it is important to consider also the geographic ancestry of both the affected individuals and the control databases, as individuals from varied geographic backgrounds can have a higher number of novel and/or rare variants.

Given the rarity of many genetic syndromes, phenotypic overlap, and the possibility for genetic heterogeneity, the use of exome sequencing is logistically difficult, although less so than whole-genome sequencing. It requires collections of individuals with the same rare phenotype and underlying (monogenic) cause. In the clinical realm, exome sequencing on any one individual suspected to have a monogenic disorder is unlikely to identify clearly the causative mutation(s) without further clarification. One possible solution to this is the use of databases of newly described exonic sequencing variants, with cross-reference to phenotypic data and/or disorder classifications. This would be similar databases that have been created for copy-number variants identified by array comparative genomic hybridization (aCGH).

While many clinically diagnosed genetic disorders with yet-to-be-determined genetic causes may appear to be transmitted in a Mendelian fashion, they may not be monogenic. Other causes for genetic disorders include, but are not limited to; cytogenetic abnormalities, genomic copy-number variants detectable by aCGH, large rearrangements within a gene, non-coding sequence variants, and inherited epigenetic modifications (Fig. 2). Since it is often not possible to identify which disorders are monogenic simply by phenotype, a battery of genetic testing is typically necessary to identify the underlying cause for genetic disorders. Importantly, identification of the genetic basis allows for subsequent availability of carrier testing, prenatal diagnosis, and pre-implantation genetic diagnosis. Additionally, understanding the genetic basis may eventually result in therapeutic interventions for affected individuals.
**HotSpots**

A) Potentially deleterious genetic variation identified by exome sequencing utilized by Ng *et al* (2009)

![Exon Diagram](image1)

B) Potentially deleterious genetic variation not identified by exome sequencing

![Genetic Variation Diagram](image2)

*Fig. 2.* Genetic variation not identified by exome sequencing.

There are currently many genetic testing tools to aid in the discovery of the underlying causes of genetic disorders. Exome sequencing will probably prove to be both an efficient and a cost-effective method to identify the genetic basis for many rare, monogenic disorders, and an important addition to the genetic testing toolbox.