MEDICO-LEGAL AND ETHICAL ISSUES IN WHOLE-EXOME SEQENCING

Wayne W. Grody, M.D., Ph.D.
UCLA School of Medicine
UCLA Medical Center

The recent advent of massively parallel or "next-generation" DNA sequencing has brought whole-genome analysis into the clinic for the first time, and many current applications are directed at children and adults with developmental or neurologic disorders that are undiagnosable by standard or single-gene tests. Thus, it is important that neurologists become familiar with this technology, what it can and cannot offer, and its technical and ethical challenges.

For most of the 25+-year history of clinical molecular diagnostics, and for the entirety of the Human Genome Project, DNA sequencing was performed on semi-automated capillary electrophoresis instruments using the biochemical method known as dideoxy-chain-termination or Sanger sequencing, after its inventor. This is a highly accurate though relatively slow method, capable of providing the DNA sequence of a targeted and limited region of about 150-200 nucleotides in a single run of 1-2 days. That length of DNA represents only a small fraction of the total length of most genes, let alone of the entire genome, which explains why the sequencing of the first human genome, under the auspices of the Human Genome Project, took 13 years and about 3 billion dollars. Obviously an effort of that magnitude could never remotely be translated into a clinical test.

One reason for the slow pace of Sanger sequencing is that it is based on daughter-strand synthesis of a small stretch of DNA that is selected by the hybridization of specific oligonucleotide primers to just that region, which serve as start-sites for DNA polymerase to make complementary strands that terminate whenever any one of the four dideoxy-nucleotide derivatives is incorporated, with the resulting sequence deduced by measuring the sizes of the terminated fragments on the capillary electrophoresis instrument. In contrast, the newer method of "next-generation" or "massively parallel" sequencing breaks up the whole genome into more than 300 million small fragments which are universally primed for DNA polymerase copying using four-color fluorescently labeled nucleotides and analyzed on instruments that take instantaneous "snap-shots" of the added nucleotides on each round of synthesis across the entire genome. Each fragment is copied, or "covered", between 10 and 100 or more times, and the aggregated fluorescent photos of all of the synthesis products comprise "cluster arrays" representing 4-5 terabytes of computer data. The alignment software then reconstructs these hundreds of millions of DNA fragments back into the entire genome. Depending on the particular instrument used, the method can provide the sequence of 3-10 gigabases of DNA in a single run,[5] easily enough to cover the entire haploid or diploid genome of an individual. While the instruments and reagents are very expensive, the net cost per sample per run has been decreasing over the last few years, and now stands at about $2500-$6000 per individual, depending upon depth of "coverage" and other factors. This is well within the range of currently accepted single-gene sequencing tests, such as for *BRCA1* and *BRCA2*. Thus, it is now a legitimate test to consider for certain clinical situations.

Most laboratories using next-generation sequencing at present are applying it to just the small subset (about 1.5%) of the total human genome that codes for proteins, comprising about 230,000 exons and designated, for short, the "exome". The procedure involves an additional step beyond those described in the previous paragraph: an exon-capture step by which the coding regions are selected out of the total genome DNA by means of hybridization, either to a microarray or in solution. Since it

believed that the majority of inherited disorders are due to mutations in the coding regions, this approach allows the laboratory to focus exclusively on those and eliminate the tremendous mass of noncoding DNA in the genome which would otherwise add greatly to the sequencing load and produce 100 times more data to interpret. The downside, however, is that the exon-capture techniques are not 100% efficient, and about 3-5% of the exons will not be captured and sequenced. So in that sense the commonly used term "whole-exome sequencing" is actually a misnomer or an exaggeration. If some of the missing exons are deemed essential for ruling out a particular suspected diagnosis, they must be specifically targeted and sequenced by any of a variety of PCR-based work-around techniques.

Depending on the diagnostic question to be answered, some laboratories will further restrict the exons being queried to just a panel of genes known to be associated with the disorder suspected, such as seizures, ataxia, or cardiomyopathy. A key advantage to this approach is that it eliminates the possibility of unwanted, off-target results (see below). But a disadvantage is that most of the existing gene panels produce a yield of positive mutation results in less than half of the patients tested with the associated phenotype.

Whether done by whole-exome or whole-genome approaches, the application of next-generation sequencing to patients with syndromic-appearing but undiagnosed disorders has already produced some dramatic successes. Some of these are in the form of new gene discovery, while others involve detection of mutations in known genes that were not suspected to be associated with the particular patient's phenotype or were not practical to query by traditional genetic testing methods. In just the past 5 years, new causative genes for at least 250 heritable disorders have been discovered by these methods. Often this one test will put an end to the long and expensive "diagnostic odyssey" that these patients go through. For that reason we hope it will be widely accepted for coverage by insurance carriers, especially when they understand that the cost, while appreciable, is less than the aggregate cost of sequencing two or three individual genes one at a time.

Probably the biggest challenge in performing whole-exome and whole-genome sequencing on a clinical basis is dealing with the huge quantity of unexpected sequence variants that are inevitably detected in every case. In just the exome alone, for every real, causative mutation detected, one also finds at least 18,000 variants in other genes, most of which are of uncertain clinical significance; in the whole genome, the number is more like 3 million. In aggregate, these variants have been dubbed the "incidentalome". Obviously, no one can go through all of them manually to decide what they mean and which ones should be reported. Instead, we have to rely on computer software to do the initial "filtering": the discarding of those variants judged by a number of rules to be likely nonpathogenic. Even after that, however, it is not uncommon to be left with several hundred variants that are potentially worrisome, and it may take the combined efforts of molecular biologists, clinical geneticists and bioinformaticists to sort through them. Even more concerning, from an ethical perspective, are the incidental discovery of "off-target" mutations -- deleterious changes in known genes that are not relevant to the patient's current phenotype and the reason for ordering the test, but may predict future disease: for example, the finding of a pathologic *BRCA* mutation in a 3-year-old girl undergoing whole-exome sequencing for autism or hearing loss or congenital seizures. The *BRCA* mutation has nothing to do with those conditions, and we would never deliberately do such predictive testing for an adult-onset disease in a young child – but what is the liability of failing to disclose this incidentally detected risk, either for the child's future or the present health of the parent who transmitted it?

The types of incidental findings in exome sequencing may be classified as follows:

- Missense variants of uncertain significance in known gene (VUS)
- Variants and deleterious mutations in unknown gene(s) (GUS)
- Variants and mutations in known but currently irrelevant genes (e.g., pharmacogenetic markers, recessive carrier states)
- Deleterious mutations in unintended target (e.g., *BRCA* mutations in a baby)

In 2013, the American College of Medical Genetics and Genomics issued guidelines for the types of incidental findings that should be reported and under what circumstances. The major recommendations were:

- Mutations in a select list of high-penetrance, potentially lethal but actionable conditions ***must be sought and reported***
- Same rules apply to sequencing of healthy parents in a "trio" or benign companion tissue when doing tumor sequencing
- These results are given to the ordering clinician who has responsibility for deciding which, when and how to convey to the patient
- The same rules apply ***whether the patient is an adult or a child***
- The patient ***cannot opt out*** from receiving these incidental findings

The conditions listed are in the following groups:

- Familial cancers (BRCA, Lynch, etc.)
- Cardiomyopathies, hereditary arrhythmias (LQT, etc.)
- Connective tissue disorders (EDS, Marfan)
- Neuroectodermal disorders (tuberous sclerosis, NF2)
- Miscellaneous (familial hypercholesterolemia, malignant hypertension)

Release of the guidelines provoked immediate controversy, surrounding perceived infringement of patient autonomy and conditions either included or not included on the list. Since then the ACMG has issued a modification allowing for an informed "opt-out" of receiving incidental findings.

At our institution we do allow for an "opt-out" during the pre-test genetic counseling and informed consent. Moreover, all of our test interpretations are performed by an interdisciplinary Genomic Data Board, consisting of laboratory directors, bioinformaticists, genetic counselors, clinical geneticists, and, if available, the ordering neurologist or other clinician. Our experience with this process over the past 5 years has revealed the following:

- Communication with the ordering clinician is essential
- Diagnostic yield is better than expected
- Trio cases (including mother and father) are much easier and more fruitful than singletons
- Generic autism and developmental delay don't yield much
- Test volume keeps growing
- Insurers are more open to coverage than anticipated

As genome-level sequencing continues to become less expensive and more widespread, it is likely that in time, many of the ethical concerns we now agonize over will become more routine, acceptable, and less alarming to physicians, patients, and insurers. We should be able to build on our ample past experience with predictive testing for emotionally charged neurologic conditions such as Huntington

disease. That too initially evoked much ethical handwringing, but the predicted adverse effects by and large never materialized. We should be able to apply the same reasoned yet pragmatic approach to genome-wide sequencing for diagnosis, carrier screening, and risk-analysis.

References

Fogel, B.L., H. Lee, J.L. Deignan, S.P. Strom, S. Kantarci, X. Wang, F. Quintero-Rivera, E. Vilain, W.W. Grody, S. Perlman, D.H. Geschwind and S.F. Nelson. Exome sequencing improves clinical diagnosis of spinocerebellar ataxia in adults and children with sporadic or familial disease. *JAMA Neurol* 2014;71:1237-1246.

Green, R.C., S. Berg, W.W. Grody, S.S. Kalia, B.R. Korf, C.L. Martin, A.L. McGuire, R.L. Nussbaum, J.M. O'Daniel, K.E. Ormond, H.L. Rehm, M.S. Watson, M.S. Williams and L.G. Biesecker. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet Med* 2013;15:565-574.

Kohane IS, Masys DR, Altman RB. The incidentalome: a threat to genomic medicine. *JAMA* 2006;296:212-215.

Lee, H., J.L. Deignan, N. Dorrani, S. Strom, S. Kantarci, F. Quintero-Rivera, K. Das, T. Toy, A. Eskin, B. Harry, M. Yourshaw, M. Fox, B.L. Fogel, J.A. Martinez-Agosto, D.A. Wong, V.Y. Chang, P. Shieh, C.G.S. Palmer, K. Dipple, W.W. Grody, E. Vilain and S.F. Nelson. Initial two-year experience utilizing clinical exome sequencing for the genetic diagnosis of rare Mendelian disorders at an academic medical center. *JAMA* 2014;312:1880-1887.

tenBosch J, Grody WW. Keeping up with the next generation: Massively parallel sequencing in clinical diagnostics. *J Molec Diagn* 2008;10:484-492.