

# Whole-Exome/Genome Sequencing and Genomics

**AUTHORS:** Wayne W. Grody, MD, PhD, FCAP, FACMG,<sup>a</sup> Barry H. Thompson, MD, MS, FAAP, FACMG,<sup>b</sup> and Louanne Hudgins, MD, FAAP, FACMG<sup>c</sup>

<sup>a</sup>*Divisions of Medical Genetics and Molecular Pathology, Departments of Pathology and Laboratory Medicine, Pediatrics, and Human Genetics, UCLA School of Medicine, Los Angeles, California;* <sup>b</sup>*American College of Medical Genomics and Genetics, Bethesda, Maryland;* and <sup>c</sup>*Division of Medical Genetics, Department of Pediatrics, Stanford University School of Medicine/Lucile Packard Children's Hospital, Stanford, California*

## KEY WORDS

cytogenetics, genetic testing, genomics, next-generation sequencing, primary care, whole-exome sequencing, whole-genome sequencing

## ABBREVIATION

DTC—direct-to-consumer

[www.pediatrics.org/cgi/doi/10.1542/peds.2013-1032E](http://www.pediatrics.org/cgi/doi/10.1542/peds.2013-1032E)

doi:10.1542/peds.2013-1032E

Accepted for publication Aug 28, 2013

Address correspondence to Louanne Hudgins, MD, FAAP, FACMG, Division of Medical Genetics, Department of Pediatrics, Stanford University School of Medicine/Lucile Packard Children's Hospital, 300 Pasteur Dr, H315, Stanford, CA 94305-5208. E-mail: [lhudgins@stanford.edu](mailto:lhudgins@stanford.edu)

PEDIATRICS (ISSN Numbers: Print, 0031-4005; Online, 1098-4275).

Copyright © 2013 by the American Academy of Pediatrics

**FINANCIAL DISCLOSURE:** The authors have indicated they have no financial relationships relevant to this article to disclose.

**FUNDING:** This effort was supported by grant UC7MC21713 from the Health Resources and Services Administration's Maternal and Child Health Bureau. The Genetics in Primary Care Institute is a cooperative agreement between the American Academy of Pediatrics and the Maternal and Child Health Bureau.

**POTENTIAL CONFLICT OF INTEREST:** The authors have indicated they have no potential conflicts of interest to disclose.

## abstract

As medical genetics has progressed from a descriptive entity to one focused on the functional relationship between genes and clinical disorders, emphasis has been placed on genomics. Genomics, a subelement of genetics, is the study of the genome, the sum total of all the genes of an organism. The human genome, which is contained in the 23 pairs of nuclear chromosomes and in the mitochondrial DNA of each cell, comprises >6 billion nucleotides of genetic code. There are some 23 000 protein-coding genes, a surprisingly small fraction of the total genetic material, with the remainder composed of noncoding DNA, regulatory sequences, and introns. The Human Genome Project, launched in 1990, produced a draft of the genome in 2001 and then a finished sequence in 2003, on the 50th anniversary of the initial publication of Watson and Crick's paper on the double-helical structure of DNA. Since then, this mass of genetic information has been translated at an ever-increasing pace into useable knowledge applicable to clinical medicine. The recent advent of massively parallel DNA sequencing (also known as shotgun, high-throughput, and next-generation sequencing) has brought whole-genome analysis into the clinic for the first time, and most of the current applications are directed at children with congenital conditions that are undiagnosable by using standard genetic tests for single-gene disorders. Thus, pediatricians must become familiar with this technology, what it can and cannot offer, and its technical and ethical challenges. Here, we address the concepts of human genomic analysis and its clinical applicability for primary care providers. *Pediatrics* 2013;132:S211–S215

## INTRODUCTION

The term “genetic testing” can encompass the detection and characterization of any analyte that is determined or predicted by one’s inheritance. In the broadest sense of the term, genetic testing includes, for example, measurements of cholesterol levels or plasma metabolites such as amino acids and microscopic examination of a peripheral blood smear for sickle-shaped red cells or spherocytes. These tests reflect the effects or signs of genetic hyperlipidemias, errors of amino acid metabolism, and hemolytic anemias, respectively. Such biochemical genetic testing has been available for many decades. However, the most fundamental form of genetic testing clearly is the direct analysis of the genetic material itself (chromosomes, genes, or the entire genome), which falls under the domains of cytogenetics and molecular genetics. Biochemical genetics, cytogenetics, and molecular genetics are the 3 laboratory subspecialties recognized by the American Board of Medical Genetics.

The advent of clinical cytogenetics dates to 1959, with the discovery that children with Down syndrome have 47 chromosomes in their cells rather than the normal 46, the abnormality being an extra copy of chromosome 21 (trisomy 21). Since that time, countless defects in chromosomal number, structure, and position have been observed and associated with various congenital (and neoplastic) disorders. In a sense, chromosomal analysis (also known as karyotype analysis) can be considered the first genome-wide test because it includes all the nuclear genetic material of the cell. Such analysis is limited, however, by the resolution of the light microscope; any deletion, duplication, insertion, or translocation that is too small to be observed under standard magnification will go undetected. More recently, the field of cytogenetics has

been enhanced by the development of molecular techniques, particularly fluorescence in situ hybridization: the use of labeled DNA probes that hybridize to intact chromosomes to reveal sub-microscopic deletions or duplications of a particular region of the genome. Even more recently, chromosomal microarrays, consisting of as many as 1 million DNA probes bound to a solid support, have been used to interrogate regions of the entire genome. This method is capable of detecting deletions and duplications of chromosomal regions at much finer resolution than afforded by standard karyotype. In fact, the power and diagnostic yield of chromosomal microarrays are so much greater that the American College of Medical Genetics and Genomics has recommended that this test replace karyotype analysis as the standard first-tier test in the diagnostic evaluation of patients with congenital malformations, nonspecific dysmorphic features, developmental delay/intellectual disability, and autism.<sup>1</sup>

Finally, a new technology called next-generation DNA sequencing has placed within reach the ultimate, most fundamental, and highest-resolution genetic analysis currently conceivable: the precise identification and ordering of all 6 billion nucleotides in the human genome. This technology, more than any other, has opened the way to true genomic medicine.

It is important to point out some semantic distinctions related to the evolving field of genomics. Most agree that the term “genetics” refers to the study of single genes in isolation. Many refer to “genomics” as the study of all genes in the genome and the interactions among them and their environments, as noted in the Executive Summary. However, others use the term “genomics” to describe whole-exome and whole-genome sequencing aimed at identifying mutations in

single genes, as we do in the current paper. Semantics aside, the ultimate goal of the field of genomics is to identify how mutations in multiple genes interact with each other and the environment, a goal that may not be achieved for several years.

## TECHNICAL ASPECTS OF DNA SEQUENCING

For most of the >25-year history of clinical molecular diagnostics, and for the entire duration of the Human Genome Project, DNA sequencing was performed on semiautomated capillary electrophoresis instruments by using the biochemical method known as dideoxy-chain-termination, or Sanger sequencing, after its inventor. This highly accurate, although relatively slow, method can provide the DNA sequence of a targeted and limited region of 150 to 200 nucleotides in a single run of 1 to 2 days. This length of DNA represents only a small fraction of the total length of most genes, let alone the entire genome, which explains why sequencing the first human genome took 13 years and cost approximately \$3 billion. 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 using the hybridization of specific oligonucleotide primers to just that region. The primers serve as start sites for DNA polymerase to make complementary strands that terminate whenever a particular dideoxynucleotide derivative is incorporated into the elongating strand, and the resulting sequence is deduced by measuring the sizes of the terminated fragments on the capillary electrophoresis instrument. In contrast, next-generation, or massively parallel, sequencing breaks up the whole genome into >300

million small fragments, which are then universally primed for DNA polymerase copying by using 4-color fluorescently labeled nucleotides and are analyzed on instruments that take instantaneous snapshots of the added nucleotides after each round of synthesis across the entire genome. Each fragment is copied, or “covered,” between 10 and  $\geq 100$  times, and the aggregated fluorescent photos of all of the synthesis products comprise “cluster arrays” representing 4 to 5 terabytes of computer data. The alignment software then reconstructs the entire genome from these hundreds of millions of DNA fragments. Depending on the particular instrument used, the method can provide the sequence of 3 to 10 gigabases of DNA in a single run,<sup>2</sup> easily enough to cover the entire haploid or diploid genome of an individual. The instruments and reagents are expensive, but the net cost per sample per run has been decreasing over the last few years and now stands at \$2500 to \$6000 per individual, depending on depth of coverage and other factors. This cost is well within the range of currently accepted single-gene sequencing tests, such as those for *BRCA1* and *BRCA2*. Thus, genome sequencing is now a legitimate test to consider for certain clinical situations.

### EXOME SEQUENCING VERSUS GENOME SEQUENCING

Most laboratories currently using next-generation sequencing are applying it to just the small subset (~1.5%) of the total human genome that codes for proteins (ie, the exome, which comprises ~230 000 exons). Exome sequencing involves an additional step beyond those described in the previous paragraph: an exon-capture step by which the coding regions are selected from the total genome DNA by means of hybridization, either to a microarray or in solution. Because most inherited

disorders are believed to be due to mutations in the coding regions, this approach allows laboratories to focus exclusively on those regions and eliminate the tremendous mass of non-coding DNA in the genome, which would otherwise add greatly to the sequencing load and produce 100 times as much data to interpret. The downside, however, is that exon-capture techniques are not 100% efficient: 3% to 5% of the exons will not be captured and sequenced. Therefore, the commonly used term “whole-exome sequencing” is actually a misnomer. If some of the missing exons are deemed essential for ruling out a particular suspected diagnosis, they must be specifically targeted and sequenced by using any of a variety of work-around techniques based on polymerase chain reaction.

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. Examples of commonly ordered gene panels, consisting of 20 to 100 genes per disorder, are shown in Table 1. A key advantage to this approach is that it eliminates the possibility of unwanted, off-target results (as discussed in The “Incidentalome” section), but it has the disadvantage that most of the existing gene panels yield positive mutation results in significantly less than one-half of the tested patients with the associated phenotype.

**TABLE 1** Disease Gene Panels That Use Next-Generation Sequencing

Hypertrophic cardiomyopathy
Dilated cardiomyopathy
Hereditary arrhythmias (channelopathies)
Retinitis pigmentosa
Albinism
Intellectual disability
DNA repair defects
Skeletal dysplasias
Disorders of sexual development
Hearing loss

### DIAGNOSIS AND DISCOVERY THROUGH NEXT-GENERATION SEQUENCING

Whether conducted by using a whole-exome or whole-genome approach, the application of next-generation sequencing to children who appear to have a syndrome but are undiagnosed has already produced some dramatic successes. Some of these are in the form of new gene discovery, whereas 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 using traditional genetic testing methods. In just the past 3 years, new causative genes for at least 150 heritable disorders have been discovered by using these methods.<sup>3</sup> Often this 1 test will put an end to the long and expensive diagnostic odyssey that these patients (and their parents) typically endure. 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 2 or 3 individual genes 1 at a time.

### THE “INCIDENTALOME”

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 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.”<sup>4</sup> 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; that is, the discarding of those variants judged, by a number of rules, as likely to be nonpathogenic. Even after that, however, it is not uncommon to be left with several hundred variants that are potentially worrisome, and sorting through them may require the combined efforts of molecular biologists, clinical geneticists, and bioinformaticists. Even more concerning, from an ethical perspective, is the incidental discovery of “off-target” mutations. These are deleterious changes in known genes that are irrelevant to the patient’s current phenotype and the reason for ordering the test but which 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. 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; however, what is the liability of failing to disclose this incidentally detected risk, either for the future health of the child or the present health of the parent who transmitted it? This example illustrates only 1 of the many difficult ethical questions raised by this technology and highlights the need for informed consent for this new testing.

## **INTEGRATION OF GENOMICS INTO PRIMARY CARE**

### **Research Discoveries and Clinical Advances From Genomic Advances, Early and Now**

There have been many clinical advances in the diagnosis of individuals with suspected genetic conditions since the aforementioned technologies became available for clinical use. The use of fluorescence in situ hybridization

revolutionized our ability to detect common microdeletion syndromes such as deletion 22q11.2 (previously known as DiGeorge syndrome and velocardiofacial syndrome), Prader-Willi syndrome, Angelman syndrome, and Williams syndrome.

When chromosomal microarrays became clinically available, the floodgates opened, and we began to identify microdeletions and microduplications that were of unknown significance. However, with time, the significance (or insignificance) of some of the microdeletions and microduplications has become clearer, as more and more individuals are found to have these copy number variants. For example, microdeletions involving 16p11.2 have been identified in individuals with varying degrees of intellectual disability and behavioral issues (eg, autism spectrum disorder). The penetrance of this microdeletion, however, seems to be incomplete, because some individuals with it are reportedly normal. This microdeletion may well become 1 of the examples of a truly genomic variation; that is, it is disease-causing only in conjunction with mutations in other genes and/or specific environmental influences.

The use of whole-exome and whole-genome sequencing has revolutionized the way researchers identify mutations in single genes (eg, Mendelian disorders). In the first example, described in 2009, whole-exome sequencing was used to identify the causative gene for Freeman-Sheldon syndrome, a disorder characterized by multiple contractures.<sup>5</sup> In a groundbreaking study, the authors used whole-exome sequencing and, by comparing sequence variants that were present only in the 4 affected individuals, were able to identify the causative gene. Since that time, this technology has been used to identify the genes responsible for at least 100 conditions. In 2012, this technology became clinically

available for individuals with disorders that are presumed to be genetic in origin. However, the effectiveness of this technology for diagnosis involving a single individual, compared with its power when the genomic information from several individuals with the same condition is available, has yet to be determined.

As stated previously, next-generation sequencing is clinically available in the form of panels for certain findings that are known to have an underlying genetic basis (Table 1). The use of such panels has proven to be a cost-effective way to simultaneously interrogate several genes from a single patient’s sample. In the past, the genetics provider would often start with 1 gene and have testing performed by  $\geq 1$  laboratory on a reflex basis.

### **Timing of Primary Care Integration of Genomics**

It is likely that whole-exome and whole-genome sequencing and genomics, in the strict sense of the word, will eventually be integrated into primary care. When and how this approach will be integrated, however, is still up for discussion. Some have proposed that it will be more cost-effective to obtain whole-genome sequencing data, store that data, and interrogate the sequence as needed. When that sequencing data should be obtained (whether in the newborn period or at some later point when genetic information may first become helpful in providing health care) is yet to be determined. We must also determine how that information will be stored so that it is “safe” and who will be able to interrogate an individual’s genome information. Clearly, the medical geneticist will continue to assist the primary care provider in deciding what genetic or genomic testing, if any, is indicated and how to interpret the results.

## DIAGNOSIS, TREATMENT, AND PREVENTION

For all medical conditions that health care providers attempt to diagnose, the goal is to maximize effective treatment and, ultimately, to prevent the condition from ever occurring in the first place. Therefore, although genomics is currently used mostly for diagnosis, as we gain a better understanding of the underlying molecular mechanisms of various genetic and genomic disorders, we should also be able to use genomics to tailor our treatment of these conditions.

The use of genomics in the prevention of genetic conditions currently involves providing accurate genetic counseling to at-risk individuals so that they can use this information for family planning. Some couples choose not to have a child if there is a risk that that child will have a genetic condition. Others choose to use currently available forms of prenatal diagnosis, including chorionic villus sampling and amniocentesis, if the mutations are known. Increasingly, preimplantation genetic diagnosis is being used by couples for whom familial mutations have been identified. In the near future, interrogating cell-free fetal DNA from the maternal serum will likely be widely clinically available and will allow for noninvasive prenatal genetic diagnoses. Prenatal genetic diagnosis is used not only to decide whether to continue a pregnancy but also to plan for delivery.

For example, if a couple knows that they are having an infant with a lethal condition, they can plan to deliver close to home where family support is nearby. If, however, the fetus is diagnosed with a genetic disorder that is likely to require multiple evaluations and procedures after birth, delivery at a tertiary care center may be indicated.

## DIRECT-TO-CONSUMER TESTING

Soon after microarray and sequencing techniques advanced to the point of general availability, commercial laboratories began to market to the public “personalized testing” of an individual’s genome, regardless of whether there was a medical indication for such testing.<sup>6</sup> This direct-to-consumer (DTC) testing may not involve appropriate pretest counseling, which can help the individual to fully understand the potential limitations of the test and, more importantly, to recognize the potential “harms” that may attend testing (the previously mentioned incidental findings, off-target mutations, and variants of unknown significance within the genome). Interpretation of test results and posttest counseling are also critical to the meaningful use of the information by the individual (and perhaps other family members). Test results provided by DTC laboratories usually are couched in terms such as “slightly increased risk” or “moder-

ately increased risk,” which individuals in a population challenged by “numeracy” may have difficulty understanding and using. Even when posttest counseling is offered, it is often rudimentary. Unfortunately, the continued increases in the speed and efficiency of genomic sequencing, the decreases in cost, and the ability to perform it on specimens easily collected at home are only going to make DTC genetic test offerings more global and prevalent in the near future. The American College of Medical Genetics and Genomics has recommended strongly that “a certified medical geneticist or genetic counselor be involved in the process of ordering and interpreting”<sup>7</sup> such tests and their results. However, in the absence of some sort of government regulation or oversight, restricting individuals’ access to such services will be difficult, if not impossible.

## CONCLUSIONS

Whole-genome/whole-exome sequencing is a powerful approach to genetic diagnosis when used in the appropriate circumstances and supported by the infrastructure of expert clinical and molecular geneticists, genetic counselors, and other patient-care services available at tertiary care centers. Sequencing should not be viewed as a platform for personal entertainment or be taken lightly, given the potential harms and challenges discussed here.

## REFERENCES

- Manning M, Hudgins L; Professional Practice and Guidelines Committee. Array-based technology and recommendations for utilization in medical genetics practice for detection of chromosomal abnormalities. *Genet Med*. 2010; 12(11):742–745
- ten Bosch JR, Grody WW. Keeping up with the next generation: massively parallel sequencing in clinical diagnostics. *J Mol Diagn*. 2008;10(6):484–492
- Bamshad MJ, Ng SB, Bigham AW, et al. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet*. 2011;12(11): 745–755
- Kohane IS, Masys DR, Altman RB. The incidentalome: a threat to genomic medicine. *JAMA*. 2006;296(2):212–215
- Ng SB, Turner EH, Robertson PD, et al. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*. 2009;461(7261): 272–276
- Bender L, Silverman LM, Dinulos MB, Nickel J, Grody WW. Direct-to-consumer genotyping: are we ready for a brave new world? *Clin Chem*. 2010;56(7):1056–1060
- American College of Medicine Genetics Board of Directors. ACMG statement on direct-to-consumer genetic testing. *Genet Med*. 2004;6(1):60

**Whole-Exome/Genome Sequencing and Genomics**  
Wayne W. Grody, Barry H. Thompson and Louanne Hudgins  
*Pediatrics* 2013;132;S211  
DOI: 10.1542/peds.2013-1032E

<b>Updated Information &amp; Services</b>	including high resolution figures, can be found at: <a href="http://pediatrics.aappublications.org/content/132/Supplement_3/S211">http://pediatrics.aappublications.org/content/132/Supplement_3/S211</a>
<b>References</b>	This article cites 7 articles, 1 of which you can access for free at: <a href="http://pediatrics.aappublications.org/content/132/Supplement_3/S211#BIBL">http://pediatrics.aappublications.org/content/132/Supplement_3/S211#BIBL</a>
<b>Permissions &amp; Licensing</b>	Information about reproducing this article in parts (figures, tables) or in its entirety can be found online at: <a href="http://www.aappublications.org/site/misc/Permissions.xhtml">http://www.aappublications.org/site/misc/Permissions.xhtml</a>
<b>Reprints</b>	Information about ordering reprints can be found online: <a href="http://www.aappublications.org/site/misc/reprints.xhtml">http://www.aappublications.org/site/misc/reprints.xhtml</a>

American Academy of Pediatrics

DEDICATED TO THE HEALTH OF ALL CHILDREN®



# PEDIATRICS®

OFFICIAL JOURNAL OF THE AMERICAN ACADEMY OF PEDIATRICS

## **Whole-Exome/Genome Sequencing and Genomics**

Wayne W. Grody, Barry H. Thompson and Louanne Hudgins

*Pediatrics* 2013;132;S211

DOI: 10.1542/peds.2013-1032E

The online version of this article, along with updated information and services, is located on the World Wide Web at:

[http://pediatrics.aappublications.org/content/132/Supplement\\_3/S211](http://pediatrics.aappublications.org/content/132/Supplement_3/S211)

Pediatrics is the official journal of the American Academy of Pediatrics. A monthly publication, it has been published continuously since 1948. Pediatrics is owned, published, and trademarked by the American Academy of Pediatrics, 345 Park Avenue, Itasca, Illinois, 60143. Copyright © 2013 by the American Academy of Pediatrics. All rights reserved. Print ISSN: 1073-0397.

American Academy of Pediatrics

DEDICATED TO THE HEALTH OF ALL CHILDREN®

