Diagnostic Clinical Genome and Exome Sequencing

Leslie G. Biesecker, M.D., and Robert C. Green, M.D., M.P.H.

SEQUENCING OF THE GENOME OR EXOME FOR CLINICAL APPLICATIONS, hereafter referred to as clinical genome and exome sequencing (CGES), has now entered medical practice. Several thousand CGES tests have already been ordered for patients, with the goal of establishing diagnoses for rare, clinically unrecognizable, or puzzling disorders that are suspected to be genetic in origin. We anticipate increases in the use of CGES, the key attribute of which — its breadth — distinguishes it from other forms of laboratory testing. The interrogation of variation in about 20,000 genes simultaneously can be a powerful and effective diagnostic method.

CGES has been hailed as an important tool in the implementation of predictive and individualized medicine, and there is intense research interest in the clinical benefits and risks of sequencing for screening healthy persons; however, current practice recommendations do not support the use of sequencing for this purpose, and for that reason we do not further address it here. We have also limited this overview of CGES to the analysis of germline sequence variants for diagnostic purposes and do not discuss the use of CGES to uncover somatic variants in cancer in order to individualize cancer therapy.

Clinicians should understand the diagnostic indications for CGES so that they can effectively deploy it in their practices. Because the success rate of CGES for the identification of a causative variant is approximately 25%, it is important to understand the basis of this testing and how to select the patients most likely to benefit from it. Here, we summarize the technologies underlying CGES and offer our insights into how clinicians should order such testing, interpret the results, and communicate the results to their patients (an interactive graphic giving an overview of the process is available with the full text of this article at NEJM.org).

TECHNICAL OVERVIEW AND LIMITATIONS OF CGES

Detailed technical descriptions of sequencing can be found elsewhere, and we provide a graphical summary of one method of CGES in Figures 1 and 2. Regardless of the specific technology that is used, the process begins with the extraction of DNA from white cells, after which the DNA is broken into short fragments, the sequences of which are determined with the use of one of various sequencing technologies. The sequencing instrument generates millions of short sequence reads, which are strings of data representing the order of the DNA nucleotides, or bases, in each fragment. These sequence reads are then aligned to specific positions in the human genome reference sequence (see Glossary) with the use of computers. Similarities and differences between the patient’s sequence and the reference sequence are tabulated, and a computerized determination of the specific genotype (A, C, G, or T) at each position in the exome or genome is performed, resulting in an output file along
with information representing the number of sequence reads generated (depth of coverage) and the accuracy of the genotype at each position. The output file is computationally filtered in accordance with the clinical objective of the test and the preferences of the laboratory. Typically, the file is filtered for variants that are rare or have not previously been reported (because it is reasoned that a common variant cannot cause a rare disease), variants predicted to cause a loss or altered function of a gene, and variants previously reported to cause disease.10,11

CGES is most useful for the detection of single-nucleotide substitutions and insertions or deletions of 8 to 10 nucleotides or smaller; it is less accurate for other types of genomic variation (Table 1). The yield of sequence reads is inherently uneven across the exome (or genome) — typical results provide adequate coverage of 85 to 95% of the targeted sequence. With exome sequencing, there is also variable coverage of flanking intronic regions, which may include disease-causing variants that affect the splicing of messenger RNA encoded by the gene (splice variants).

**INDICATIONS FOR ORDERING CGES**

CGES is currently indicated for the detection of rare variants in patients with a phenotype suspected to be due to a mendelian (single-gene) genetic disorder, after known single-gene candidates have been eliminated from consideration or when a multigene testing approach is prohibitively expensive. Patients can be of any age but are commonly children, since many genetic conditions are manifested in childhood; evaluations are performed because parents are searching for the cause or for information to guide management and treatment and desire accurate information regarding the risk of recurrence, as noted below.

The preparation for ordering CGES should include four key elements: gathering information on family history, systematically evaluating the patient’s phenotype, searching medical literature and databases, and obtaining informed consent. A thorough family history should be obtained to assess whether there are similar or related phenotypes in other family members, as well as to evaluate and assess the inheritance pattern. The patient (and other apparently affected family members) should be evaluated for other potentially relevant manifestations. For example, if the primary presentation is autism and CGES is being considered by a neurologist, the patient should also be carefully examined for respiratory, cardiac, renal, skin, and dysmorphic abnormalities. With a family history and a comprehensive phenotype in hand, a literature review or syndrome database search should be performed (Table 2) to determine whether the patient’s presentation matches a rare but stab-
A

DNA

Exon Intron Exon Exon Intron Exon Intron

Gene A

Intergenic noncoding sequence

Gene B

Base pairs

Fragmentation

B

Artificial DNA Linkers Added to End of Each Fragment, the Fragments Denatured, and Coding Sequences Selected

Fragment 1

Fragment 2

Fragment 3

C

DNA linkers

Noncoding sequences discarded

Replication and Amplification of Exon Fragments

Fragment 1

Primer

Forward strand

Reverse strand

D

Replication and Amplification

DNA Sequencing

E

Imaging

Analysis

Results of exome sequencing

Patient sequence

Sequence reads

Reference sequence

Sequence reads

Reference sequence
Figure 1 (facing page). Schematic Overview of Exome Sequencing.
Exome sequencing targets the approximately 1% of the genome that is made up of exons, which encode protein sequence. The DNA from
the patient (Panel A) is isolated and broken into fragments (Panel B); the DNA fragments are coupled to artificial DNA linker segments (Panel C),
and the fragments are selected with the use of artificial DNA or RNA baits that are complementary to targeted DNA (not shown). The sequencing
process starts with the binding of the end of each DNA fragment to a solid matrix and in situ amplification (Panel D), and the DNA frag-
ments are then sequenced on the slide in a series of reactions in which a complementary nucleotide with one of four colored fluorescent dyes
is added to each cluster of identical molecules (Panel E). The identity of the colored fluorescent indicator of each cluster is imaged with a laser
and a camera coupled to a microscope, the fluorescent indicator is removed, and the cycle is repeated to generate a nucleotide sequence read
that is 75 to 150 nucleotides in length. The sequence reads are aligned to a reference DNA sequence (Panel F), and a genotype call for each
position is made. In this example, most of the positions are homozygous reference sequence, but one position is called as heterozygous A/T.
This figure illustrates one widely used sequencing technology, but it is not intended to endorse that technology over other methods.

Figure 2. Schematic Comparison of Exome and Genome Sequencing.
Panel A shows the targeted nature of exome sequencing, with sequence reads concentrated over the coding portions
of genes. This is in contrast to genome sequencing, shown in Panel B, in which the sequence reads are nearly ran-
domly distributed over the entire genome. Each approach has advantages over the other, some of which are listed in
the two panels.

Advantages of Exome Sequencing:
- Provides higher coverage of exons
- Costs less
- Is currently offered by more laboratories

Advantages of Genome Sequencing:
- Is more sensitive and accurate than exome sequencing for
detecting structural variation, such as insertions, deletions,
and translocations (although both have limited sensitivity)
- Includes non-exonic regulatory regions, although these are
usually irrelevant to mendelian variation
- Identifies common variants in intronic and intergenic
regions (most common complex-disease risk variants and
some pharmacogenetic variants)
There is a spectrum of genomic variants, from nucleotide insertions and deletions of at least 8 to 10 bp through copy-number variants, that are less effectively assayed by current CGES technology.

Aneuploidy Down’s syndrome, Turner’s syndrome

Structural variants Chromosomal translocations associated with spontaneous abortions

Epigenetic alterations Prader–Willi syndrome, Beckwith–Wiedemann syndrome

* There is a spectrum of genomic variants, from nucleotide insertions and deletions of at least 8 to 10 bp through copy-number variants, that are less effectively assayed by current CGES technology.

Evaluating a CGES Result

The outcomes of CGES analysis vary widely. In some CGES reports, a single causative variant is asserted to be the likely cause of the disease, whereas in others, multiple candidate variants are identified that must then be evaluated by the ordering clinician or by consultants. In many cases, no plausible variants are identified.

The two main considerations for evaluating CGES results are their analytic validity and their clinical validity. Analytic validity is a measure of the likelihood that the patient actually has the particular genotype shown in the CGES results—that is, the accuracy of the test. Clinical validity, which is much more complicated and challenging to assess, is the determination that a particular disease is truly caused by variants in a particular gene and that the specific variant that has been detected is indeed pathogenic.

Positive CGES findings are highly accurate, but the false negative rate varies according to the genomic region. For this reason, CGES is not yet a substitute for targeted sequencing of suspected genes or gene panels that have been optimized for a particular condition. Most laboratories validate positive CGES results with well-established methods such as polymerase-chain-reaction amplification and Sanger sequencing. Confirmatory Sanger sequencing should be considered when any major medical intervention is being contemplated on the basis of a CGES result, if such confirmation is not routinely provided by the CGES laboratory.

As noted above, determining the clinical validity of CGES results is more challenging than determining their analytic validity. The general approach is to compare variants implicated by CGES with databases of known variation, which are in turn based on reports that describe causal variants as well as associations between variants and disease types that could narrow the focus of genetic testing before CGES is undertaken. If CGES is performed, it may lead to the discovery of a known pathogenic variant or a previously undescribed and apparently pathogenic variant in a gene known to cause human disease. Some laboratories may also report novel, apparently pathogenic variants in genes not yet known to cause human disease. In such cases, additional clinical and laboratory research may be necessary to establish the pathogenicity of the variants.

The strategy for selecting family members to undergo the sequencing and analysis will be influenced by the expected inheritance patterns (e.g., dominant or recessive) and whether other family members with and without similar features are available for phenotyping and genetic testing. In some scenarios, it can be important to test unaffected relatives (Table 3). When a de novo variant is suspected, the relatives most commonly tested are the biologic parents of the patient; if neither parent has the variant and biologic parentage is confirmed, then the variant is indeed de novo. When a recessively or dominantly inherited condition is suspected, parents, siblings, or even distant relatives may be tested, depending on the particular family history and with the understanding that if two family members are more distant, the number of potential candidate variants will be reduced. Thus, sequencing in two affected cousins will yield fewer candidate variants than sequencing in two affected siblings. In other situations, CGES performed in only the affected child, followed by genotyping of just a few variants in affected and unaffected relatives, may show cosegregation of the variant and the disease, which supports the pathogenicity of the variant. Some CGES laboratories can assist by offering advice about the testing strategy for a given clinical scenario, but this is no substitute for a robust understanding of these issues.
and phenotypes. In the literature, however, there are many false attributions of disease to variants, a problem that is in part due to the conflation of association with causation. Clinicians reviewing the results of sequencing should be aware of the possibility of a false attribution of pathogenicity to a variant and should realize that the chances of false attribution are increased in CGES because thousands of genes are tested simultaneously.

The clinical usefulness of identifying the variant that is the cause of a previously undiagnosed syndrome or heritable disorder varies. In some cases, it can lead to a specific treatment or management strategy that dramatically changes the clinical outcome. In the majority of cases in which the finding does not change clinical management, treatment, or prognosis, it may still be useful because it can end an expensive, potentially invasive, and stressful diagnostic odyssey. The identification of the causative variant may provide accurate estimates of recurrence risk and facilitate preconception intervention or prenatal diagnosis for the affected patient or affected or at-risk relatives. In adult-onset disease, one of the most useful outcomes of successfully identifying the causative variant is the subsequent detection of presymptomatic, at-risk siblings for whom screening or preventive therapy might improve the clinical outcome. Examples include enhanced surveillance or prophylactic surgery for patients found to have a genetic susceptibility to cancer.

Pretest counseling is particularly important, to maintain realistic expectations for finding the causative variant and to alert the patient or family that in most cases, a positive result is unlikely to change treatment or management decisions or to improve the prognosis. In addition, the patient should be advised that incidental findings unrelated to the reason for testing may be found and reported, as described below. It may also be important to discuss the cost of the test with the patient. As is the case with many medical services, assessment of the cost is complicated by many factors. The published billing charge for CGES in most laboratories is in the range of $4,000 to $15,000 per patient, with some laboratories offering lower per-person charges for family testing. To put this in perspective, the per-person charge for sequencing of an exome may be only two to four times the published billing charge for some single-gene sequencing tests, which is why exome sequencing can be more efficient in a number of clinical scenarios. Some laboratories have reported that third-party payers are reimbursing for this testing, but practices vary widely, and patients should understand this in advance.

**Table 2. Examples of Online Databases to Assist Clinicians in Differential Diagnosis or Candidate-Gene Identification for Rare Syndromic Disorders before CGES Is Performed.**

| Free access (or an available free-access version) |
| HuGE Navigator (http://hugenavigator.net/HuGENavigator) |
| Human Gene Mutation Database (www.biobase-international.com/product/hgmd) |
| Online Mendelian Inheritance in Man (www.omim.org) |
| Phenomizer (http://compbio.charite.de/phenomizer) |
| SimulConsult (www.simulconsult.com) |
| **Subscription or fee required for access** |
| Isabel (www.isabelhealthcare.com/home/default) |
| London Medical Databases (http://lmdatabases.com) |
| POSSUM (www.possum.net.au) |

**Interpreting and Communicating CGES Results**

Clinicians should review the CGES results delivered by the laboratory geneticist and place the findings into context with other relevant medical considerations. Sometimes an identified variant will spur additional history taking or an additional examination of the patient, which may reveal clinical features of a previously unrecognized syndrome or lead to the conclusion that the variant is not related to the disorder in the patient (Table 4).

In some cases, the CGES report from the testing laboratory identifies a causative variant (or two variants for a recessive disorder) in a single gene that is considered sufficiently pathogenic and specific that a diagnostic association with a heritable disorder is strongly supported. Such a conclusion by the laboratory geneticist is typically based on the integration of the submitted clinical information with information on diseases associated with the identified variant. In this case, as with all laboratory tests, the order-
Table 3. Genetic Considerations in Deciding Which Relatives Should Undergo CGES.

| Autosomal dominant inheritance suspected |
| Testing is most valuable in the trio consisting of a child and both biologic parents if the disorder may be caused by a de novo variant of a gene associated with a disorder that is caused by heterozygous (monoallelic) mutations. |
| If a phenotype with autosomal dominant inheritance within a family is being evaluated, sequences from two distant relatives with the phenotype will be more valuable than sequences from two close relatives with the phenotype. Sequences from distant relatives are recommended because there will be fewer variants shared solely by chance in close relatives. |
| Autosomal recessive inheritance suspected |
| Testing can also be valuable in the trio of a child and both biologic parents if gene variants inherited in an autosomal recessive pattern are being considered. |
| If there is consanguinity, then testing the trio may be less helpful, because the child’s sequence will already have large areas of homozygosity. |

Table 4. Examples of How to Evaluate the Pathogenicity of a Variant.

| Variant supported as pathogenic |
| A CGES report for a boy with apparently isolated intellectual disability identified a novel, previously unidentified variant in the gene ATRX. This gene is associated with the α-thalassemia mental retardation syndrome. In response to this report, the clinician orders hematologic testing, which identifies a subtle thalassemia phenotype. This additional testing strongly supports the variant identified in the CGES result as the cause of the intellectual disability in the patient. |
| Variant not supported as pathogenic |
| A CGES report identified a variant in the gene NF2 in a 2-month-old infant with congenital, bilateral sensorineural hearing loss. The variant is present in the Human Gene Mutation Database as a disease-causing variant associated with neurofibromatosis 2. However, that database entry is based on a single report that did not specify whether the patient with the variant was a case patient, a patient with a suspected case, or a control. A review of neurofibromatosis 2 in GeneReviews (www.ncbi.nlm.nih.gov/books/NBK1201/) shows that this disorder typically causes unilateral hearing loss with an onset in young adulthood, not bilateral deafness with an onset in infancy. This post-hoc assessment — the absence of support in the literature for causality of the variant combined with data from GeneReviews on the clinical features known to be caused by mutations in NF2 — suggests that the evidence for the pathogenicity of this variant is weak, and the variant is unlikely to explain the child’s phenotype. |

CGES can generate results unrelated to the indication that prompted sequencing, and these results may be clinically useful.\textsuperscript{19,22} As with all testing, it is important to constrain such evaluations to avoid unnecessary future testing and expense. To this end, the American College of Medical Genetics and Genomics has recommended that the laboratories providing CGES routinely seek and report to the ordering clinician specific variants in a minimum set of 56 clinical features.
genes representing 24 disorders that are highly medically actionable. It is estimated that 1 to 3% of patients undergoing CGES have such a finding, which may be outside the expertise of the clinician who ordered the test. A clinician with expertise in the identified disorder should review the personal and family history for other manifestations of the disorder. Such a result may necessitate referral for proper counseling and medical guidance. In cases in which the personal or family history may suggest a specific genetic disease but no etiologic or likely etiologic variant is found, it is essential to communicate to the patient that CGES is not a comprehensive test for all disease-associated variants and that an absence of incidental findings does not mean that specific, indicated testing is unnecessary. Some laboratories routinely perform incidental-finding analysis, and others offer an opt-out for such an analysis; in either case, the ordering clinician should be aware of this issue and should obtain consent and counsel patients appropriately.

**SUMMARY**

CGES is a useful diagnostic test for a number of clinical situations, and it is already being used by clinical geneticists and other specialists. The indications and approaches we outline here are sure to evolve over time, as more data are generated for various clinical disorders, data interpretation is improved, and CGES is studied in new clinical situations (e.g., in neonatal medicine). Clinicians can feel confident ordering this test if they become comfortable with the evaluations, both before and after testing, and the processes that are required to maximize its usefulness, and if they are familiar with its limitations.

The opinions expressed in this article reflect the views of the authors and may not represent the opinions or views of any institutions with which they are affiliated.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

We thank the members of the Clinical Sequencing Exploratory Research Consortium, for valuable ideas and discussion on this topic, and Shamil Sunyaev, Ph.D., for helpful suggestions.