

## Gene testing revolution: Disease prediction results skyrocket for whole genome and whole exome sequencing

How can scientists best identify common molecular defects that lead to disease? What are the latest options to screen patients with suspected genetic disorders? These are among the biggest challenges in medicine and genomic research.

Now [new research](#) suggests that focusing on the whole exome sequence—the protein-coding region of our DNA that makes up about 1.5 percent of each person’s genome—may offer the best balance between cost and effectiveness in testing patients.

On June 26, 2000 then President Bill Clinton announced that the first draft of the human genome would “revolutionize the diagnosis, prevention and treatment of most, if not all, human diseases.” The idea was that with the ability to obtain human genetic sequences, scientists would be able to identify the genetic variants that either caused or increased the risk for all sorts of common diseases. Clinics could then test for these mutations to predict future illness and start preventative treatment before you, the patient, ever suffered any symptoms.

The main approach adopted by the research community to find these disease-causing mutations was to identify disease-associated single nucleotide variants through a series of what are called [Genome Wide Association Studies](#) (GWAS). GWAS use a streamlined method that looks at a very small set of predefined points—approximately 100,000 bases out of the total 3 billion—scattered across the length of the genome. Then, for each point, researchers check if there is a specific single nucleotide variant (an A instead of a C or G) that is seen more often among samples from individuals with the inherited disease than those without. These disease-associated single nucleotide variants are also known as point mutations. For example, most direct-to-consumer (DTC) genetic test results, like those from [23andMe](#), report probabilities of disease risk based on findings from GWAS. But how valuable is that to any one individual?

Those who follow the research-side of genomic medicine may have heard that, despite the initial hope, nearly ten years of GWAS [have failed to find specific genetic mutations](#) that actually predict disease. Nearly 2000 variants have been associated with 300 diseases; however, for almost all of these associations the probability of disease in individuals who harbor the mutation varies so widely that knowledge of the mutation is useless for guiding clinical intervention. It turns out that the mutations being identified by GWAS are not the key drivers of disease.

So, rather than helping to guide clinical interventions and reassure patients, GWAS results sometimes made things worse or provided only flimsy predictions of uninteresting ailments. Telling someone that the global population has, for example, an eight percent chance of developing a specific disease is not reassuring to a patient who wants to know what disorders she might be susceptible to. Researchers and clinicians who depended upon this breakthrough research appeared to have hit a brick wall.

Now both the research and clinical picture is brightening. Those who follow the research even closer will

have heard, with little explanation, that a new breed of studies based on genome sequencing may ultimately be able to identify genetic disease causes that the GWAS approach has not. The focus is not on our collective genome but on an individual's actual genome—either her entire genome or just her exome, which is the portion of the genome known to encode proteins (approximately 1.5 percent of the human genome).

GWAS is very limited for at least two reasons. It only looks at approximately 0.01 percent of the possible sites of the genome for disease-causing mutations and it only focuses on one type of mutation: a point mutation. Other types of disease-causing variations include the spurious insertion or deletion of swathes of a genome as well as gene duplications. These are completely missed by single point mutation searches and may partly explain why so many GWAS studies have failed to find a disease-causing variant.

Perhaps a point mutation only causes disease if it is coupled with one of these DNA insertions/deletions nearby. Perhaps a disease is not caused by a specific mutation in a gene, but rather by having an additional copy of the gene. What's more, the point mutations are taken completely out of context with no information of the surrounding region that could explain why the mutation would be causing the disease. These patterns can only be seen by sequencing swathes of the genome to get a view of the region surrounding a specific mutation.

In recent years, many prominent geneticists, led by [David Goldstein](#) of Duke University, have been focusing on whole genomic sequencing. Far more comprehensive than single variant association studies, these new forms of genomic research are stirring hopes that we may yet find genetic causes of disease that have been unobservable in GWAS. By comparing whole genomes of affected and unaffected individuals scientists would not only be able to survey the entire landscape for all types of variants, but also see the location of the disease-associated variant in relation to other important genes it may affect. This additional information can then possibly tell us something about the mechanism by which disease occurs.

As [noted](#) by biologist Jacqueline Beal, while a single point variant association study merely implicates a mutation in the 'crime' of causing disease, sequencing the entire region allows researchers to gather evidence of "how the 'crime' was committed, i.e., exactly how the protein was rendered nonfunctional and why a patient might be susceptible to or actually have the disease in question."

Despite these advantages, the whole sequence approach has been impractical for most researchers due to the cost and time necessary to sequence a single genome, much less those of the hundreds of people necessary for a decent disease-variant association study. However, new technologies and international efforts to pool genomic data such as the [1000 genome project](#) are now making whole genome comparison studies increasingly common. Indeed, whole genome sequencing has become a buzz term of sorts with familiar promises to change the pursuit of the causes/treatments of genetic disease.

Even so, the cost of whole genome sequencing has remained impractical in some clinical settings. For this reason, many experts believe that streamlined versions of whole genome sequencing—focusing on the protein-coding exome or sequencing to look for specific gene markers- are necessary for getting results in the clinic. Whole exome sequencing is not able to identify the non-coding variants associated

with diseases, which can be found using [whole genome sequencing](#), but the benefits, particularly costs, might be worth the trade off in many circumstances.

Clinical gene testing has undergone significant evolution, progressing from single variant, full sequence of individual genes, and selected gene panel, to the now increasingly popular whole exome/genome sequencing techniques. Whole exome or whole genome testing approaches using next-generation technology appears to be ideal for more complex diseases, as these techniques enable the evaluation of millions of sequences concurrently. Recent studies include those that have given new insights into [epilepsy](#) and [bladder cancer](#).

A [study](#) published early this month in the New England Journal of Medicine suggests that whole-exome sequencing is often the best balance of cost and effectiveness, offering a success rate as high as a 25 percent in solving hereditary disease mysteries.

“For years we’ve known that whole-exome sequencing can identify new disease-causing mutations,” Yaping Yang, a clinical geneticist at the Baylor College of Medicine in Houston and a study coauthor, told Nature Medicine. “But this puts it on the map as a tool for clinical medicine.”

Whole exome sequencing currently costs about \$7,000, and that price is dropping quickly. It’s estimated that about half of current health plans cover the cost of whole exome sequencing; for insurance companies, identifying diseases or disease proclivities early is far more cost effective than treating patients with an advanced disorder.

When looking for a specific condition or disease, the even more targeted single gene mapping approach could prove to be the appropriate testing technique. Dr. Richard Gibbs, Professor of Molecular & Human Genetics and director of the Baylor College of Medicine Human Genome Sequencing, [compares](#) the three options to photograph resolution; patients can get high-resolution photographs (whole genome sequencing), medium-resolution (exome/genes only) or a view from 1,000 feet (gene markers).

Now that the data is available, the next few years of research will tell if the promise of having genetic predictors of disease can be fulfilled after all.

***Katherine Wendelsdorf is a computational biologist based in San Francisco, CA.***

#### **Additional Resources:**

- “[A Decade Later, Genetic Map Yields Few New Cures](#),” New York Times
- “Whole-exome sequencing rises to top in largest clinical application for undiagnosed disease,” Nature
- “[Clinical Whole-Exome Sequencing for the Diagnosis of Mendelian Disorders](#),” New England Journal of Medicine