

# Tackling hereditary diseases with the help of grid computing

FEATURE | MARCH 27, 2013 | BY AUDREY GERBER

*It is said that you cannot escape your genes. But with the help of grid computing, a leading provider of molecular diagnostic products and services from Israel is making it easier to outrun them.*

The key to understanding, treating and eventually preventing hereditary diseases lies in identifying and mapping the genetic mutations which cause them, and understanding the underlying cascade of biological events that can occur when mutations are present. In order to do this, researchers need to decipher DNA sequences.

By determining the precise order of the four [nucleotides](#) within a strand of DNA, scientists are uncovering the basic building blocks of life and revealing, quite literally, what it is that 'makes us tick' — or, more importantly, what happens when one of these mechanisms that makes us tick goes awry.

Determining the sequence of four nucleotides, AGCT, is not quite as simple as it sounds though. The coding sequence of a single gene can be made up of thousands of nucleotides, and many genes may be associated with a single disease. That's a lot of data to sequence. Rapid speed sequencing attained by capillary-electrophoresis-based Sanger sequencing, invented in 1997, greatly accelerated this process, giving scientists the tools they needed to sequence genomes of numerous types and species of life — including the human genome. This information is behind some of the greatest discoveries in genetics, medicine and pharmaceuticals over the last decade.

However, [Sanger sequencing](#) has inherent limitations in throughput, scalability, and speed. The advent of an entirely new technology, '[next-generation sequencing](#)' (NGS), offers a fundamentally different approach that is ushering in a new age of genomic science and completely new cost paradigms that make genetic technologies more accessible.

In principle, NGS technology is similar to [capillary electrophoresis](#) (CE). The bases of a small fragment of DNA are sequentially identified from signals emitted as each fragment is re-synthesized from a DNA template strand. NGS, however, extends this process across millions of reactions simultaneously. This rapid sequencing of large stretches of DNA base pairs may span entire genomes at once, producing huge gigabases of data from these base pairs in a single sequencing run.

So what's the next step? Although NGS may be used to sequence the whole genome, selected genomic sequences or genes may be enriched to sequence specific genes only. This approach is being adopted to sequence all the genes associated with a specific disease — at the same cost as sequencing a single gene by CE, and with a much lower effect on patient privacy. [Pronto Diagnostics](#), a Tel-Aviv-based developer of molecular diagnostic products and services, is working to extend the power of affordable desktop NGS instruments and bring more powerful diagnostic capabilities to clinical feasibility.

However, NGS data output has more than doubled each year since it was invented, meaning that huge computing power is needed to take full advantage of this technology. In 2007, a single sequencing run produced a maximum of around one gigabase of data. By 2011, that rate nearly reached a terabase of data in a single sequencing run.

Pronto Diagnostics does not have in-house bioinformatics capabilities of the scale required for NGS-based research. In fact, few commercial companies do outside of global pharmaceutical conglomerates or academic-based laboratories and institutions. Consequently, the company turned to IsraGrid, Israel's National Grid Initiative (NGI), for assistance. IsraGrid is a cooperative initiative of three government ministries: Industry & Trade, Finance and Defense, and Israel's Council for Higher Education. It was initiated in the framework of the National Infrastructures for R&D Forum, spearheaded by leading high-tech industrialists, to provide Grid and Cloud computing infrastructure for important research. And to extend capacity, IsraGrid is a partner in the European Grid Initiative (EGI), providing full access to this enormous resource.

Prior to sequencing DNA, techniques — broadly termed 'target enrichment (TE) strategies' — are often used by researchers to selectively capture genomic regions of interest from DNA samples. In order to develop TE assays and accompanying analysis tools for additional disease groups, and to create a database of non-pathogenic genomic variants, Pronto Diagnostics is aligning TE-NGS results of selected genes with the human genome, and analyzing the data from many perspectives.

A typical TE-NGS results file is a text file between three to seven gigabases in size, which must be compared and aligned to the human genome sequence, which is a 3.17-gigabase file. This process could take several days on a standard quad core personal computer, but on the grid it can be allocated and divided into many parallel threads and completed in up to 12 hours — or overnight. Moreover, a typical NGS run sequences many DNA samples in parallel, and the grid enables parallel analysis rather than performing one run and only upon its completion starting another. Also, many additional manipulations of the data in huge files, such as sorting, comparing to other large data files for annotation or filtering and more, can also be divided into smaller steps that can be run in parallel and in exponentially less time. To date, Pronto Diagnostics used approximately 60 computing cores for each grid job that was submitted, adding up to a total of some 100,000 CPU hours — and growing as the research continues.

In contrast to less costly and commonly used assays that scan only the known mutations along the tested genes, TE-NGS will enable the identification of novel mutations. Because conventional Sanger sequencing is so expensive, most laboratories and providers only sequence the exons of one or two genes. The TE-NGS approach enables the parallel sequencing of all the genes known or suspected to be associated with the disease, including introns that include known mutations. Pronto Diagnostic's work is leading toward diagnostic grade NGS analysis assays to discover and diagnose all the genetic mutations associated with these diseases, rather than just the more common ones, and making this both an accurate and affordable option for more and more patients.

**Average:**

Your rating: None Average: 4 (3 votes)



As easy as AGCT! Image courtesy [George Gastin](#), [Wikimedia Commons](#).

"Access to the computing power of Israel's national grid initiative is saving us time and allowing us to leverage valuable resources to the maximum. What would have taken us months to process takes just days. We hope to reach time to market objectives and be able to offer a very cost-effective and powerful diagnostic tool that will make a real difference in people's lives." *Lilach Friedman, senior genome analyst, Pronto Diagnostics*

IsraGrid is Israel's National Grid Initiative supported by Israel's Council of Higher Education and the Ministry of Industry and Commerce and acts as an operating unit of IUCC — Israel's [Inter-University Computation Center](#). Part of the [European Grid Infrastructure \(EGI\)](#), IsraGRID provides core services to enable efficient e-science research in a wide range of scientific fields, including physics, life sciences and more via access to production quality European e-infrastructure based on grid/cloud technologies.

[About the Author »](#)

---

### Audrey Gerber

Audrey Gerber is the Communications Officer for [IUCC](#), Israel's [Inter-University Computation Center](#) and [IsraGrid](#).

**RELATED TERMS:** [DNA](#) [Europe](#) [genetics](#) [genomics](#) [grid computing](#) [Israel](#) [sequencing](#) [biology](#) [Middle East](#)

---

## Comments

[ADD NEW COMMENT](#)

---

### Post new comment

**Subject:**

**Comment: \***

By submitting this form, you accept the [Mollom privacy policy](#).

SAVE	PREVIEW
------	---------