NOVARTIS

Surfing the wave of big data analytics

The cause of a rare, inherited, often fatal kidney disease in many patients remained elusive for years, despite multiple attempts to solve the mystery. Then in 2012, a team of researchers cracked some of the unsolved cases in just six weeks.

By Tia O'Brien | Oct 27, 2013

The difference? This time the team, a collaborative effort by Novartis and the Necker Children's Hospital-Imagine Foundation in Paris, applied cutting-edge DNA sequencing technology and sophisticated analytics to the problem. The same approaches that Google uses to compile and search through billions of Web pages allowed researchers to rapidly analyze the genomes of patients and their relatives. In three families, the team pinpointed a surprising cause of the disease. Previously undetected mutations in a single gene (called LMX1B) triggered focal segmental glomerulosclerosis (FSGS), a disease that scars the kidneys' filtering system.

"Big data was the game changer," says one of the team leaders, Joseph Szustakowski, head of Bioinformatics in Biomarker Development at the Novartis Institutes for BioMedical Research (NIBR) in Cambridge, Mass.

> The hope is that if you have a well-designed experiment, a well-formulated hypothesis and a solid plan for analysis, then the needle pops out of the haystack when you query the data.
>
> **Joseph Szustakowski**, Head of Bioinformatics in Biomarker Development at the Novartis Institutes for BioMedical Research

Bioinformatics, a specialized branch of data science, is making a positive impact in modern medical research and drug discovery. For patients, it is helping to accelerate both the development of promising treatments and the potential of personalized medicine — therapies based on an individual's unique genetic profile.

In just 10 years, the time and cost required to sequence genomes has plummeted. Scientists find themselves with potentially valuable new data, but the torrent of information is relentless. If printed on standard office paper and stacked, the raw sequencing data of just one patient's genome would top an 80-story building. Until researchers mine these vast data pools, breakthroughs remain trapped inside, awaiting discovery. "Without the analysis piece, the data is meaningless," says Stephen Cleaver, executive director of Informatics Systems at NIBR in Cambridge.

Data scientists join drug research teams

To make sense out of this wave of data, scientists are developing sophisticated ways to store, retrieve and analyze it. A new breed of "data scientist" such as Cleaver and Szustakowski is working to re-invent the traditional drug research team. Instead of biologists, chemists and clinicians working in silos, pharmaceutical companies such as Novartis are assembling collaborative, cross-disciplinary teams. These teams include data scientists, drawing on their expertise in computer science and statistics to sift through information and attempt to extract answers to pressing questions. They collaborate with biologists and clinicians to develop a clear hypothesis and then put it to the test.

For the FSGS experiment, the team hypothesized that the heritable form of the disorder was caused by a rare mutation in a protein-coding gene. To prove it, they began by selecting key individuals in an affected family and sequencing their DNA. Next, the team ran the raw DNA sequences through large clusters of computers, comparing them with reference genomes to generate a list of mutations present in FSGS patients. The team took a closer look at the list, querying the data to determine which mutations were meaningful.

"The hope is that if you have a well-designed experiment, a well-formulated hypothesis and a solid plan for analysis, then the needle pops out of the haystack when you query the data," says Szustakowski. "And that's exactly what

happened in the FSGS case."

How large was the FSGS haystack? Szustakowski compares it to looking for a 1-meter needle in a haystack that stretched from Earth to the sun. Researchers now are exploring whether the same mutations linked to FSGS play a role in other kidney disorders. They must also determine how the mutations affect the activity of the gene and the protein it produces, knowledge that might eventually lead to the development of a targeted therapy.

Maximizing the potential of big data analytics and bioinformatics

In addition to helping facilitate rational drug discovery, big data has the potential to someday open up new frontiers for improved patient care, ranging from personalized medicine to faster, safer, less expensive clinical trials and other innovations.

Maximizing bioinformatics' potential will require overcoming several hurdles. While these FSGS cases could be explained by a single mutation, many diseases are more complex. Other disorders — such as rheumatoid arthritis and autism — are thought to stem from a combination of rare and common genetic variants as well as environmental factors. Data scientists must continue to develop new tools and methods for analysis to meet these challenges.

But there's a severe shortage of talent. Google, Facebook and other Silicon Valley stars siphon away top candidates. "We can't find enough data scientists to keep up with the demand for our services within NIBR," says Cleaver. When he recruits, he touts the satisfaction of designing large, complex datasets that directly affect patients. "It could come down to saving someone's life," he says.

Over in Szustakowski's lab, he recalls the afternoon when, in the course of three hours, a scientist exploring other disorders tracked down several suspect mutations. "Each time I'd hear a 'Yes!' and 'Joe, you've got to come see this!' " Such groundbreaking discoveries are the reason he opted for a career that could power a paradigm shift in medicine.

Source URL: *https://uat2.novartis.com/stories/surfing-wave-big-data-analytics*

List of links present in page

1. https://uat2.novartis.com/stories/surfing-wave-big-data-analytics
2. https://uat2.novartis.com/tags/category/our-labs
3. https://uat2.novartis.com/tags/authors/tia-obrien
4. http://www.nibr.com/
5. https://uat2.novartis.com/careers/career-search?
   search_api_fulltext=bioinformatics&items_per_page=10&job_division%5B%5D=DIV_RE&field_job_posted_date=All