

Embracing big data to understand complex diseases ^[1]

Discovery ^[2]

Author: Elizabeth Dougherty

In 1967, in response to a life-shortening heart disease epidemic in the Western world, the Icelandic Heart Association (IHA) launched the Reykjavik Study, a research program intended to get at the roots of heart disease and find ways to stop it. Researchers recruited 30 000 Icelanders to join the study, which morphed into another study that still runs to this day.

Researchers from Novartis and IHA recently mined data from the multi-decade studies – and gathered additional data – to make a new discovery that sheds light on how heart disease and other diseases develop with age. It turns out that the blood-stream contains multiple complex regulatory networks of proteins that orchestrate health, a finding published online ^[3] in *Science* on August 2, 2018. The Novartis-IHA team found 27 networks, each a unique set of proteins whose levels rise and fall together in the blood stream in synchrony, a bit like a chorus singing a distinct song. When a network isn't operating in perfect harmony, disease can develop.

“We're talking about hundreds of proteins that are produced by multiple organs all changing their levels together as a single regulatory group,” says co-first author John Lamb, Director of Genetics at the Genomics Institute of the Novartis Research Foundation (GNF). “These regulatory networks are very strongly related to genetics and disease.”

The networks can't be seen or heard in a physical exam. They aren't traceable with a dye and they can't be inferred by reading the genome. Rather, they became evident only when the research team leveraged big data – in this case, a complex and massive conglomeration of data from medical exams, gene sequencing and measures of proteins in blood culled from IHA's data banks.

“This is a different way of thinking. It's not classic linear thinking,” says co-first author Valur Emilsson, head of systems medicine at IHA and professor at the University of Iceland. “Diseases are complex, so you cannot shy away from the complexity.”

The research team believes these regulatory networks could hold the keys to finding new ways to detect, monitor or even treat age-related diseases that take a major toll on individuals, communities and societies. Heart disease alone affects more than 17 million people and results in nearly a third of deaths worldwide each year, according to the World Health Organization.

Big data roots

Back in 1967, genomic sequencing and sophisticated imaging platforms were not available to scientists. Nevertheless, the IHA researchers forged ahead, tracking the health of individuals

participating in the Reykjavik Study by conducting physical exams at regular intervals.

In doing so, they learned a lot about cardiovascular risk factors. By the mid-2000s, cardiovascular mortality in Iceland had dropped by 80%.

By then, medical technology had advanced. The team launched a new study, the AGES-Reykjavik (Age, Gene/Environment Susceptibility-Reykjavik) Study to take advantage of it. This study was co funded by IHA and the National Institute on Aging.

The AGES-Reykjavik Study recruited nearly 5 500 of the original Reykjavik Study participants. In addition to sequencing the genomes of participants and continuing to collect clinical data, the team also collected blood samples and used modern imaging tools to perform scans of every organ system, measuring white and gray matter in the brain, brown and white fat in the abdomen, atherosclerosis in the heart and more.

“The imaging is practically a virtual autopsy,” says Vilmundur Gudnason, director of the IHA Research Institute and professor at the University of Iceland. “We have enormously detailed information on our participants’ organs and can identify disease at very early stages, before it becomes noticeable by a doctor.”

Big data evolution

Meanwhile, Lori Jennings, co-lead author on the Science study, was working with Lamb and a team of Novartis scientists in La Jolla, California, in the US to study how cells communicate. They wanted to identify every protein that could end up with an active role in the bloodstream, including proteins that cells make and intentionally release into the blood.

Jennings and the team developed a high-throughput system to coax cells to make and secrete thousands of proteins and to capture and purify them so that they could be studied and characterized.

This work dovetailed with the efforts of a team of scientists from the Novartis Institutes for BioMedical Research (NIBR) in Cambridge, Massachusetts, as well as SomaLogic, a company based in Boulder, Colorado, in the US. SomaLogic had created technology to accurately detect protein levels in blood serum. Sensitivity had been a challenge because some proteins in blood are plentiful and others are rare.

Some proteins, for example, are 100 million times more abundant than others. Technologies traditionally used to detect proteins in serum, such as mass spectrometry, miss the rare proteins.

The SomaLogic technology at that time recognized 1 000 proteins, a large number but still only a fraction of those suspected to appear in serum. So the Novartis teams, which had knowledge about the missing proteins, collaborated with the company to expand the technology’s coverage.

By 2015, the collaborative team had developed a version of the technology that measured the levels of more than 4 000 proteins in a single droplet of serum.

Order from complexity

Now armed with the new technology, IHA went to work to expand its dataset by measuring protein levels in blood samples that had been collected from the Icelanders in the AGES-Reykjavik Study.

The resulting dataset was incredibly rich. For each participant, it included DNA data, a decades-long history of health and disease, and information about the levels in blood of more than 4 000 proteins. “Proteins were a missing factor in the past because we didn’t have the technology to measure them,” says Emilsson.

The protein networks the research team discovered appear to bridge the gap between risk genes and diseases such as heart disease and metabolic syndrome. Each of the 27 networks contains one or a handful of hub proteins. Gene mutations that alter the level of a hub protein – throwing it out of tune from the rest of the chorus – have a substantial influence on a network’s behavior. Some mutations make disease more likely, others less.

“Biology by definition is a high-dimensional, complex setting. In this dataset alone there may be a million gene variants per person, over 4 000 proteins, and about 200 clinical measures of disease,” says Lamb. “We drew on all of the advanced computing, mathematics and statistics skills we had to uncover and understand the patterns we’re seeing.”

These protein networks might help explain medical cases that have baffled physicians for years.

“Why does one person who is overweight and has other risk factors for liver disease have a pristine liver while an otherwise healthy person develops liver disease? There must be a genetic susceptibility,” says Jennings, who is now Director of Disease Biomarker Research at NIBR. “These protein networks and their link between gene variants and disease could help us answer that question.”

For instance, one out-of-tune protein could drive the development of liver disease despite an individual’s overall wellness. A medicine that brings that protein back in tune could treat or even prevent the condition. More research will reveal the potential. “We’ve only just scratched the surface,” says Lamb.

Main image: GarryKillian/Shutterstock

Tags:

[Chronic Illness](#) [4]

[Digital](#) [5]

[Diseases](#) [6]

[Novartis Institutes for BioMedical Research](#) [7]

[New Scientific Discoveries](#) [8]

[Scientific Research](#) [9]

Source URL: <https://www.novartis.com/stories/discovery/embracing-big-data-understand-complex-diseases>

Links

[1] <https://www.novartis.com/stories/discovery/embracing-big-data-understand-complex-diseases>

- [2] <https://www.novartis.com/stories/discovery>
- [3] <http://science.sciencemag.org/content/early/2018/08/01/science.aag1327>
- [4] <https://www.novartis.com/tags/chronic-illness>
- [5] <https://www.novartis.com/tags/digital>
- [6] <https://www.novartis.com/tags/diseases>
- [7] <https://www.novartis.com/tags/novartis-institutes-biomedical-research>
- [8] <https://www.novartis.com/tags/new-scientific-discoveries>
- [9] <https://www.novartis.com/tags/scientific-research>