IBM

Healthcare

Jefferson
Philadelphia University +
Thomas Jefferson University

HOME OF SIDNEY KIMMEL MEDICAL COLLEGE

Business challenge

What causes some people to develop diseases and not others? The attempt to find an answer is driving groundbreaking research and leading pioneers to challenge traditional approaches to treatment.

Transformation

The Computational Medicine Center at Jefferson is breaking new ground in the understanding of disease by analyzing huge amounts of biological data with the help of high-performance computing.

Dr. Isidore Rigoutsos
Director of the Computational Medicine Center
Jefferson (Philadelphia University + Thomas Jefferson University)

Business benefits:

Pushes the boundaries of knowledge, anticipating new breakthroughs in healthcare

Supports the development of diagnostics and therapies that could boost positive outcomes

Removes barriers to scientific exploration through data-driven research

Jefferson
Deeper understanding of disease enables radically new approaches to diagnosis and treatment

Jefferson (Philadelphia University + Thomas Jefferson University) is a distinctive comprehensive national university setting a new standard for 21st-century professional education. It has 7,800 students, more than 4,000 faculty members and offers approximately 160 undergraduate and graduate programs on multiple campuses. Its unique Nexus Learning model focuses on collaborative inter-professional and transdisciplinary approaches to learning supported by design and systems thinking, innovation, entrepreneurship, empathy, and the modes of thought central to the liberal arts and scientific inquiry.

"In a relatively short time, we recovered years' worth of data, which had been generated by dozens of people in the Center. And, before long, we were again able to shift our attention back to our research work."

Dr. Isidore Rigoutsos,
Director of the Computational Medicine Center
Jefferson (Philadelphia University + Thomas Jefferson University)

Share this

Delving deep into the mysteries of disease

Disease can strike anyone at any time. Now, an emerging complex network of variables could provide clues about why it affects some people and not others. Specifically, Jefferson scientists have discovered that non-coding RNAs, which are functional RNA molecules that are transcribed from DNA but not translated into proteins, play an even bigger role in human disease than previously thought.

Dr. Isidore Rigoutsos is the Founding Director of the Computational Medicine Center and Professor in the Department of Pathology, Anatomy and Cell Biology, the Department of Biochemistry & Molecular Biology and the Department of Cancer Biology at Jefferson (Philadelphia University + Thomas Jefferson University), as well as a member of the Sidney Kimmel Cancer Center. He picks up the story: "We know proteins play a critical part in the normal functioning of the human body, helping cells do what they need to do. We also know that non-coding RNAs contribute to human conditions and diseases such as cancer.

"Since 2002, my work has been focused on the study of non-coding RNAs, including microRNAs, pyknons,

piRNAs, and, in recent years, transfer RNA fragments. In particular, I am interested in the questions surrounding the biogenesis of such non-coding RNAs, their mechanism of action and the identification of their targets, and in clarifying their roles in the onset and progression of disease. I find regulatory sequences that are tissue-specific and human-specific particularly intriguing."

With a diverse academic background that spans physics and computer science, the co-creation of the Computational Biology Center in 1992 and establishment of the Bioinformatics and Pattern Discovery Group in 1998 when employed by IBM Research, and creation of Jefferson's Computational Medicine Center, Dr. Rigoutsos is uniquely positioned to discuss the importance of data-driven approaches to this area of research.

"For more than 25 years, I have been developing and using computational approaches to analyze very large biological datasets, study genomic architecture, and understand the genetics of disease," says Dr. Rigoutsos. "Many of the existing diagnostics and disease treatments are based on a relatively simple model of how cells and non-coding RNAs work. We suspected that the reality is a lot more complicated than that, and set out to follow the data and see where it led us."

Letting data lead the way

In 2010, Dr. Rigoutsos established the Computational Medicine Center at Jefferson with the goal to understand how non-coding RNAs relate to disease. As data-driven analysis was central to the Center's activities, the organization engaged IBM in early 2010 to help it deploy powerful, resilient infrastructure to support its research.

"For data-driven discovery to succeed, you need lots of data, you must be able to analyze it quickly and you need to be able to retrieve older data constantly," comments Dr. Rigoutsos. "Working with IBM, we built a networked, parallel file system that would allow us to store large amounts of data, access it quickly, and back it up reliably."

"The resilience of the IBM solution was tested a little over a year ago when an accident severed a large portion of our storage system. Thanks to IBM Spectrum Protect we were able to bounce back with no data loss."

Dr. Isidore Rigoutsos, Director of the Computational Medicine Center, Jefferson (Philadelphia University + Thomas Jefferson University)

The team's focus has been on 'short' non-coding RNAs, i.e. molecules with 50 or fewer nucleotides. In the early 2000s, there was only one known such category, the microRNAs, whose typical length is 22 nucleotides. To make a microRNA, a specific region of DNA, generally 70-100 nucleotides long, is transcribed into an initial RNA molecule. This longer RNA is processed and, eventually, produces the "active" 22-mer microRNA. What makes microRNAs important is their ability to regulate simultaneously the abundance of multiple proteins in the cell. Pioneering work in the early 2000s by Drs. George Calin and Carlo Croce who were at Jefferson at the time showed that the dysregulation of microRNAs can cause disease.

For more than two decades – microRNAs were originally discovered in 1993 – it was believed that each region of DNA that harbors the "recipe" for the precursor microRNA gives rise to a single active microRNA, in all tissues and in all people. Nearly 10 years ago, scientists observed that each such recipe in the cell makes a multitude of different active microRNAs at the same time: they called these products 'microRNA isoforms' or 'isomiRs.' The sequences of any two isomiRs generally differ from one another by only a few nucleotides. The Jefferson team showed that isomiRs are produced in a controlled manner.

"We began by looking at the short RNA profiles that were part of the 1000 Genomes Project, and focused on the active microRNAs that we could find. Fortunately, all these RNA profiles were from a single cell type and this made the analysis easier," recalls Dr. Rigoutsos. "After studying data from more than 450 people, we discovered that the variations in the produced isomiRs that we encountered in this one cell type depend on attributes that had never been considered before. Specifically, we found that a person's sex, race, and population origin affect the number of the distinct isomiRs that are produced and their abundance.

"This is striking, as these molecules are regulators of protein production, and previous research – at Jefferson, no less – had linked them already to human diseases. The datasets from the 1000 Genomes Project allowed us to discover that some of the attributes of an individual determine the production of isomiRs, introducing a new level of complexity, but also deepening our understanding."

Next, the team extended their analysis to include cancer patients, and confirmed that more variables needed to be added to the mix. Dr. Rigoutsos elaborates: "We used to think that each DNA recipe makes the same active microRNA in every tissue in which it is used. Our work showed that this was not true either: the isomiRs that were produced from a

given recipe and their relative abundance also changed from tissue to tissue. Eventually, we extended our analysis to short non-coding RNA profiles from more than 10,000 cancer patients, which had been collected by the U.S. National Institutes of Health. The data represented 32 different types of cancer and allowed us to confirm en masse that the isomiRs that are produced from each recipe differ among cancers.

"These findings suggest that what we used to view as a single disease may actually be multiple diseases at the molecular level that differ from one person to the next. Some of the differences in isomiR production are likely consequential because the molecules at hand, the microRNAs, are powerful regulators of protein production."

The team did not stop there. They repeated the same analytical steps for another category of short non-coding RNA regulators that has taken research by storm in recent years, the so-called fragments of transfer RNA (tRNA), or tRFs. As with microRNAs, the Jefferson scientists demonstrated that the production of tRFs by tRNA is regimented, and not random. Just like with the isomiRs, the team showed that an individual's sex, race, population origin, as well as tissue, disease, and disease subtype affect which tRFs are produced and at what levels of abundance. The team's tRF findings are particularly relevant for the study of disease because fast-accumulating evidence shows that tRFs have extensive regulatory roles as well.

No-assumption analytics pay off

Dr. Rigoutsos and his team discovered that the number of protein regulators within the human body is substantially higher than previously believed, and the landscape is a lot more complicated. Their findings have direct impact on attempts to deploy precision medicine. Precision medicine is an emerging approach for disease treatment and prevention that considers individual variability in genes, environment, and lifestyle for each person.

"For a long time, the community has tried to explain disease through a reductionist view of regulatory regions in DNA," adds Dr. Rigoutsos. "But it turns out that each of these regions doesn't produce just one protein regulator, but many more, so we need to radically adapt our thinking."

By advancing the understanding of human disease, the team is laying the groundwork for the development of new and more effective diagnostics and treatments. It is also demonstrating the value of a data-driven approach to medical research.

"Traditionally, research in life sciences, and many other fields, was pursued through hypothesis-driven investigations," comments Dr. Rigoutsos. "When you let the data lead the way, you can be less constrained and can entertain bolder journeys that are not limited by what is already known in the literature. High-performance computing is the all-important catalyst that makes such scientific explorations possible. Our own research is a great example of the power of this approach. In fact, our findings were a big surprise, even to us."

The Computational Medicine Center has the high-performance, resilient infrastructure it needs to dive deep into data. Dr. Rigoutsos concludes: "The resilience of the IBM solution was tested a little over a year ago when an accident severed a large portion of our storage system. With IBM Spectrum Protect, we were able to bounce back with no data loss. In a relatively short time, we recovered years' worth of data, which had been generated by dozens of people in the Center. And, before long, we were again able to shift our attention back to our research work. Built-in automation, high availability features and seamless integration between the different components made that possible."

Solution components

- IBM® System Storage® DCS3700
- IBM Spectrum Protect™
- IBM Spectrum Scale™
- IBM Storwize® V5030
- IBM TS3310

Take the next step

To learn more about IBM Storage solutions, please contact your IBM representative or IBM Business Partner, or visit the following website: ibm.com/storage

Connect with us