September 23, 2016 | By Samantha Jones

Debashis Sahoo: From Computers to Cancer Genomes

Debashis Sahoo

Debashis Sahoo, an assistant professor in both the Department of Pediatrics and the Department of Computer Science and Engineering at the University of California San Diego, has been selected as a participating member of the Cancer Genomes and Networks program at UC San Diego Moores Cancer Center.

Members of the Cancer Genomes and Networks research program focus upon three thematic areas: genome instability, human cancer genetics and systems biology. Sahoo, who will focus on the latter two, plans to develop computational models for human cancer and predict important biomarkers and therapeutic targets.

"We have shown that Boolean analysis – a sophisticated data analysis method – provides a platform for such predictions," said Sahoo. "A part of this work is published in the *New England Journal of Medicine*. We will explore the application of this new method in many different types of human cancer. This appointment will provide me resources and numerous collaborative opportunities with cancer experts. Working with the members of the Moores Cancer Center enables a computer scientist like me to develop lifesaving strategies for human cancer."

Sahoo's journey into the world of cancer genomics was borne from an initial curiosity not about cancer cells, but computers. He studied computer science as an undergraduate at the Indian Institute of Technology in Kharagpur, India, initially focusing on theoretical computer science and, specifically, formal verification (FV), an area of computer science that has had a notable impact on hardware and software design. The goal of FV is to either prove or disprove, using formal methods of mathematics, whether or not an algorithm does exactly what it is supposed to do and nothing more, which can yield practical data with the hope of more efficiently solving problems in complex systems.

After completing his undergraduate education, Sahoo headed to Stanford to begin his Ph.D. in the Department of Electrical Engineering. Almost four years into his program, he started doubting the real-world applicability of his work. "Deep into my Ph.D. program I decided to change my focus," he said. "I wanted to focus on much more practical aspects. I was thinking: 'how I can make an impact [on] society?' If I just develop theories and no one implements them, it won't work."

Sahoo hoped to work on cancer biology, but as he knew nothing about cancer he needed to play some serious catch-up. Knowing it would be no small undertaking, he dove right in, enrolling in as many cancer-related courses as he could at Stanford, taking all of the undergraduate pre-medical requirements. He showed great promise in the field, even early on. "After one year, I taught cancer biology to undergrads," he said. These experiences gave Sahoo the confidence to move forward with his goal of using computer science to more efficiently progress medical discovery. "Having both FV and cancer biology training gave me ideas about how to come up with algorithms."

He joined the UC San Diego School of Medicine in 2014 and received an additional appointment at the Jacobs School of Engineering in 2016.

Catching the Eye of the Scientific Community

Sahoo began with B cells., Also referred to as B lymphocytes, they are a type of white blood cell, the progeny of hematopoietic stem cells. A hematopoietic stem cell can be considered the mother of all blood cells, as it is from these cells that each blood cell type arises. During development, stem cells can both self-renew, making more of themselves, and differentiate, becoming a different and more specific cell type.

Sahoo theorized that he could devise a way to computationally predict B cell differentiation, and went to a leader in the hematopoiesis field, Dr. Irving Weissman, to share his idea. Weissman, along with the members of his laboratory at Stanford's Institute for Stem Cell Biology and Regenerative Medicine, were studying how a hematopoietic stem cell can give rise to other blood cell types, such as B cells. Weissman is recognized as one of the premier hematopoiesis researchers of our time, his lab being the first to identify blood stem cells in mice and humans.

"When I presented [Dr. Weissman] with this idea of a computational model that predicts what happens in B cell differentiation, he was surprised," Sahoo recalled. "I said, 'I can make predictions based on public data,' and he said, 'Let's test it'." Sahoo ran the test in Weissman's lab and gave the resulting predictions to Weissman's most established researchers. All of the gene predictions were gathered from human data sets and then tested on mice, yet they still yielded the expected results.

Sahoo said that Weissman, who had been working on blood cell differentiation for decades, was blown away. Sahoo recalled Weismann saying to him, "What I couldn't have done in 40 years you have done in a few seconds." He subsequently invited Sahoo to join his team.

At the time, Sahoo was a graduate student at Stanford. Three years after this meeting with Weissman, he finished his Ph.D. titled, "Boolean analysis of high-throughput biological datasets (PDF)," under David Dill, and Sylvia Plevritis. Following graduation, he left for Stanford's Institute for Stem Cell Biology and Regenerative Medicine, where he was offered a research fellow position to work alongside prominent scientists of varying research backgrounds. There, Sahoo became fully immersed in modeling biology as a complex physical system.

Computer Science Meets Cancer Biology

Cancer arises when the normal process of cell differentiation goes awry; specialized cells that are damaged are replenished by cells that are often referred to as "cancer stem cells." The cells are so-named because they differentiate in a way that mimics what we see in normal

tissue, yet their division is uncontrolled and disorganized, leading to tissue invasion that is often deadly.

"When people are treated and then the cancer comes back, it's because [the treatment] doesn't kill the stem cells," said Sahoo. "If you kill the stem cells you can potentially cure the cancer."

Sahoo was motivated to understand what makes these cells differentiate in such a sporadic fashion. "If you understand the path of differentiation," he explained, "you can find markers of stem cells as well as differentiated cells, and identify specific cancer stem cell markers."

Working on bladder cancer, Sahoo discovered biomarker Keratin 14, or KRT14, which can be used to help predict patient survival. A biological marker, or biomarker, is something that indicates a certain condition, such as cancer, helping a doctor to determine a patient's prognosis. In this case, the gene KRT14 was shown to significantly associate with poor patient survival in bladder cancer.

Currently, the cheapest and easiest way to diagnose a cancer is through simple immunohistochemistry, a method that tests for specific proteins in cells using antibodies known to bind to those proteins. Because genes function primarily by being expressed in their protein form, this method is a way of testing whether or not a gene is still being expressed. Now that the KRT14 gene is recognized as a bladder cancer biomarker, KRT14 protein can be tested for to help predict patient survival.

In high-risk bladder cancer, a patient's bladder is often removed. If the cancer returns, that indicates that the cancer has moved to another region of the body. Sahoo and his colleagues showed that, for a patient whose bladder had been removed, if the cells in the bladder were KRT14 positive, the cancer has more than likely spread to other areas in the body. This was shown to predict a patient's prognosis better than anything had before. It was so exceptional, in fact, that it beat one of the best analyses of the time, called "stage" and "grade." The "stage" of a tumor describes its size and spread from its place of origin, and the "grade" defines the appearance of cancerous cells.

Sahoo's work was ultimately published in the *Proceedings of the National Academy of Sciences* in 2015, and presented as a new way of identifying high-risk bladder cancer patients, yet discovering this bladder cancer biomarker was only the beginning for Sahoo.

Moving Toward a Future of Big Data and Big Discovery

The impact of Sahoo's work lay in the fact that the algorithm he developed was remarkably general. Using this algorithm, one can study both normal and cancerous processes in any tissue, all based on public data, allowing for the prediction of biomarkers in other forms of cancer. Sahoo expects that its eventual impact will be monumental.

Despite this finding, moving toward clinical trials is a lengthy process. "I think it will still take five or six years to happen," said Sahoo. Yet he remains optimistic. "If you predict something and it works in practice, it's the best thing you can have. That was my goal when I changed my field, and seeing it implemented in the clinic was my dream."

More recently, Sahoo applied this same algorithm to colon cancer data, searching for biomarkers of cells differentiating in colon tissue. Of the candidate genes identified, CDX2 ranked as most likely to be implicated. Sahoo and his colleagues found that there was a decreased five-year disease-free survival rate for CDX2-negative colon cancer patients, compared to CDX2-positive colon cancer patients. This result ultimately led the team to suggest that patients with stage II or stage III CDX2-negative colon cancer could benefit from chemotherapy and surgery, not just surgery alone.

Sahoo is currently using his algorithm to work in other fields of biomedical research. He is studying the biology behind HIV and the flu, with the hope of identifying properties that allow for rapid evolution, which makes these viruses so deadly. "We can potentially target some of these viruses in unique ways to treat them," he said. "In some cases, we may be able to eradicate diseases."

Sahoo says he's also interested in researching the microbiome, or the microorganism landscape that exists throughout our bodies. In recent years, the field of microbiome research has exploded, as researchers have found microbiome associations with disease and overall fitness. Sahoo's computational approach has the potential to shed light on these associations. He hopes to use properties of the microbiome that are invariant across tissues of humans to find an underlying, common, biological link among all bacteria that favors an environment, and then use those findings as a tool for disease management.

In terms of future work, Sahoo firmly believes that collaboration is critical. "UC San Diego provided me with a good mix of computer science and biology, helping me establish my career," he said. Sahoo sees collaboration, amongst other things, as a way to leverage the expertise of leaders in the field. "If I talk to people, I'm sure we'll come up with a better way to manage whatever study they are doing."

MEDIA CONTACT

Tiffany Fox, 858-246-0353, tfox@ucsd.edu

UC San Diego's Studio Ten 300 offers radio and television connections for media interviews with our faculty, which can be coordinated via studio@ucsd.edu. To connect with a UC San Diego faculty expert on relevant issues and trending news stories, visit https://ucsdnews.ucsd.edu/media-resources/faculty-experts.