CANCER BYTES

How U-M researchers are using big data to transform cancer care

BY IAN DEMSKY

THE OLDEST KNOWN MEDICAL TEXT ON CANCER DATES BACK MORE THAN 3,500 YEARS TO AN EGYPTIAN TREATISE ON SURGICAL CASES. ACCORDING TO THE UNKNOWN PHYSICIAN, POSSIBLY THE LEGENDARY IMHOTEP, A PATIENT MAY PRESENT WITH BULGING TUMORS OF THE BREAST. IF THESE ARE HARD AND COOL TO THE TOUCH, THE TEXTURE OF "GREEN HEMAT FRUIT," THEN THE WORST CAN BE EXPECTED.

"THERE IS NO TREATMENT," THE ANCIENT AUTHOR NOTES.

The idea of individual doctors examining, diagnosing and tending to individual patients is as old as human history. In recent years, however, something new and exciting has been happening — physicians and researchers, including those at the University of Michigan Rogel Cancer Center, are tapping into the wealth of previously inaccessible information in patients' individual cancers and bringing that data together to help improve care for millions. The umbrella term from this work has come to be known as big data.

"Cancer researchers have actually been using big data and cutting-edge computational methods for quite some time, relative to other chronic diseases," says Bhramar Mukherjee, Ph.D., chair and John D. Kalbfleisch Collegiate Professor of Biostatistics and associate director for quantitative data sciences at the Rogel Cancer Center. "That's because there's so much meaningful data to work with — tumor sequence data, germline genetic data, and all of the -omics: like gene expression, methylation, proteomics, metabolomics and studies of the microbiome. We are rapidly entering the age of omni-omics."

One of U-M's strengths, she notes, lies in the breadth and depth of expertise across campus, which gives rise to powerful collaborations between data scientists, those studying cancer biology and genetics in the laboratory, and clinicians with firsthand experience caring for patients.

A risky business

Not everyone who smokes cigarettes gets cancer. Neither are genetics a clear predictor — 8 out of 10 women who develop breast cancer have no family history of the disease.

Whether an individual develops cancer depends on a complex interplay between underlying genetic factors, environmental factors like exposure to toxins, and lifestyle choices, such as diet, smoking, alcohol use, exercise and tanning.

Mukherjee and her team have a goal of enabling real-time precision prevention in cancer. They are using data drawn from the electronic health records of more than 65,000 U-M patients who have opted into a study called the Michigan Genomics Initiative. The researchers want to better understand not simply which genetic mutations give rise to specific cancers, but also what other medical conditions and test results might serve as early warning signs.

Access to patients' entire health records has provided a rich history spanning years, Mukherjee says, with an average of 27 doctor visits and 31 diagnostic codes associated with each patient.

"We're looking for cases where we can agnostically find a disease, or a cluster or pattern of diseases, that arise long before an actual cancer diagnosis," Mukherjee says. "This is particularly important for cancers for which we do not have good screening tools — like pancreatic cancer and ovarian cancer."

Bhramar Mukherjee, Ph.D.

Ultimately, the goal is to go even further, pairing data from patients' medical records and blood samples with a broad spectrum of information to paint a unique portrait of the factors that give rise to disease: tumor data, dental records, neighborhood information, air quality data, prescription claims, death records and so on.

A central pillar of the group's current analyses is a risk-prediction tool known as a polygenic risk score. As the name implies, it's made by aggregating small contributions to disease risk from individual genetic flaws into an easier-to-work-with grouping that represents a family of mutations.

"The idea is to increase the predictive power of our studies by collapsing multiple risk variants into a single biomarker," says Lars Fritsche, Ph.D., an assistant research scientist in Mukherjee's group, who has led much of the research.

The second pillar is the wealth of unique patient data from the Michigan Genomics Initiative, which launched in 2012.

In 2019, Mukherjee's research team published a study in *PLOS Genetics* with two major findings. Using skin cancer as a test case, they found that polygenic risk scores constructed from large, public datasets accurately lined up with actual patient data — that is, patients with a high genetic risk score for skin cancer tended also to have been diagnosed with skin cancer.

But the more interesting finding was that the data also surfaced other conditions and traits among these patients. For example, the data showed actinic keratosis (scaly skin growths caused by sun damage and known to be pre-cancers) showing up in patients who were later diagnosed with skin cancer. The data was able to map how early the keratosis tends to show up before skin cancer is found.

By showing that the approach can accurately find known cancer precursors, the researchers believe in the future it will help uncover associations that would otherwise go overlooked.

More broadly, the researchers anticipate polygenic risk scores may soon start to be used in the clinic, particularly for monitoring patients at highest risk.

"The question is, can we put some patients on high alert, maybe introduce screening at an earlier time or recommend a behavioral change that can have the biggest impact?" Mukherjee says.

Re-modeling prostate cancer

For years, clinicians have relied on a standard set of tools to figure out which prostate cancer cases are likely to be aggressive and spread: prostate-specific antigen, or PSA, level, tumor stage and Gleason score, a grading system based on how biopsied cells look under the microscope.

Daniel Spratt, M.D., Laurie Snow Research Professor of Radiation Oncology at Michigan Medicine, is part of a team of researchers who want to improve predictions by incorporating patients' genomic data into the equation.

They developed a clinical genomic model that adds in data from a commercial, 22-gene test called Decipher. The model significantly improved upon the predictive power of the standard models used to classify men into a six-tiered system of low, intermediate and high risk, according to findings that the research team published in the *Journal of Clinical Oncology*.

"We found it reclassified about two-thirds of men," says Spratt, the study's lead author. "That tells me that without genomics, we're wrong fairly often about how aggressive a patient's disease is going to be. And that means there are a significant number of patients we can spare the cost and side effects of treatment."

Michigan Genomics Initiative

65,000 patients

EACH PATIENT:

~27 visits
~31 diagnostic codes

tumor data | dental records | neighborhood information

The group's findings showed about half of patients classified with favorable, intermediate risk under National Comprehensive Cancer Network guidelines would be reclassified as low risk by the genomics model. Therefore, the model could increase the confidence of doctors and patients for pursuing active surveillance rather than immediately jumping to surgery or hormone therapy. Meanwhile, about 15% of low risk patients were actually at higher risk when genomics were factored in.

The study also identified a new cohort of patients at very high risk for metastasis who would likely benefit from more aggressive treatments and who would be good candidates for experimental clinical trials.

"Prostate cancer is being left behind in the era of precision medicine," Spratt says. "Adding genomics to our clinical toolkit could radically change the way we perceive and treat localized prostate cancer."

Daniel Spratt, M.D., attends a tumor board discussion with other U-M cancer specialists.

Pan-Cancer Atlas

10 years | 2.5 petabytes of data

33 cancer types | 11,000 tumors

published in 27 papers

air quality data | prescription claims | death records

DATA

Working closely with Todd Morgan, M.D., chief of urologic oncology at Michigan Medicine, Spratt is undertaking new research to further validate the genomics model and develop the data needed to recommend its use as a new standard for clinical practice.

Mapping the future of cancer research

The goal of The Cancer Genome Atlas was simple: to create a new, more comprehensive understanding of how, where and why cancer arises.

The decade-long project combined the efforts of researchers across the globe, assembling and analyzing 2.5 petabytes of data spanning 33 cancer types and 11,000 tumors — enough data to fill up 2.5 million iPhones. Its final report, the Pan-Cancer Atlas, was published in 27 papers across the Cell Press family of journals in 2018.

Thomas Giordano, M.D., Ph.D., Henry Clay Bryant Professor of Pathology at Michigan Medicine, served on the Pan-Cancer Atlas steering committee and co-led TCGA projects on thyroid and adrenal cancer.

U-M researchers also contributed tumor samples and led several other projects, including in esophageal cancer, and head and neck squamous cell carcinoma.

"Overall, we learned a great deal about the molecular basis of cancer and discovered new mutations, which have helped improve molecular diagnostics," Giordano says. "We learned a lot about the consequences of specific mutations and how they drive the biology of tumors. Nearly all of the projects identified three to five molecular subgroups of cancer, some of which had not been previously identified."

In papillary thyroid cancer, for example, two mutations involving proteins in the same signaling pathway were previously believed to be essentially functionally equivalent. But the research data revealed different signaling properties, which has led to changes in how these tumors are classified, and thus improved the care and treatment patients receive, Giordano says.

"Today, when researchers study thyroid cancer — or many other types of cancer — they can compare their results to the atlas and have a robust dataset for comparison," he says.

A good example can be seen in the work of ovarian cancer researcher Analisa DiFeo, Ph.D., associate professor of pathology and obstetrics and gynecology at Michigan Medicine.

She's trying to develop compounds to inhibit a specific microRNA that's enriched in recurrent ovarian cancer tumors — with the idea that blocking it may be able to help prevent the cancer's return.
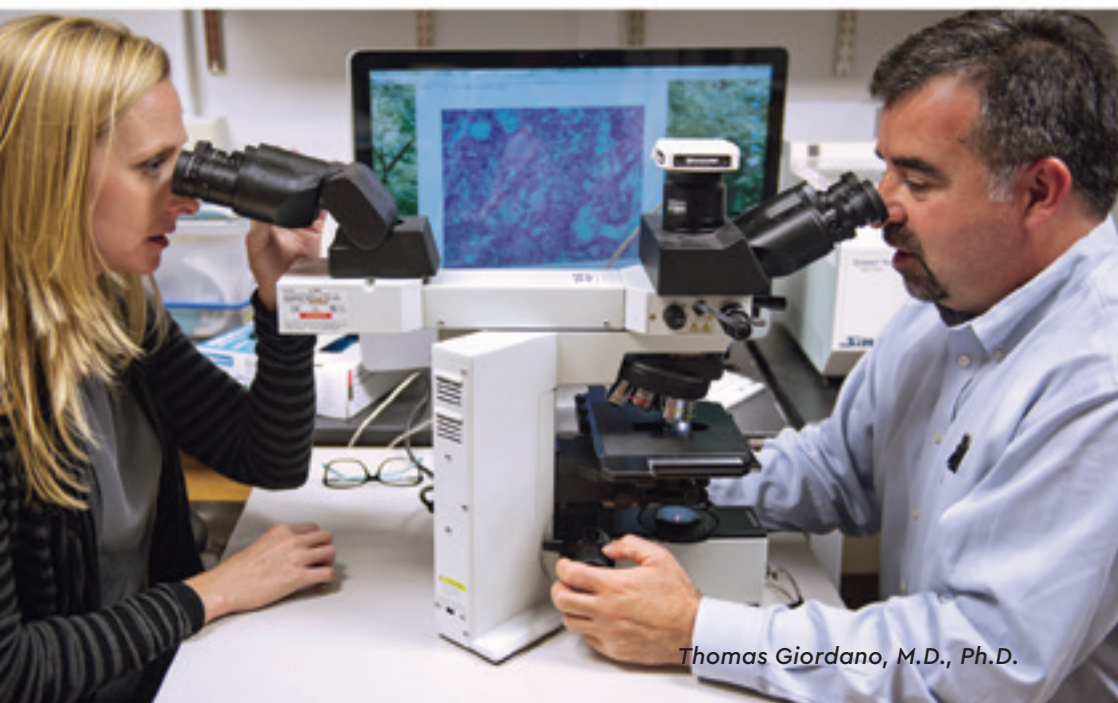
Thomas Giordano, M.D., Ph.D.

In the atlas, DiFeo can see that patients with amplification of this same microRNA across cancer types have significantly worse outcomes.

"So the hope is that if we can find a drug to target this microRNA, it won't only apply to ovarian cancer, but could potentially apply to many other types of cancer," she says.

David G. Beer, Ph.D., professor emeritus of surgery at Michigan Medicine, found a promoter duplication in a protective enzyme in the genome of Caucasians that is associated with reduced expression and may help to explain why they have higher rates of the most common type of esophageal cancer.

"Risk factors like obesity and reflux are happening at the same rate for African Americans and Caucasians, but African Americans are not getting cancer," he says. "It's not just the presence of the genomic duplication, but also these other factors contributing to the damage of the esophagus that contribute to increased risk of developing cancer."

Beer and his colleague Laura Kresty, Ph.D., associate professor of thoracic surgery at Michigan Medicine, have seen promising initial results from research into whether cranberry-derived flavonoids, a type of phytonutrient, could help reduce levels of DNA damage caused by reflux.

"Fundamentally, this data is already changing how patients are treated," Giordano says. "And it's an amazing resource to build upon in the years to come."

Building a Better Mouse Map

Just as The Cancer Genome Atlas has molecularly characterized thousands of human tumors across many cancer types, Kathleen Cho, M.D., is among a group of U-M faculty working to create its equivalent in mice. A Mouse Cancer Genome Atlas, if you will.

"Those of us who work in the field have recognized the tremendous value that the human genome cancer atlas, the TCGA, has provided," says Cho, Peter A. Ward Professor of Pathology, professor of internal medicine and head of gynecologic pathology at Michigan Medicine. "At U-M, we have a lot of strength in developing genetically engineered mouse models of specific cancer types. The question is, how well do the mouse tumors recapitulate their human counterparts? Or, among several leading mouse models, how do we know which one is best?"

In trying to recreate human disease as closely as possible, scientists consider several factors in engineering genetic defects into mice: Do the cancers occur in the right place? Do they look like their human counterparts under the microscope? Do they distribute themselves across the body in the same way when they metastasize? Does variation among individual mice follow a similar pattern to human patients?

"And if we were to look at the DNA, the RNA and the proteins in the mouse tumors, do we see the same kinds of alterations at the molecular level that we see in human tumors as described by the TCGA?" Cho says.

The ultimate goal, she says, would be to develop a new resource for cancer researchers around the globe.

"In my field, for example, there have been several different models developed for high-grade serous carcinoma, the most common form of ovarian cancer," Cho says. "And, of course each group thinks their model is the best one. But if you wanted to move forward with testing a novel therapy or prevention strategy, ideally you'd like to choose the model that has been proven to be the best mimic of its human counterpart.

"It's a lot less expensive to do preclinical studies in animal models than to do human trials, but they do require resources and are time-consuming — so you don't want to be putting your eggs in the wrong basket," she continues. "Ultimately, an inferior model could lead to a sub-optimally designed human clinical trial."

Kathleen Cho, M.D.