the center for quantitative biology

Where Big Data and Basic Science Meet

:: BY TIM DEAN

Like many leading researchers, Michael Whitfield's interest in science began early and has taken a few unexpected turns along the way. "I started by chasing snakes and lizards," he recalls with a chuckle, "but then in college, I worked in a molecular biology lab and got hooked."

Whitfield went to graduate school for biochemistry and biophysics, but as he was finishing his PhD the Human Genome Project was ramping up. He decided to pursue a post-doctoral fellowship in functional genomics in the joint laboratories of David Botstein and Patrick Brown at Stanford University that were leading efforts to measure all of the genes in the human genome in a single experiment, using a process called DNA microarrays.

"I had begun to do some really interesting work on cancer and the human cell cycle, but then I was approached by a private foundation and its very charismatic founder, Sharon Monsky, who had scleroderma and convinced me I should work on this poorly understood disease," says Whitfield, who is today a professor of biomedical data science and of molecular and systems biology at Dartmouth's Geisel School of Medicine, and serves as chair of the Department of Biomedical Data Science at Geisel.

Since joining Dartmouth in 2003, Whitfield has built a successful multidisciplinary research program that uses advanced genomic and computational techniques to better characterize and develop more effective diagnostics and therapies for scleroderma. A rare autoimmune disease of the connective tissue that can cause severe, even life-threatening symptoms—including fibrosis (hardening) of the skin, vascular dysfunction, and inflammation and scarring of internal organs—scleroderma affects primarily women between the ages of 30 and 60.

"Scleroderma is a heterogeneous disease and its heterogeneity has been a major reason why it has been difficult to treat," he explains. "We're now applying novel methods in the lab to understand the molecular underpinnings of the disease, and importantly, to understand the variability across patients. When we combine these approaches, we can figure out the best way to treat a patient based on molecular data about their own personal disease."

Elizabeth Sergison and Fred Kolling prepare samples for a single-cell genomics assay using the 10X Genomics platform. (photo by Kata Sasvari)

the center for quantitative biology

> "Single-cell genomics is a rapidly evolving field; having a dedicated center with resources behind it gives us the ability to bring on new technologies, test them, and figure out which ones are best for our needs."

NEW CENTER LAUNCHED

Last year, with the launching of the Center for Quantitative Biology (CQB), Whitfield seized on an exciting opportunity to make precision medicine technologies more broadly accessible to Dartmouth's diverse scientific and clinical communities.

Supported by a 5-year, $12.5 million grant from the National Institutes of Health (NIH), the Center is funded as an Institutional Development Award (IDeA) Center for Biomedical Research Excellence (COBRE) from the NIH's National Institute of General Medical Sciences. The IDeA program builds research capacities in states that historically have had low levels of NIH funding by supporting basic, clinical, and translational research, faculty development, and infrastructure improvements.

The rapid advancement of high-throughput genomic (genes), proteomic (proteins), metabolomic (nutrients), and immune profiling technologies now provides a breadth and depth of data on individual cells that can be explored to examine basic biological processes, changes in populations of cells or organisms, and the molecular basis of disease.

"The scientific theme of our Center focuses on these 'omics,' in studies that range from whole organisms and tissue biopsies to the detailed genomic analyses of single cells," says Whitfield, principal investigator on the grant, who is co-leading the CQB's Research Administration and Mentoring Core with Jay Dunlap, PhD.

"In essence, what we're trying to do is make high-level computation available to basic scientists that might not have it in their labs, while also helping people who are excellent computational biologists get access to really good wet-lab data to analyze—that partnership is at the heart of our efforts."

The activities of the Center are organized around and guided by four major goals: recruiting and developing new faculty with expertise in computational and experimental approaches, accelerating the interdisciplinary and collaborative research projects of junior faculty, further mentoring of junior quantitative biologists, and developing shared services—through the formation of the Single-Cell Genomics Core and Data Analytics Core—that provide the systems and infrastructure needed to support the Center's work.

Being able to build on the strength of existing programs and leverage shared resources, such as those made available by Norris Cotton Cancer Center (NCCC) and Dartmouth College, has helped the CQB to gather early momentum.

"Single-cell genomics is a rapidly evolving field; having a dedicated center with resources behind it gives us the ability to bring on new technologies, test them, and figure out which ones are best for our needs," says Fred Kolling, PhD, who along with Whitfield serves as co-principal investigator of the Single-Cell Genomics Core. "Things are changing so quickly we need to have our hands in all of these things to stay on the cutting edge."

For example, says Kolling, when the COBRE was started about a year ago, a concept known as spatial transcriptomics, which allows investigators to do single-cell sequencing in the context of a three-dimensional tissue such as a tumor biopsy, was just in its infancy. "It's very clear now that it will be the next big step in the next year or so," he says. "It requires an up-front investment in technology, equipment, and expertise, but that's exactly what the Center is for—to see what that next application is and say, 'Great. We can bring that here.'"

Michael Whitfield meets with Tammara Wood (left) and Kristin Harrington about activities within the CQB's Research Administration and Mentoring Core.

Kata Saswari
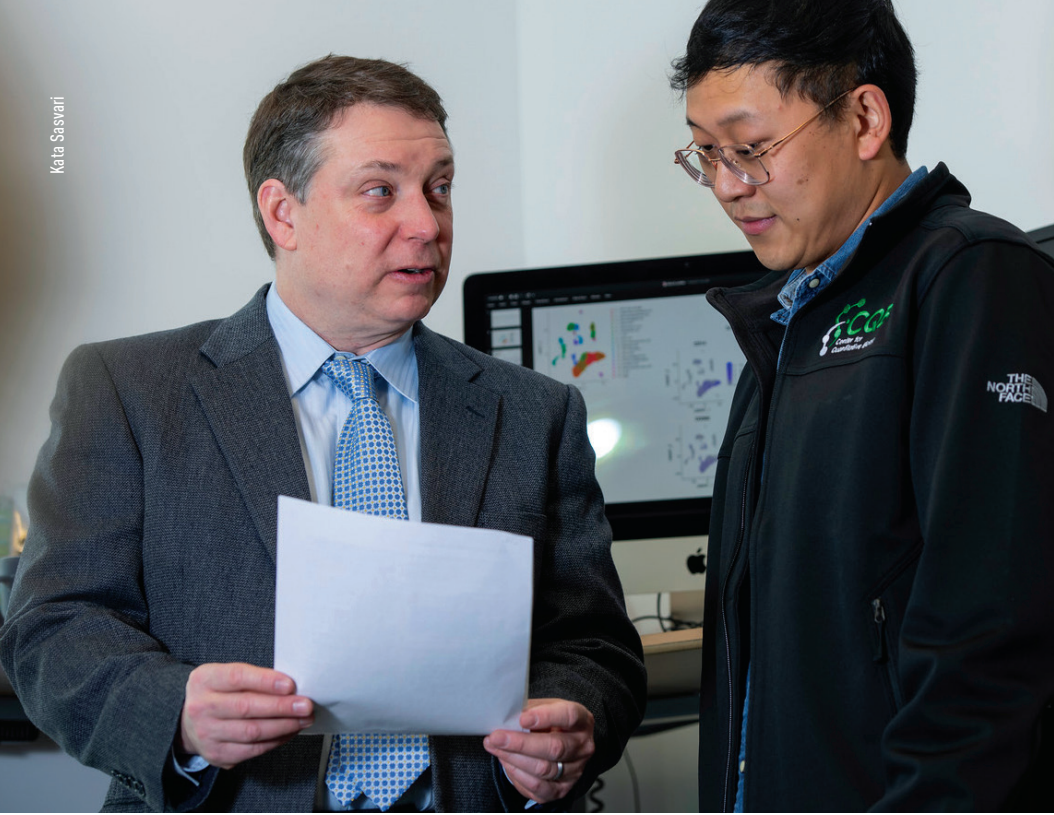
Michael Whitfield (left) looks at single-cell genomics data with research scientist Frank Wang.

A DIFFERENT TACK

In other research, Daniel Schultz, PhD, a project leader on the COBRE, is using microfluidics devices built in his lab along with advanced single-cell sequencing techniques to shed light on a very different problem—the evolution of antibiotic resistance mechanisms.

"In one of the models we study, we look at tetracycline resistance in E. coli bacteria, which helps us understand how bacteria turn these resistance mechanisms on and off in their natural environments in the soil and water," explains Schultz, an assistant professor of microbiology and immunology at Geisel who is part of Dartmouth's Neukom Cluster in Computational Science.

"And then we use pseudomonas aeruginosa infections in cystic fibrosis (CF) patients as a model from a disease perspective, to see what kinds of genetic changes occur within the cells as the bacteria adapts to the human lung environment," he says. "We're able to evolve the bacteria in the lab under different dynamic conditions, by exposing them to varying levels of drugs, to see how well they respond."

Often, says Schultz, when bacteria become optimized to resist a certain drug regimen, there are tradeoffs. "Bacterial cells often have to respond very quickly and it's a costly process for them to undergo—we want to understand where those resulting weak spots may be so we can exploit them and develop more effective drugs for these difficult-to-treat infections."

Just as having a dedicated center for generating single-cell genomic data "brings together a lot of expertise and makes sure that we get high-quality data out in a timely and cost-efficient manner—the same is true on the data analytics side," explains Kolling.

"The Data Analytics Core, headed by James O'Malley, are our partners and do an excellent job in this regard. And that's been critical because, while new technologies are coming out for generating single-cell data, probably five times as many are coming out for how to analyze it."

When asked to describe the advantage of being able to generate single-cell genomics data over more traditional approaches, Kolling employs a simple analogy. "It's like the difference between being able to examine the individual pieces of fruit in a fruit salad and those same pieces of fruit after they've been blended together to make a fruit smoothie," he says.

"Similarly, if you go to do a normal bulk analysis of a blood sample, which contains, for instance, dozens of immune cell types, you're going to get a mish mash of all the different types of cells that are in that sample. That analysis can provide you with some useful information, but to do personalized medicine you need to know the type of cell that you're targeting in order to develop the treatment—that's what single cell does."

Using this approach has allowed cancer immunologist Mary Jo Turk, PhD, the O. Ross McIntyre, MD Endowed Professor of Microbiology and Immunology at Geisel, to make some impressive gains in characterizing the role that cytotoxic T cells play (in both mouse and human models) in fighting melanoma.

"We're now finding that patients who respond very well to immunotherapy can sustain these small but important populations of T cells that may have attacked a tumor up to nine years ago, and that they still have these T cells circulating through their blood and in their skin, where melanoma arises," explains Turk, a Core user and mentor on the COBRE who also serves as co-director of the Immunology and Cancer Immunotherapy Program at NCCC.

"The only way we could have found them was using a technique called single-cell transcriptomics, which analyzes all of the RNA molecules within a cell, that's supported by the Core—it's really enabling us to do fascinating things with our human work."

Transcriptomic analyses, Turk says, will be key to developing more effective therapies in the future. "Once we can more fully understand what programs these cells and makes them do what they do and persist for so long, we're hoping, within the next five years or so, to be part of efforts to build a better T cell and show that they can go to the right place in the body, hang out for years, and keep patients healthy."

"We're able to evolve the bacteria in the lab under different dynamic conditions, by exposing them to varying levels of drugs, to see how well they respond."

Kata Sasvari

the center for quantitative biology

Rob Strong

Robert Frost and Mary Jo Turk review single-cell genomics data from a melanoma research project.

While genome-wide single cell profiling—which enables the measurement of DNA, RNA, and protein-related processes for thousands to hundreds of thousands of individual cells—now provides unprecedented insights into the biology of complex tissues, it can be extraordinarily challenging.

"In a sense, it's turning big data into even bigger data," says Robert Frost, PhD, who as leader of the CQB's project on gene set testing is developing novel methods for analyzing single-cell data. "And the complexity, noise (biological variations and fluctuations), and sparsity of data at the single-cell level all contribute to making it difficult to interpret and analyze."

Gene set testing is a good approach for contending with these challenges, he explains, "because it allows us to identify small groups of genes that share the same function and operate in the same pathways—so, it's easier to interpret and gives us more accurate and robust results that are less susceptible to noise and sparsity."

Frost is working with several cancer researchers in Geisel's Department of Microbiology and Immunology and the Cancer Center to characterize subsets of immune cells within tumors. "There's a really compelling story for using single-cell profiling techniques to study the tumor microenvironment because there are many different types of cells there and we want to be able to pull each cell out separately, type the cell, and then look at some characteristics of what's happening with its biological program," he says.

"That can lead us to understanding more about interactions between immune cells and between immune cells and tumor cells, and to asking questions like, 'Why are these T cells dormant or deactivated in this case?' and 'Is there another pathway we could target with a drug that would compensate for what the cancer is doing to shut down or evade the immune response?'"

ENHANCING COLLABORATION

In an effort to strengthen such collaborations and continue pushing the field forward, Frost and Kolling host bi-monthly single-cell journal club meetings, which include experimentalists, data scientists, grad students, post-docs, technicians, and technology experts from outside the organization.

"They've really served as a forum for people to come in and present research, discuss papers, and share their expertise on emerging technologies—and we're all benefiting as a result," says Frost.

Similarly, regular meetings between different COBRE teams, both within Dartmouth and externally at collaborating institutions across the region, are helping to ensure that the CQB's Research Administration and Mentoring Core runs smoothly and efficiently.

"Being able to meet with other COBRE teams who are at later phases of their grants and have been generous enough to share their ideas, as well as some of the lessons they've learned along the way, has been invaluable," says Tammara Wood, who assists Whitfield with managing the day-to-day operations of the CQB.

One of the most important goals of the Core, and of the Center as a whole, is to oversee the mentoring and development of current as well as new junior COBRE investigators, and to support their transition to becoming independent researchers.

"Some people can succeed without mentorship, but I believe successes are always greater when you have a dedicated mentor—someone who can read your grants, go to your presentations, and provide honest, supportive feedback—and that's what COBRE encourages," says Turk, who mentors Frost and is part of a large group of senior faculty members who have a mentorship role

"Some people can succeed without mentorship, but I believe successes are always greater when you have a dedicated mentor..."

with the CQB. "If we can bring our junior investigators along to have strong, vibrant research labs, then it benefits all of our programs here."

Whitfield couldn't agree more. "I think it all goes back to building a community of people—including experimentalists, computational experts, and clinician researchers—who have a shared interest in developing single-cell methods and using genomic analyses to further their biological questions. And this has been a great place to do that.

"You know, the environment at Geisel has allowed me, in my own research, the support and freedom to pursue a high-risk project that has the potential to change the lives of patients, and to make connections across schools, departments, and fields of study," he says. "It's tremendously gratifying to be part of a new Center that can offer many others across our institution those same kinds of opportunities."

TIM DEAN IS A SENIOR WRITER FOR *DARTMOUTH MEDICINE.*

Daniel Schultz, holding one of his microfluidics devices used to study antibiotic responses in bacteria.

Rob Strong

COLLABORATION ACROSS CAMPUS

Shutterstock

In addition to the work highlighted in the main story, the Center is engaging in a number of other innovative projects—both within the COBRE grant and across Dartmouth—that while diverse are centered around the CBQ's technologies. These include research by:

- Matthew Havrda, PhD (Molecular & Systems Biology), a Core user who is applying single-cell techniques to better understand Parkinson's disease;

- Brock Christensen, PhD (Epidemiology), a Core user and mentor on the COBRE who is beginning to examine cells from Dartmouth's biorepository that are thought to play a key role in pediatric brain cancers;

- Feng Fu, PhD (Dartmouth College Arts & Sciences), a project leader on the COBRE who is developing mathematical models to predict outcomes in clinical trials for cancer;

- Aaron McKenna, PhD (Molecular & Systems Biology), a Core user and technical adviser to the COBRE who is using advanced genome editing techniques to trace the lineage of where cells come from;

- Matt Mahoney, PhD (The Jackson Laboratory, Bar Harbon, ME), a project leader on the COBRE who is employing machine learning technologies to predict the gene expression in hard-to-access tissues like the brain.